Supplementary Information
Box. Novel computational concepts in metabolomics

Untargeted metabolomics

Big Data Streaming

Uploading data onto informatic software does not require technical expertise, however a major challenge has been the time needed to upload, convert and process metabolomic data files on data processing servers. A solution to the time demands of this upload has been presented in a streaming approach, where data are automatically uploaded from the instrument as they are acquired. This concept of data streaming for complete data processing from upload-convert-process reduces the mean wait time seven fold[1].

Automated workflows

As one of the most laborious steps in the computational workflow is metabolite identification, another concept that has been under development is means to expedite this process [2]. Typically quantitative information is first gathered by running statistical analyses on the obtained mass spectrometry (MS) data, which stratifies ions of interest, giving information about those metabolites associated with the stimulus and then a separate tandem MS experiment to gain the qualitative metabolite identification. In order to implement an autonomous workflow the qualitative and quantitative data are acquired simultaneously via automated tandem MS matching to the METLIN metabolite database[3]. This type of analysis can save weeks of mass spectrometry time, as well as manual and somewhat biased inspection of the data. The autonomous workflow is also now part of the XCMS Online platform.

Targeted metabolomics

Targeted analysis has been generally carried out in a systematic manner, where the metabolites of interest are known, optimized on the instrument (typically a tandem quadrupole mass analyzer) using a standard, and quantified using a concentration curve. This is a hypothesis-driven research where multiple pathways are targeted for quantification of tens to hundreds of metabolites in a single run. Many studies have used targeted metabolomics to analyze metabolites that are part of important biological pathways (glycolysis, TCA cycle, glutaminolysis and many more). In contrast to untargeted analysis, which requires a number of different computational tools for analysis, targeted metabolomics requires minimal data processing. However, there have been some recent novel computational workflows introduced that are

transforming targeted analysis, where hundreds of metabolites can be analyzed in a large-scale high-throughput manner, with links to libraries and other statistical software[4-6]. Other new methods that are coming to the forefront include ways to implement multiple reaction monitoring-based metabolite profiling, thus allowing targeted analysis in a global manner; there has also been an impetus to develop autonomous targeted workflows[7-10].

Data sharing

An important concept which has been brought about by the National Institutes of Health Common Fund's Metabolomics program and from the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI) is the importance of sharing data from untargeted metabolomics datasets, including not only the results but also the experimental metadata[11]. To overcome the major issue of different vendor data formats, all raw mass spectrometry data must be converted into a standardized format (mzML). For detailed metadata upload (ISA-Tab) has proven very successful[12]. At present the main resources for data sharing include XCMS Online[13] and the metabolomeXchange portal. MetabolomeXchange houses datasets from the MetaboLights database[14], Metabolomic Repository Bordeaux[15], and the Metabolomics Workbench. Many of the studies listed in these resources are aimed at elucidating mechanisms of biological action.

References

1. Rinehart, D. et al. Metabolomic data streaming for biology-dependent data acquisition. *Nat Biotechnol* **32**, 524-7 (2014).
2. Benton, H.P. et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal Chem* **87**, 884-91 (2015).
3. Smith, C.A. et al. METLIN - A metabolite mass spectral database. *Ther Drug Monit* **27**, 747-751 (2005).
4. Cai, Y.P., Weng, K., Guo, Y., Peng, J. & Zhu, Z.J. An integrated targeted metabolomic platform for high-throughput metabolite profiling and automated data processing. *Metabolomics* **11**, 1575-1586 (2015).
5. Tsugawa, H., Kanazawa, M., Ogiwara, A. & Arita, M. MRMPROBS suite for metabolomics using large-scale MRM assays. *Bioinformatics* **30**, 2379-2380 (2014).
6. Tsugawa, H. et al. MRM-DIFF: data processing strategy for differential analysis in large scale MRM-based lipidomics studies. *Front Genet* **5**, 471 (2014).
7. Nikolskiy, I., Siuzdak, G. & Patti, G.J. Discriminating precursors of common fragments for large-scale metabolite profiling by triple quadrupole mass spectrometry. *Bioinformatics* **31**, 2017-23 (2015).

8. Luo, P. et al. Multiple Reaction Monitoring-Ion Pair Finder: A Systematic Approach To Transform Nontargeted Mode to Pseudotargeted Mode for Metabolomics Study Based on Liquid Chromatography-Mass Spectrometry. *Anal Chem* **87**, 5050-5055 (2015).
9. Misra, B.B. & van der Hooft, J.J. Updates in Metabolomics Tools and Resources: 2014-2015. *Electrophoresis* (2015).
10. Gu, H., Zhang, P., Zhu, J. & Raftery, D. Globally Optimized Targeted Mass Spectrometry: Reliable Metabolomics Analysis with Broad Coverage. *Anal Chem* **87**, 12355-62 (2015).
11. Rocca-Serra, P. et al. Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* **12**, 14 (2016).
12. Sansone, S.A. et al. Toward interoperable bioscience data. *Nat Genet* **44**, 121-6 (2012).
13. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal Chem* **84**, 5035-9 (2012).
14. Haug, K. et al. MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res* **41**, D781-6 (2013).
15. Ferry-Dumazet, H. et al. MeRy-B: a web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. *BMC Plant Biol* **11**, 104 (2011).