EXPERT REVIEW OF PROTEOMICS
https://doi.org/10.1080/14789450.2018.1476143

Taylor & Francis
Taylor & Francis Group

MEETING REPORT

Integrative omics - from data to biology

Hassan Dihazi[a,b], Abdul R. Asif[a,c], Tim Beißbarth[a,d], Rainer Bohrer[a,e], Kirstin Feussner[f,g], Ivo Feussner[f,g], Olaf Jahn[a,h], Christof Lenz[a,c,i], Andrzej Majcherczyk[a,j], Bernhard Schmidt[a,k], Kerstin Schmitt[a,l], Henning Urlaub[a,c,i] and Oliver Valerius[a,l]

[a]Göttingen Proteomics Forum (GPF), Göttingen, Germany; [b]Nephrology and Rheumatology, University Medical Center Göttingen, University of Göttingen, Göttingen, Germany; [c]Institute for Clinical Chemistry/UMG-Laborateries, University Medical Center Göttingen, University of Göttingen, Göttingen, Germany; [d]Department of Medical Statistics, University Medical Center Göttingen, University of Göttingen, Göttingen, Germany; [e]Gesellschaft für Wissenschaftlische Datenverarbeitung mbH, Göttingen, Germany; [f]Göttingen Metabolomics and Lipidomics Platform (GMLP), Göttingen, Germany; [g]Department of Plant Biochemistry, Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen, Göttingen, Germany; [h]Proteomics Group, Max Planck Institute of Experimental Medicine, Göttingen, Germany; [i]Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany; [j]Büsgen-Institute, Section Molecular Wood Biotechnology and Technical Mycology, University of Göttingen, Göttingen, Germany; [k]Department of Cellular Biochemistry, University Medical Center Göttingen, Göttingen, Germany; [l]Institute for Microbiology and Genetics, University of Göttingen, Göttingen, Germany

ABSTRACT

Introduction: Multi-omic approaches are promising a broader view on cellular processes and a deeper understanding of biological systems. with strongly improved high-throughput methods the amounts of data generated have become huge, and their handling challenging.

Area Covered: New bioinformatic tools and pipelines for the integration of data from different omics disciplines continue to emerge, and will support scientists to reliably interpret data in the context of biological processes. comprehensive data integration strategies will fundamentally improve systems biology and systems medicine. to present recent developments of integrative omics, the göttingen proteomics forum (gpf) organized its 6th symposium on the 23rd of november 2017, as part of a series of regular gpf symposia. more than 140 scientists attended the event that highlighted the challenges and opportunities but also the caveats of integrating data from different omics disciplines.

Expert commentary: The continuous exponential growth in omics data require similar development in software solutions for handling this challenge. Integrative omics tools offer the chance to handle this challenge but profound investigations and coordinated efforts are required to boost this field.

ARTICLE HISTORY
Received 1 March 2018
Accepted 10 May 2018

KEYWORDS
Multi-omics approaches; integrative omics; bioinformatic; systems biology; biostatistics

1. Introduction

The göttingen proteomics forum (GPF) is a local network of scientists from different biological and medical backgrounds with shared expertise in proteomics and mass spectrometry. Its mission is to spread current developments within the proteomics field to life science researchers in the Göttingen area. Over the last years, advances in omics technologies paved the way for this technique to enter clinical practice. Despite the progress achieved and the tools developed in this direction [1–4], the integration of the data generated by this different techniques, which can deliver a more comprehensive and useful picture on diseases, is still challenging. The handling of increasing amounts of omics data and their integration to obtain a more comprehensive picture on biological/medical processes, have greatly increased the demand for cleaver software solutions for data evaluation and integration. After having devoted the last event of our regular symposia series 'Advances in Bioanalytical Mass Spectrometry' to the emerging field of imaging mass spectrometry [5], our topic this year was integrative omics, moving from single to multidimensional omics data sets and their exploitation to deeper understand biological processes. In four sessions, invited and local speakers covered topics addressing multi-omics, bioinformatics, biostatistics, and systems biology. The program and abstracts of all sessions can be found on the homepage of the GPF (http://gpf.gwdg.de/GPF/GPF-events.html).

2. Biological networks in health and disease

Biological networks connect a multitude of individual cellular processes and depend on the controlled intracellular localization and distribution of biomolecules, with their regulated spatiotemporal interactions finally resulting in organism responses. The networks' architectures and dynamics determine cell/organism states and phenotypes. Severe disturbances of balanced regulatory circuits might perturb whole networks and result in disease states. During the last years, multi-omics approaches have increasingly addressed the challenges of data integration and contributed to a deeper understanding of biochemical and molecular processes underlying larger biological networks. . For example, overlapping and integration of data from distinct global omics-screens in comprehensive databases with new interfaces (i.e. bioLOGIC from the Svejstrup lab) led to the uncovering of new factors and cellular pathways involved in the UV-induced DNA damage response [4]. Also for personalized medicine, the development of large biobanks with different layers of omics-

CONTACT Hassan Dihazi dihazi@med.uni-goettingen.de Department of Nephrology and Rheumatology, University Medical Center, Georg-August University Göttingen, Robert-Koch-Strasse 40, D-37075 Germany

data from millions of individuals will enable profound analyses to better understand the cell biology associated with disorders and diseases, and with that for prognosis and treatment even prior to the onset of disease symptoms (numerous prospects reviewed in Karczewski and Snyder [1]).

The first session of our meeting consisted of talks by Bernhard Küster (Proteomics and Bioanalytics, Technical University Munich, Germany), Douglas Armstrong (School of Informatics, University of Edinburgh, Scotland), and Tim Beißbarth (Department of Medical Statistics, University Medical Center Göttingen, Germany). Bernhard Küster presented advantages of applying chemical affinity proteomics to identify and characterize targets of clinically evaluated kinase inhibitors as therapeutic drugs as well as so far poorly characterized kinase inhibitors [6]. These insights will not only promote drug discovery efforts, but will also support the understanding and evaluation of side effects during clinical application, and finally improve clinical decisions. Douglas Armstrong illustrated how already published data together with the help of bioinformatics tools can be used to extract valuable information on protein–protein interaction networks. In the presented study, he used data sets published in 30 different studies dealing with synaptic proteomes. With this approach, the data repository of synaptic protein/protein interactions and underlying functional networks was significantly improved [7]. The session ended with a presentation that addressed bioinformatics and biostatistics approaches and was contributed by GPF member Tim Beißbarth. He focused on bioinformatic solutions for the systematic identification of signaling pathways and their consecutive transcriptional responses stimulated through individually activated receptors. The advantages of the developed methodology for integrative omics was demonstrated for activated B-cell receptor signaling by integrating the phosphoproteome and transcriptome data sets.

3. Bioinformatic and biostatistic tools

Nowadays, advanced computational science has a profound impact on understanding and solving complex networks in life sciences and biomedical research. Targeted omics-analysis of biological processes often results in extended and complex data sets demanding for advanced bioinformatics and statistical tools for their cellular/biological interpretation. The tools available for the omics community, however, confront researchers with new challenges in data analysis. The second session of the GPF symposium focused on the presentation of a variety of tools that support scientists to accurately deal with big data. The session included presentations by Martin Eisenacher (Bioinformatics and Biostatistics for Mass Spectrometry-based Proteomics, Ruhr-University Bochum, Germany) and Jürgen Cox, developer of the MaxQuant and Perseus computational platforms for comprehensive analysis of large-scale (prote)omics data (Max-Planck-Institute for Biochemistry, Munich, Germany).

Martin Eisenacher presented a customer-oriented platform with bioinformatic tools, analysis workflows, and solutions for the medical research needs: BioInfra.Prot. It provides, for example, a protein inference algorithm (PIA) that deals with the issue of protein ambiguity in proteomics and that offers improved overall protein identification [8]. The second presentation of Jürgen Cox introduced the actual features of the MaxQuant/Perseus softwares for large-scale (prote-)omics data analyses. MaxQuant is an integrated computational software with algorithms specifically developed for high-resolution quantitative mass spectrometry analysis, and it enables the processing of multidimensional mass-spectrometric data sets. The Perseus software is a downstream analysis suite with multiple tools for the further biostatistics and bioinformatics evaluation of the primary (e.g. shotgun) proteomics results. It includes various statistical methods and connected visualization tools. Perseus also enables the combination and integration of different omics data (e.g. next-generation sequencing data) and thus helps to extract biological information from different data sets.

4. Omics approaches in metabolism

The third session of the GPF symposium gave insights to the rapidly emerging field of metabolomics, its importance and applications and the use of the metabolite information in integrating omics data. The session included talks from Asaph Aharoni (Department of Plant & Environmental Sciences, Faculty of Biochemistry, Weizmann Institute of Science, Rehovot, Israel), Alexander Karabatsiakis (Clinical & Biological Psychology, Institute of Psychology and Education, Ulm University, Germany) and Kirstin Feussner (Albrecht-von-Haller-Institute of Plant Sciences, Department of Plant Biochemistry, University of Göttingen, Germany).

Metabolic pathways in a biological system are tailored to satisfy its needs. In his research, Asaph Aharoni uses integrated omics approaches to unravel networks of genes and proteins involved in the regulation of plant metabolism [9]. In his talk, he highlighted several advanced tools for studying metabolomes and the advantage of integrating metabolomics, genetics and informatics data. The advantage of the combination of the different tools in a single study was shown in investigations combining coexpression analysis and metabolomics to unravel the secondary metabolism of nightshades (Solanaceae). The second talk was given by Alexander Karabatsiakis [10]. He highlighted the advantage of combining metabolomics with lipidomics data to investigate the mechanisms underlying the physical and mental health problems in adults who suffer from childhood maltreatment (CM). The data presented showed that with this integrative omics approach it was possible to identify markers of inflammation and oxidative stress, which allowed to differentiate between individuals with and without CM. Moreover, the data presented suggest pathways that may be involved in the effect of CM on health and disease. The third talk provided new insights into approaches that combine different omics workflows and was given by Kirstin Feussner, representing the Göttingen Metabolomics & Lipidomics Platform (GöMLP, weblink). Here, nontargeted metabolomics combined with lipidomics, hormone analyses, and transcriptomics data by the software tool MarVis gave insights into plant stress adaptation and allowed to identify new wound-related markers and pathways.

5. Multi-omics data integration

The last session of the GPF-symposium dealt with multi-omics data integration. It included talks by Oliver Kohlbacher (Center for Bioinformatics, Eberhard Karls University Tübingen and Max

Table 1. Collection of tools for omic-data integration.

Tool name	Website	Reference
VANTED	http://vanted.ipk-gatersleben.de/	[13]
Confero	http://sourceforge.net/projects/confero/	[14]
DAnTE	http://omics.pnl.gov/software/DanteR.php	[15]
MADMAX	http://madmax2.bioinformatics.nl	[16]
MCIA	http://www.bioconductor.org/packages/release/bioc/html/omicade4.html	[17]
InCroMAP	http://www.ra.cs.uni-tuebingen.de/software/InCroMAP/http://cran.r-project.org/web/packages/qtlnet/	[18]
Integromics	http://www.integromics.com/	[19]
MONA	http://sourceforge.net/projects/monaos/	[20]
NetGestalt	http://www.netgestalt.org/	[21]
Joint NMF	http://zhoulab.usc.edu/SNMNMF/	[22]
iCluster	http://www.mskcc.org/mskcc/html/85130.cfm	[23]
JIVE	https://genome.unc.edu/jive/	[24]
Joint Bayes Factor	https://sites.google.com/site/jointgenomics/	[25]
sMBPLS	http://zhoulab.usc.edu/sMBPLS/	[26]
SNPLS	http://page.amss.ac.cn/shihua.zhang/	[27]
MDI	http://www2.warwick.ac.uk/fac/sci/systemsbiology/research/software/	[28]
BCC	http://people.duke.edu/%7Eel113/software.html	[29]
PARADIGM	http://sbenz.github.com/Paradigm	[30]
SNF	http://compbio.cs.toronto.edu/SNF/	[31]
Lemon-Tree	http://lemon-tree.googlecode.com	[32]
ATHENA	https://ritchielab.psu.edu/software/athena-downloads	[33]
FSMKL	https://github.com/jseoane/FSMKL	[34]
iBAG	http://odin.mdacc.tmc.edu/~vbaladan/	[35]
Anduril	http://csbi.ltdk.helsinki.fi/anduril/	[36]
MKGI	https://ritchielab.psu.edu/software/	[37]

Planck Institute for Developmental Biology, Tübingen, Germany) and Philip Stegmaier (geneXplain, Wolfenbüttel, Germany). Oliver Kohlbacher's talk entitled 'Going multi-omics – Many issues and (perhaps) a few solutions' highlighted challenges in analyzing and integrating data from multi-omics data sets. Despite the enormous efforts and the success achieved in developing tools for multi-omics data (Table 1 presents a collection of tools developed for Multi-omics data interpretation and integration), the talk made clear that data management, data integration, and improved algorithms are urgently needed and that the reproducibility of multi-omics analyses is still a big issue. Integrated data analysis workflows can automate the analysis and thus enable high-throughput analyses as well as support reproducible science. Workflow-enabled software tools like OpenMS [11] have been designed with this goal in mind. Another component required for integrated data analysis are data management solutions. qPortal [12] is a web-based tool that provides a convenient web-based interface to access big omics data sets in a user-friendly manner and is directly integrated with high-performance computing environments.

The second talk was on Multi-omics upstream analysis and was given by Philip Stegmaier. He presented tools for the analysis of multi-omics data sets. The approach is based on so called 'upstream analysis' [38] which investigates DNA promotor regions to identify potential transcription factors, which may be responsible for expression changes.

6. Future directions and challenges

The comprehension of systems biology and systems medicine will significantly profit from the increasing number of omics projects in biological and medical research, the increasing power and capacities of omics-data acquisition, and the ongoing boost in computational power. The development of software solutions for data handling and integration of various types of omics-experiments, for format conversions, and the expansion of public repositories are in continuous progress so that obstacles in data accessibility and adaptability will diminish. The channeling of different types of technologies into common analysis pipelines also requires that individual centers of different omics expertises, nowadays often core facilities, form new networks for collaborations to perform multiomics projects. Especially efforts in expanding existing bioinformatics and biostatistics platforms are required to ensure state-of-the-art analysis tools, modern data-processing infrastructure, and the capability to develop individual project-adapted analysis strategies. The existing conventions and guidelines for the quality management of individual omics data should be adapted for multiomics data analyses. The documentation of data analysis should be accompanied by detailed comprehensible descriptions of the data cross-correlation. Pilot studies and directions in software development and network initiatives were, therefore, discussed at the GPF symposium by leading scientists in the field.

7. Conclusion

Driven by the advances in high-throughput technologies, omics data are growing exponentially at present. State-of-the-art bioinformatics tools and strategies that enable integration of different omics data in complex biological systems are key requirements to understand molecular systems including the underlying biological processes, and to promote translational research. From the presentations at the GPF symposium, it became obvious that integrative omics is a highly promising but also challenging field in life sciences and that profound investigations and coordinated efforts in bioinformatics and biostatistics are required to connect individual developments. One main challenge that persists is the heterogeneity of data formats delivered by the different techniques. Integration of more than two different omics data formats is still not routine and requires optimized software tools together with well-trained scientists to generate reasonable and comprehensible analysis workflows for big data.

Funding

We are grateful for financial support from the Deutsche Forschungsgemeinschaft (DFG, GRK 2172 PRoTECT - Plant Responses To Eliminate Critical Threats) and the Universitätsbund Göttingen. We would like to thank the following companies for their contributions and support for the realization of the symposium: Agilent, Bruker, Sartorius, Sciex, Thermo Fisher Scientific, and Waters Cooperation.

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript

apart from those disclosed. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

1. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nat Rev Genet. 2018;19:299–310.
2. Ryan CJ, Cimermančič P, Szpiech ZA, et al. High-resolution network biology: connecting sequence with function. Nat Rev Genet. 2013;14:865–879.
3. Cowen L, Ideker T, Raphael BJ, et al. Network propagation: a universal amplifier of genetic associations. Nat Rev Genet. 2017;18:551–562.
4. Boeing S, Williamson L, Encheva V, et al. Multiomic analysis of the UV-Induced DNA damage response. Cell Rep. 2016;15(7):1597–1610.
5. Dihazi H, Bohrer R, Jahn O, et al. 5th Symposium of the Göttingen proteomics forum "Mass spectrometry with spatial resolution: MALDI-imaging and laser capture microscopy". Expert Rev Proteomics. 2013;10(1):17–20.
6. Klaeger S, Heinzlmeir S, Wilhelm M, et al. The target landscape of clinical kinase drugs. Science. 2017;358:6367.
7. Alfieri A, Sorokina O, Adrait A, et al. Synaptic interactome mining reveals p140Cap as a New Hub for PSD proteins involved in psychiatric and neurological disorders. Front Mol Neurosci. 2017;10:212.
8. Turewicz M, Kohl M, Ahrens M, et al. BioInfra.Prot: a comprehensive proteomics workflow including data standardization, protein inference, expression analysis and data publication. J Biotechnol. 2017;261:116–125.
9. Sonawane PD, Pollier J, Panda S, et al. Aharoni plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. Nat Plants. 2016;3:16205.
10. Karabatsiakis A, Hamuni G, Wilker S, et al. Metabolite profiling in posttraumatic stress disorder. J Mol Psychiatry. 2015;3(1):2.
11. Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods. 2016;13(9):741–748.
12. Mohr C, Friedrich A, Wojnar D, et al. qPortal: a platform for data-driven biomedical research. PloS One. 2018;13(1):e0191603.
13. Klukas C, Schreiber F. Integration of -omics data and networks for biomedical research with vanted. J Integr Bioinform. 2010;7(2):112.
14. Hermida L, Poussin C, Mb S, et al. Confero: anintegrated contrast data and gene set platform for computational analysis and biological interpretation of omics data. BMC Genomics. 2013;14:514.
15. Polpitiya AD, Qian WJ, Jaitly N, et al. DAnTE: a statistical tool for quantitative analysis of omics data. Bioinformatics. 2008;24 (13):1556–1558.
16. Lin K, Kools H, de Groot PJ, et al. MADMAX – manage-ment and analysis database for multiple omics experiments. J Integr Bioinform. 2011;8(2):160.
17. Meng C, Kuster B, Ac C, et al. A multivariate approach to the integration of multi-omics datasets. BMC Bioinform. 2014;15:162.
18. Wrzodek C, Eichner J, Buchel F, et al. InCroMAP: integrated analysis of cross-platform microarray and pathway data. Bioinformatics. 2013;29(4):506–508.
19. Le Cao KA, Gonzalez I, Dejean S. Integromics: an R package to unravel relationshipsbetween two omics datasets. Bioinformatics. 2009;25(21):2855–2856.
20. Sass S, Buettner F, Mueller NS, et al. A modular framework for gene set analysisintegrating multilevel omics data. Nucleic Acids Res. 2013b;41(21):9622–9633.
21. Shi Z, Wang J, Zhang B. NetGestalt: integrating multidimensional omics data overbiological networks. Nat Methods. 2013;10(7):597–598.
22. Zhang S, Li Q, Liu J, et al. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. Bioinformatics. 2011;27:i401–i409.
23. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics. 2009;25:2906–2912.
24. Lock EF, Hoadley KA, Marron JS, et al. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. Ann Appl Stat. 2013;7:523–542.
25. Ray P, Zheng L, Lucas J, et al. Bayesian joint analysis of heterogeneous genomics data. Bioinformatics. 2014;30:1370–1376.
26. Li W, Zhang S, Liu CC, et al. Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. Bioinformatics. 2012;28:2458–2466.
27. Chen J, Zhang S. Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data. Bioinformatics. 2016;32:1724–1732.
28. Kirk P, Griffin JE, Savage RS, et al. Bayesian correlated clustering to integrate multiple datasets. Bioinformatics. 2012;28:3290–3297.
29. Lock EF, Dunson DB. Bayesian consensus clustering. Bioinformatics. 2013;29:2610–2616.
30. Vaske CJ, Benz SC, Sanborn JZ, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics. 2010;26:i237–i245.
31. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11:333–337.
32. Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with lemon-tree. PLoS Comput Biol. 2015;11: e1003983.
33. Kim D, Li R, Dudek SM, et al. ATHENA: identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. BioData Min. 2013;6:23.
34. Seoane JA, Day INM, Gaunt TR, et al. A pathwaybased data integration framework for prediction of disease progression. Bioinformatics. 2014;30:838–845.
35. Wang W, Baladandayuthapani V, Morris JS, et al. iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data. Bioinformatics. 2013;29(2):149–159.
36. Ovaska K, Laakso M, Haapa-Paananen S, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. Genome Med. 2010;2:65.
37. Kim D, Li R, Lucas A, et al. Using knowledge-driven genomic interactions for multi-omics data analysis: metadimensional models for predicting clinical outcomes in ovarian carcinoma. J Am Med Inform Assoc. 2016;24:577–587.
38. Koschmann J, Bhar A, Stegmaier P, et al. Upstream analysis: an integrated promoter-pathway analysis approach to causal interpretation of microarray data. Microarrays (Basel). 2015;4(2): 270–286.