Journal of proteome research

Letter

pubs.acs.org/jpr

OMICS Studies: How about Metadata Checklist and Data Publications?

Eugene Kolker* and Elizabeth Stewart

Seattle Children's Research Insitute, 1900 Ninth Avenue, Seattle, Washington 98101, United States

ABSTRACT: Data fully utilized by the community resources promote progress rather than repetition. Effective data sharing can accelerate the transition from data to actionable knowledge, yet barriers to data sharing remain, both technological and procedural. The DELSA community has tackled the sharing barrier by creating a multi-omics metadata checklist for the life sciences. The checklist and associated data publication examples are now jointly published in *Big Data* and *OMICS: A Journal of Integrative Biology*. The checklist will enable diverse datasets to be easily harmonized and reused for richer analyses. It will facilitate data deposits, stand alone as a data publication, and grant appropriate credit to researchers. We invite the broader life sciences community to test the checklist for feedback and improvements.

Checklist Version	Version 1.0
Experiment Information	**Description**
Lab Name	Lab conducting the experiment
Date	Checklist submission date
Author Information	Name, organization, contacts
Title of Experiment	One-sentence description of the particular experiment
Project	Project name, ID, organization
Funding	Funding sources for the project
Digital ID	Multiple IDs may be listed, such as those to GEO, MOPED, PRIDE, DOIs, etc.
Abstract	A short description of the experiment briefly stating the goals of the research and principal outcomes if any (100 words or less)
Experimental Design	**Description**
Organism	For example, human, mouse
OMICS Type(s) Utilized	For example, proteomics, metabolomics
Reference	Published articles that utilize these data, their PubMed identifier (PMID), or other relevant IDs or links
Experimental Design	Design Specifications; type of replication (biological, technical, time

KEYWORDS: *metadata, multi-omics, OMICS, checklist, community, data publication, life sciences, Big Data*

We learn it early in childhood − sharing is good. As researchers, we know that sharing data serves not only the common good but also the individual. Data fully utilized by the community resources promote progress rather than repetition. The individual can gain greater insight with less effort as others contribute to the goal. Yet the barriers to effective sharing can be significant, starting with the often unrecognized effort it takes to get the data into an accessible repository. In addition, that effort will only realize its full impact if not only the data but also rigorously and consistently collected metadata are shared.

Recently, the Nature Publishing Group unveiled a publication checklist to ensure more complete reporting of specific experimental information regarding experimental design, statistical analysis, and reagents.[1] The Data-Enabled Life Sciences Alliance (DELSA Global) has endorsed this checklist as an example of the type of enabling tool that is needed to more fully utilize shared data.[2] Other types of data-sharing tools include, for example, the single omics checklists such as Minimum Information About a Proteomics Experiment (MIAPE) by the Human Proteome Organization[3−5] and the Minimum Information About a Microarray Experiment (MIAME) by the Microarray Gene Expression Data Society.[6]

The DELSA community has tackled the sharing barrier by creating a multi-omics metadata checklist for the community. The checklist and associated data publication examples are being jointly published in *Big Data* and *OMICS: A Journal of Integrative Biology*.[7−10] The checklist is a simple yet comprehensive tool to ensure that shared data are both easily harmonized and reused for richer analyses. It will facilitate data deposits, stand alone as a data publication, and grant appropriate credit to researchers. The specifics of each omics data-type will fit into the comprehensive structure of the checklist, thus building common bridges between data sets. Data shared through diverse established databases such as Pride,[11] ProteomeXchange,[11,12] MOPED,[13] GEO,[14] and Metabolights[15] can be found faster with a common format searchable checklist.

Life sciences research cannot capture the complete picture of interactions without combining complementary multidisciplinary data. Single omics studies are used to their fullest potential when combined with complementary omics studies. Rich multi-omics data sets produced by these studies, if described with proper metadata, will significantly accelerate life sciences research. We invite the broader life sciences community to test the checklist for feedback and improvements.

AUTHOR INFORMATION

Corresponding Author

*E-mail: eugene.kolker@seattlechildrens.org.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Editorial. Announcement: Reducing our irreproducibility. *Nature* **2013**, *496*, 398.

(2) Kolker, E.; Ilkay, A.; Bourne, P.; Faris, J.; Fox, G.; Frishman, D.; Geraci, C.; Hancock, W.; Lin, B.; Lancet, D.; Lisitsa, A.; Knight, R.; Martens, L.; Mesirov, J.; Özdemir, V.; Schultes, E.; Smith, T.; Snyder,

Received: November 26, 2013

   dx.doi.org/10.1021/pr4011662 | *J. Proteome Res.* XXXX, XXX, XXX−XXX

M.; Srivastava, S.; Toppo, S.; Wilmes, P. In praise of open research measures. *Nature* **2013**, *498*, 170.

(3) Taylor, C. F.; Paton, N. W.; Lilley, K. S.; Binz, P.-A.; Julian, R. K., Jr; Jones, A. R.; Zhu, W.; Apweiler, R.; Aebersold, R.; Deutsch, E. W.; Dunn, M. J.; Heck, A. J. R.; Leitner, A.; Macht, M.; Mann, M.; Martens, L.; Neubert, T. A.; Patterson, S. D.; Ping, P.; Seymour, S. L.; Souda, P.; Tsugita, A.; Vandekerckhove, J.; Vondriska, T. M.; Whitelegge, J. P.; Wilkins, M. R.; Xenarios, I.; Yates, J. R., III; Hermjakob, H. The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **2007**, *25*, 887−893.

(4) Paik, Y. K.; Omenn, G. S.; Thongboonkerd, V.; Marko-Varga, G.; Hancock, W. S. Genome-wide Proteomics, Chromosome-centric Human Proteome Project (C-HPP), Part II. *J. Proteome Res.* **2014**, *13*, 1−4.

(5) Omenn, G. S. The strategy, organization, and progress of the HUPO Human Proteome Project. *J. Proteomics*. **2014**, 10.1016/j.jprot.2013.10.012.

(6) Alvis, B.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, C. A.; Causton, H. C.; Gaasterland, T.; Glenisson, P.; Holstege, F. C.; Kim, I. F.; Markowitz, V.; Matese, J. C.; Parkinson, H.; Robinson, A.; Sarkans, U.; Schulze-Kremer, S.; Stewart, J.; Taylor, R.; Vilo, I.; Vingron, M. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **2001**, *29* (4), 365−371.

(7) Kolker, E.; Özdemir, V.; Martens, L.; Hancock, W.; Anderson, G.; Naderson, N.; Aynacioglu, Ş.; Baranova, A.; Campagna, S. R.; Chen, R.; Choiniere, J.; Dearth, S. P.; Feng, W.-C.; Ferguson, L.; Fox, G.; Frishman, D.; Grossman, R.; Heath, A.; Higdon, R.; Hutz, M. H.; Janko, I.; Jiang, L.; Joshi, S.; Kel, A.; Kemnitz, J. W.; Kohane, I. S.; Kolker, N.; Lancet, D.; Lee, E.; Li, W.; Lisitsa, A.; Llerena, A.; MacNealy-Koch, C.; Marshall, J.-C.; Masuzzo, P.; May, A. L.; Mias, G.; Monroe, M.; Montague, E.; Mooney, S.; Nesvizhskii, A.; Noronha, S.; Omenn, G.; Rajasimha, H. K.; Ramamoorthy, P.; Sheehan, J.; Smarr, L.; Smith, C. V.; Smith, T.; Snyder, M.; Rapole, S.; Srivastava, S.; Stanberry, L.; Stewart, E.; Toppo, S.; Uetz, P.; Verheggen, K.; Voy, B. H.; Warnich, L.; Wilhelm, S. W.; Yandl, G. Towards more transparent and reproducible omics studies through a common metadata checklist and data publications. *OMICS: J. Integr. Biol.* **2014**, *1*, 10−14.

(8) Snyder, M.; Mias, G.; Stanberry, L.; Kolker, E. Metadata checklist for the integrated personal omics study: proteomics and metabolomics experiments. *OMICS: J. Integr. Biol.* **2014**, *1*, 81−85.

(9) Kolker, E.; Özdemir, V.; Martens, L.; Hancock, W.; Anderson, G.; Naderson, N.; Aynacioglu, Ş.; Baranova, A.; Campagna, S. R.; Chen, R.; Choiniere, J.; Dearth, S. P.; Feng, W.-C.; Ferguson, L.; Fox, G.; Frishman, D.; Grossman, R.; Heath, A.; Higdon, R.; Hutz, M. H.; Janko, I.; Jiang, L.; Joshi, S.; Kel, A.; Kemnitz, J. W.; Kohane, I. S.; Kolker, N.; Lancet, D.; Lee, E.; Li, W.; Lisitsa, A.; Llerena, A.; MacNealy-Koch, C.; Marshall, J.-C.; Masuzzo, P.; May, A. L.; Mias, G.; Monroe, M.; Montague, E.; Mooney, S.; Nesvizhskii, A.; Noronha, S.; Omenn, G.; Rajasimha, H. K.; Ramamoorthy, P.; Sheehan, J.; Smarr, L.; Smith, C. V.; Smith, T.; Snyder, M.; Rapole, S.; Srivastava, S.; Stanberry, L.; Stewart, E.; Toppo, S.; Uetz, P.; Verheggen, K.; Voy, B. H.; Warnich, L.; Wilhelm, S. W.; Yandl, G. Towards more transparent and reproducible omics studies through a common metadata checklist and data publications. *Big Data* **2013**, *1* (4), 196−201.

(10) Snyder, M.; Mias, G.; Stanberry, L.; Kolker, E. Metadata checklist for the integrated personal omics study: proteomics and metabolomics experiments. *Big Data* **2013**, *1* (4), 202−206.

(11) Vizcaíno, J. A.; Côté, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **2013**, *41* (Database issue), D1063−D1069, http://www.ebi.ac.uk/pride/.

(12) Hermjakob, H.; Apweiler, R. The Proteomics Identifications Database (PRIDE) and the ProteomeXchange Consortium: making proteomics data accessible. *Expert Rev. Proteomics* **2006**, *3*, 1−3, http://www.proteomexchange.org.

(13) Higdon, R.; Stewart, E.; Stanberry, L.; Haynes, W.; Choiniere, J.; Montague, E.; Anderson, N.; Yandl, G.; Janko, I.; Broomall, W.; Fishilevich, S.; Lancet, D.; Kolker, N.; Kolker, E. MOPED enables discoveries through consistently processed proteomics data. *J. Proteome Res.* **2014**, *13* (1), 107−113, http://www.moped.proteinspire.org.

(14) Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A. NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* **2013**, *41* (Database issue), D991−D995, http://www.ncbi.nlm.nih.gov/geo/.

(15) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendrakar, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; González-Beltrán, A.; Sansone, S.-A.; Griffin, J. L.; Steinbeck, C. MetaboLights– an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **2013**, *41* (Database issue), D781−D786, http://www.ebi.ac.uk/metabolights/about.