OMICS A Journal of Integrative Biology
Volume 18, Number 1, 2014

DOI: 10.1089/omi.2013.0148

Letter to the Editor

Metadata Checklist for the Integrated Personal OMICS Study: Proteomics and Metabolomics Experiments

Michael Snyder,[1–3] George Mias,[1,3] Larissa Stanberry[3–5] and Eugene Kolker[3–5]

To the Editor:

The integrative personal omics profiling study introduced a novel, integrative approach based on personalized, longitudinal, multi-omics data. The study collected genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from a single individual over a 14-month period. The results revealed various medical risks and extensive dynamic changes in diverse molecular components and biological pathways across healthy and diseased conditions. The current letter to the editor is a data publication that provides the checklists for the metadata of the proteomics (Table 1) and metabolomics (Table 2) datasets of the study. The proposed checklist was recently developed and endorsed by the Data-Enabled Life Sciences Alliance (DELSA Global). We call for the broader use of data publications using the metadata checklist to make omics data more discoverable, interpretable, and reusable, while enabling appropriate attribution to data generators and infrastructure science builders.

Table 1. Proteomics Metadata Checklist

Checklist Version	*1.0 (ref: Kolker, 2014)*
Experiment information	*Description*
Lab name	Synder Lab, Department of Genetics, Stanford University
Date	October 24, 2013
Author information	George Mias
Title of experiment	Integrated Personal Omics Profiling
Project	Integrated Personal Omics Profiling
Funding	Stanford University, NIH training grant, NIH/NLM training grant T15-LM007033, NIH/NIGMS R24-GM61374; the Spanish Ministry of Science and Innovation Projects SAF2008-05384 and CSD2007-00017; European Union FP7 Projects 2007-A-201630 (GENICA) and 2007-A-200950 (TELOMARKER); European Research Council Advanced Grant GA232854; the Körber Foundation, the Fundación Marcelino Botín, and Fundación Lilly (España); NIH/NHLBI training grant T32 HL094274; NIH/NHLBI KO8 HL083914; NIH New Investigator DP2 Award OD004613; the Breetwor Family Foundation. G.M.'s research is supported by the National Human Genome Research Institute of the National Institutes of Health under award number K99HG007065, and previously T32HG000044.
Digital ID	1_2013; MOPED (moped.proteinspire.org) experiment: snyder_personal_omics_profiling; Peptide Atlas: PASS00062; Open Science Data Cloud keyservice.opensciencedatacloud.org/ark:/31807/synder-001
Abstract	This study presented an integrative personal omics profile that combined genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from a single individual over a 14-month period. The study revealed various medical risks and uncovered extensive, dynamic changes in diverse molecular components and biological pathways across healthy and diseased conditions. The current checklist provides the metadata for the proteomics part of the study.

(*continued*)

[1]Department of Genetics and [2]Stanford Center for Genomics and Personalized Medicine, Stanford University, Stanford, California.
[3]Data-Enabled Life Sciences Alliance (DELSA Global), Seattle, Washington.
[4]Bioinformatics and High-Throughput Analysis Laboratory, Seattle Children's Research Institute, and [5]Predictive Analytics, Seattle Children's, Seattle, Washington.

Table 1. (Continued)

Experimental design	
Organism	Human
OMICS type(s) utilized	Proteomics
Reference	Cell 148(6), 1293-1307, 2012, (PMID 22424236) (Chen, 2012)
Experimental design	Longitudinal data collected on a single subject
Sample description	Samples were taken on a single individual over a 14-month period
Tissue/cell type ID*	Peripheral blood mononuclear cells
Localization ID	Cell
Condition ID	Healthy state, respiratory syncytial virus (RSV), and human rhinovirus (HRV) infections (time specific)
Experimental methods	
Sample prep description	Whole-blood samples were collected at each time point, and peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation at $400 \times g$ for 25 min using the Lymphocyte Separation Media (MP Biomedicals). Serum and plasma were also collected for each time point. Samples were lysed in 10X volume of buffer containing 4% sodium dodecyl sulfate, and 100 mM dithiothreitol in 100 mM Tris/HCl (pH 8.0). Samples were incubated at 95°C for 5 min and sonicated. Detergent was removed from the samples using the Filter Aided Sample Preparation (FASP) with YM-30 micron filter units (Cat No. MRCF0R030, Millipore). About 200 μL of 8 M urea in 0.1 M Tris/HCl (pH 8.5) was added, and samples were centrifuged at $14{,}000 \times g$ at 20°C for 15 min. This step was repeated three times. About 50 μL of 0.05 M iodoacetamide in 8 M urea was added to the filters, and the samples were incubated in darkness for 1 hr. Sample was washed three times with 100 μL of 200 mM ThAB. Protein concentration was measured using the Bradford method. Finally, trypsin (Promega, Madison, WI) was added at a protein-to-enzyme ratio of 50:1. Samples were incubated overnight at 37°C. Peptides were collected by centrifugation and labeled with TMT 6-plex Reagent. The TMT Label Reagents were equilibrated to room temperature. For 0.8 mg vials, 41 μL of anhydrous acetonitrile was added to each tube and 41 μL of the TMT Label Reagent was added to each 25–100 μg sample. The reaction mixture was incubated for 1 hr at room temperature. About 8 μL of 5% hydroxylamine was added to the sample and incubated for 15 min. Samples were combined at equal amounts and dried by speed vac.
Platform type	LC MS/MS
Instrument name	Waters NanoAquity 2D nLC and (LTQ)-Orbitrap Velos (Thermo Fisher Scientific)
Instrument details	The LC system was directly coupled in-line with a linear trap quadrupole (LTQ)-Orbitrap Velos instrument (Thermo Fisher Scientific) via a Thermo nanoelectrospray source. The source was operated at 2.2–2.4 kV to optimize the nanospray, with the ion transfer tube at 200°C.
Instrument protocol	*Peptide separation:* The protein sample was resuspended in 100 mM ammonium formate at pH 10 and loaded to the LC system. Peptides were separated by reverse phase chromatography at a high pH in the first dimension, followed by an orthogonal separation at a low pH in the second dimension. An online dilution of the effluent was performed after the first dimension to ensure that no peptides were lost before the second dimension. In the first dimension, the mobile phases were solvent A, 20 mM ammonium formate at pH 10, and solvent B, acetonitrile. Peptides were separated on an Xbridge 300 mm × 5 cm C18 5.0 mm column (Waters) using 14 discontinuous step gradients at 2 mL/min. Acetonitrile concentration for each step was adjusted to ensure nearly equivalent peptide load and MS intensity for each second-dimension run. To maximize peptide recovery, the fractions were diluted online using 0.1% formic acid in water at 20 mL/min and then trapped by Symmetry 180 mm × 2 cm C18 5.0 mm trap column (Waters). In the second dimension, peptides were loaded to an in-house packed 75 mm ID/15 mm tip ID x 20 cm C18-AQ 3.0 mm resin column with solvent A (0.1% formic acid in water). Peptides were separated with a linear gradient from 5% to 30% solvent B (0.1% formic acid in acetonitrile) at a flow rate of 300 nL/min in 180 min. Each sample separation was repeated three times. *Proteomics MS analysis:* The mass spectrometer was run in a data-dependent mode. One survey scan acquired in the Orbitrap mass analyzer with resolution 60,000 at m/z 400 was followed by MS/MS of the 10 most intense peaks with charge state of 2+ and above an intensity threshold of 5,000 counts. MS/MS fragmentation was done in the high collisional cell with normalized collision energy of 40% eV and activation time of 0.1 sec. The MS/MS scan was acquired in the Orbitrap at a resolution of 7,500. Dynamic exclusion was enabled to minimize repeated sequencing. Peaks selected for fragmentation more than once within 30 sec were excluded from selection (10 ppm window) for 60 sec.

(*continued*)

TABLE 1. (CONTINUED)

Data processing	
Processing/normalization methods/software	Mass tolerance was 10 ppm for the precursor ion and 0.02 Da for fragment ions. Cystine carbamidomethylation was included as a fixed modification, with n-terminal and lysine TMT 6plex modification and methionine oxidation as variable modifications. Up to two missed cleavages were allowed. Only unique peptides with minimum 6 amino acid length were considered for protein ID. The median value of different peptide ratios was used for protein quantitation. Spectra were obtained from three TMT-labeled samples. Proteins were identified at a false discovery rate <0.01 and requiring at least two unique peptides per protein. For relative quantitation, each time point was compared with a healthy time point, day 255, and all ratios were normalized to have a unit mean. After protein identification, the three sets were matched using a replicated common ratio present in all three. QC assessment required a CV <0.13 for the replicates; that the reference (day 255) mass tag be always present in all three samples; and that a minimum of 2/3 points be present for all proteins identified. The log-2 relative ratios were vector normalized to 1; a nonparametric bootstrap distribution ($n>100{,}000$ samples) was constructed by sampling each time point with replacement.
Sequence/annotation database	IPI human database, v 3.75
ID method/software	Protein Discoverer (Thermo)
ID/expression measures	Log2 expression, expression ratios
Data analysis method/ software	Clustering, pathway analysis, custom R, Mathematica, Python scripts
I/O data file formats	Tab delimited csv
Additional Information	None

CV, coefficient of variation; ID, identification.

TABLE 2. METABOLOMICS METADATA CHECKLIST

Checklist Version	*1.0 (ref: Kolker, 2014)*
Experiment information	*Description*
Lab Name	Snyder Lab, Department of Genetics, Stanford University
Date	October 24, 2013
Author Information	George Mias, Somallee Datta
Title of Experiment	Integrated Personal Omics Profiling
Project	Integrated Personal Omics Profiling
Funding	Stanford University, NIH training grant, NIH/NLM training grant T15-LM007033, NIH/NIGMS R24-GM61374; the Spanish Ministry of Science and Innovation Projects SAF2008-05384 and CSD2007-00017; European Union FP7 Projects 2007-A-201630 (GENICA) and 2007-A-200950 (TELOMARKER); European Research Council Advanced Grant GA232854, the Körber Foundation, the Fundación Marcelino Botín, and Fundación Lilly (España); NIH/ NHLBI training grant T32 HL094274; NIH/NHLBI KO8 HL083914; NIH New Investigator DP2 Award D004613; the Breetwor Family Foundation. G.M.'s research is supported by the National Human Genome Research Institute of the National Institutes of Health under award number K99HG007065, and previously T32HG000044. NSF/DBI award 0969929, NIH/ NIDDK awards U01-DK-089571 and U01-DK-072473, The Robert B. McMillen Foundation, The Gordon and Betty Moore Foundation, and Seattle Children's Research Institute.
Digital ID	2_2013; Open Science Data Cloud keyservice.opensciencedatacloud.org/ark:/31807/synder-002
Abstract	This study presented an integrative personal omics profile that combined genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from a single individual over a 14-month period. The study revealed various medical risks and uncovered extensive, dynamic changes in diverse molecular components and biological pathways across healthy and diseased conditions. The current checklist provides the metadata for the metabolomics part of the study.
Experimental design	
Organism	Human
OMICS type(s) utilized	Metabolomics
Reference	Cell 148(6), 1293-1307, 2012, (PMID 22424236) (Chen, 2012)
Experimental design	Longitudinal data collected on a single subject
Sample description	Samples were taken on a single individual over a 14 month period

(*continued*)

Table 2. (Continued)

Tissue/cell type ID	Blood serum
Localization ID	Cell
Condition ID	Healthy state, RSV and HRV infections (time specific)
Experimental methods	
Sample prep description	About 100 μL of the serum sample was used for the metabolomics study. Metabolites were extracted by adding four times volume of equal-volume mixture of methanol, acetonitrile, and acetone that were prechilled at −20°C. To maximize metabolite extraction, samples were vortex at 4°C for 15 min at 2 min intervals. Proteins were precipitated by incubating the sample at −20°C for 2 hr. Samples were then centrifuged at 10,000 rpm at 4°C for 10 min. The supernatant was collected and dried for metabolomics analysis. For each time point, three of the 100 μL samples were analyzed in triplicate.
Platform type	LC-MS and LC-MS/MS
Instrument name	Agilent 1260 LC system and Agilent 6538 Q-TOF MS
Instrument details	Coupled in-line with Agilent 6538 Q-TOF MS with electrospray ionization.
Instrument protocol	The LC mobile phases consisted of 0.2% acetic acid in water (solvent A) and 0.2% acetic acid in methanol (solvent B). The extract was resuspended in 50% methanol and sonicated for 5 min. The sample was loaded to an Agilent SB-aq 1.8 μm, 2.1 × 50 mm analytical column with a SB-C8 3.5 μM, 2.1 × 30 mm guard column in front. Columns were heated to 60°C with a flow rate of 0.6 mL/min. A linear gradient from 2% to 98% solvent B in 13 min was used for metabolites separation. To assure the mass accuracy of the recorded ions, continuous internal calibration ions were infused in-line through the dual electrospray ionization (ESI) source using an isocratic pump at flow rate of 0.05 mL/min. Internal calibrants at *m/z* 121.0509 and 922.0098 were used in positive ion mode and *m/z* of 119.0362 and 980.0164 were used in negative ion mode. The Q-TOF was operated at a source condition of 3.75 kV with drying gas 9 L/min and nebulizer gas 45 psi at 300°C. The instrument was run at extended mass range to 1,700 *m/z*. The fragmentor voltage was 125 V and skimmer at 47 V. The data were acquired at a scan rate of 1.5 spectra/sec for MS. MS/MS was run at targeted mode at a scan rate of 3 spec/sec with 10 spec/sec for MS. Collision energy of 20 V, a fixed isolation window of 4 *m/z*, and retention time window of 0.25 min. Each sample was run at MS mode first at both positive and negative modes, and the differentially expressed metabolites were selected for MS/MS experiment.
Data processing	
Processing/normalization methods/software	The Molecular Feature Extractor in QA was used to search for features that have common elution profile and groups ions into one or more compounds containing *m/z* values that are related. For the chromatography alignment, only ions with intensity above 5,000 counts and retention time window within 0.2 min were selected. Ions not present in all files were filtered out. For samples from the same time point, the median value was used.
Sequence/annotation database	METLIN human metabolites
ID method/software	MassHunter Workstation software (Agilent Technologies), including Qualitative Analysis (QA v3.01) and Mass Profiler Professional (vB.02); Mass tolerance = 10 ppm
ID/expression measures	Spectra from profiling at each time point were obtained with three technical replicates and aligned for mass and retention time. The aligned spectra were filtered for a minimum of 2/3 time points being present for each identified mass. Data with CV < 0.4 were retained.
Data analysis method/ software	Clustering, pathway analysis, custom R, Mathematica, Python scripts.
I/O data file formats	Compound Exchange Format (CEF)
Additional Information	None

ID, identification.

Acknowledgments

Research reported in this publication was supported by Stanford University, the National Institutes of Health under the NIH training grant, NIH/NLM training grant T15-LM007033, NIH/NIGMS R24-GM61374; the Spanish Ministry of Science and Innovation Projects SAF2008-05384 and CSD2007-00017; European Union FP7 Projects 2007-A-201630 (GENICA) and 2007-A-200950 (TELOMARKER); European Research Council Advanced Grant GA232854; the Körber Foundation, the Fundación Marcelino Botín, and Fundación Lilly (España); NIH/NHLBI training grant T32 HL094274; NIH/NHLBI KO8 HL083914; NIH New Investigator DP2 Award D004613; and the Breetwor Family Foundation to M.S. G.M.'s research is supported by the National Human Genome Research Institute of the National Institutes of Health under award number K99HG007065, and previously T32HG000044. E.K.'s research is supported by the National Science Foundation (NSF) under the Division of Biological Infrastructure award 0969929, National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (NIH) under awards

U01DK089571 and U01DK072473, Seattle Children's Research Institute (SCRI), The Robert B. McMillen Foundation, EMC, Intel, and The Gordon and Betty Moore Foundation award to E.K. This support is very much appreciated. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Spanish Ministry of Science and Innovation Projects, the European UnionFP7 Projects, the European Research Council, the Körber Foundation, the Fundación Marcelino Botín, Fundación Lilly, or the Breetwor Family Foundation. The omics metadata checklists presented here are being published in parallel in *Big Data*, Volume 1, Number 4, due to its broad and community-wide importance.

Author Disclosure Statement

The authors declare that no conflicts of financial interest exist.

Address correspondence to:
Eugene Kolker
Bioinformatics and High-Throughput Analysis Laboratory
Seattle Children's Research Institute
Predictive Analytics
Seattle Children's Hospital
1900 Ninth Avenue
Seattle, WA 98101

E-mail: eugene.kolker@seattlechildrens.org

References

Chen R, Mias G, Li-Pook-Than J, et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

Kolker E, Özdemir V, Martens L, et al. (2014). Toward more transparent and reproducible omics studies through a common metadata checklist and data publications. OMICS 18(1) In this issue. Published in parallel: Big Data (2013) 1, 196–201.