OMICS A Journal of Integrative Biology
Volume 18, Number 10, 2014
© Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2014.0090

Letter To The Editor

Metadata Checklist: Identification of *CHI3L1* and *MASP2* as a Biomarker Pair for Liver Cancer Through Integrative Secretome and Transcriptome Analysis

Wenchao Ding,[1] Qingchong Qiu,[1] Guanfeng Liu,[1] Jie Liu,[1] Ruifang Mao,[1] and Biaoyang Lin[1–3]

To the Editor:

This letter to the editor is a dataset publication that provides the checklist (Table 1) for the metadata of the study entitled above (Wang et al., 2009). The checklist utilized here was recently developed and endorsed by the Data-Enabled Life Sciences Alliance (DELSA Global) (Kolker et al., 2014). The data are available at MOPED (the Model Organism Protein Expression Database, https://www.proteinspire.org/MOPED/) (Montague et al., 2014) with the digital ID: wang_liver_cancer (Table 1). We call for the broader use of data publications using the metadata checklist to make omics data more discoverable, interpretable, and reusable.

Availability of metadata is crucial to cultivate an innovation ecosystem where data generators will receive due credit while building effective, reproducible, and accountable linkages on the trajectory from data-to-knowledge-to-innovation in postgenomics medicine and integrative biology (Ozdemir et al., 2014).

Table 1. Multi-Omics Metadata Checklist

Checklist version *Experiment information*	*Version 1.0 (2014)* *Description*
Lab Name	Systems Biology Division, Zhejiang-California International Nanosystems Institute (ZCNI), Zhejiang University, Hangzhou, China
Date	06/12/2014
Author Information	Wenchao Ding,[1] Qingchong Qiu,[1] Guanfeng Liu,[1] Jie Liu,[1] Ruifang Mao,[1] Biaoyang Lin[1–3] [1]Systems Biology Division and Propriumbio Research Center, Zhejiang-California International Nanosystems Institute (ZCNI), Zhejiang University, 866 Yu Hang Tang Road, Zi Jin Gang Campus Zhejiang University, Hangzhou, Zhejiang, China 310058 [2]Swedish Medical Center, Seattle, WA 98122, USA. [3]Dept. of Urology, University of Washington, Seattle, WA 98195, USA.
Title of Experiment	Identification of *CHI3L1* and *MASP2* as a Biomarker Pair for Liver Cancer Through Integrative Secretome and Transcriptome Analysis
Project	wang_liver_cancer
Funding	This work was supported by grants from the National Infrastructure of Natural Resources for Science and Technology, Ministry of Science and Technology, China (2012AA022705, 2008DFA11320, 2006AA02Z4A2), China Postdoctoral Science Foundation (No.2013M531451), grant 2012R10021 from Zhejiang Province, China.
Digital ID	MOPED: wang_liver_cancer

(*continued*)

[1]Systems Biology Division and Propriumbio Research Center, Zhejiang-California International Nanosystems Institute (ZCNI), Zhejiang University, Zi Jin Gang Campus Zhejiang University, Hangzhou, Zhejiang, China.
[2]Swedish Medical Center, Seattle, Washington.
[3]Department of Urology, University of Washington, Seattle, Washington.

Table 1. (Continued)

Checklist version *Experiment information*	*Version 1.0 (2014)* *Description*
Abstract	Hepatocellular carcinoma (HCC) is the fifth most frequent neoplasm with more than 500,000 new cases diagnosed yearly. Novel liver cancer biomarkers are needed. By tandem mass spectrometry, we analyzed the secretomes of 12 individual paired samples of liver cancer and adjacent normal tissues and identified 1528 proteins with >2 unique peptide hits. The false discovery rate was 3.4%. Using spectral counting, we found 87 proteins in the HCC group and 86 proteins in the normal group that showed fivefold overexpression. These proteins provided a rich source of biomarker candidates. We presented a novel paradigm in combining biomarkers that include an up-regulated cancer biomarker and a down-regulated organ-enriched marker, and identified chitinase-3-like protein 1(*CHI3L1*) and mannan-binding lectin serine peptidase 2 (*MASP2*) as the top biomarker pair for HCC diagnosis using integrative transcriptomics and proteomics analysis. Using ELISA assays, we further evaluated this biomarker pair in a separate cohort of 25 serum samples of liver cancer patients and 15 age-matched normal controls. The combined marker pair (CHI3L1/MASP2 ratio) performed better than either marker alone with an AUC of 0.97 for liver cancer diagnosis. Further validation of the biomarker pair in HCC patients versus disease controls and independent cohorts is warranted.
Experimental design	*Description*
Organism	Human
OMICS Type(s) Utilized	Proteomics
Reference	Identification of *CHI3L1* and *MASP2* as a biomarker pair for liver cancer through integrative secretome and transcriptome analysis. [Proteomics Clin. Appl. 2009;3:541–551]
Experimental design	HCC and benign adjacent paired tissues (at least 2 cm away from the edge of HCC tissues) were collected from 12 HCC patients who underwent hepatectomy or liver transplantation at the First Affiliated Hospital, Zhejiang University with IRB approval. None of these patients received antineoplastic therapy prior to surgery. Additional serum samples were obtained from the serum bank at the First Affiliated Hospital, Zhejiang University.
Sample description	A total of 12 paired tissue samples (12 HCC samples, 12 samples benign tissue surrounding HCC) were used for analysis. Samples were subject to LC-MS/MS and Western Blot analysis. For confirmation of biomarker candidates, serum samples were subject to ELISA assays.
Tissue/Cell type ID	Hepactic cancer cell [BTO:0000608], Hepatocyte [BTO:0000575]
Localization ID	Cell [GO:0005623]
Condition ID	HCC, Benign
Experimental methods	*Description*
Sample prep fescription	The paired tissues were transferred to a Petri dish containing 20 mL of PBS and were finely minced into 2–3-cubic millimeter pieces using scissors. Thereafter the tissue pieces were re-suspended in 50 mL PBS and were poured over the stainless steel filter (200-mm diameter) to discard single cells and cell debris. The collected tissue pieces were washed three times with PBS and were re-suspended in 20 mL serum-free DMEM (Sigma, St Louis, MO) in a Petri dish. The tissue pieces were cultured at 37°C in a cell culture incubator (Thermo Scientific, Milford, MA) with 5% CO_2. Supernatants from tissue culture were collected at 24 h after tissue culture. The supernatants were centrifuged at 2000 X g for 10 min to remove any cells or cell debris that might be contained in the supernatants. The samples were concentrated about 20-fold by a Speedvac (Labconco Centrivap Concentrator, Kansas City, MO) and were re-suspended in 25 mM ammonium bicarbonate (NH_4HCO_3, Sigma, St Louis, MO). Of each sample, 60 mg was separated on 12% SDS polyacrylamide gels. Gels were stained with Colloidal CBB. Proteins in the gel were digested with trypsin using the Pierce In-Gel Tryptic Digestion Kit protocol (Pierce Biotechnology, Rockford, IL).
Platform type	LC-MS/MS
Instrument name	LTQ-Orbitrap
Instrument details	LTQ-Orbitrap software: BioWorks 3.2, Xcalibur

(*continued*)

TABLE 1. (CONTINUED)

Experimental methods	*Description*
Instrument protocol	Tryptic peptide mixture was separated by the Ettan MDLC nanoflow/capillary LC system (GE Healthcare, Pittsburgh, PA) equipped with a trapping column (Dionex/LC Packings m-Precolumn Cartridge P/N 160454 C18 PepMap 100, 5 mm, 100 Å, 300-mm id x 5 mm, Sunnyvale, CA) and a nanocolumn (Dionex/LC Packings P/N 160321 15060.075-mm id, C18 PepMap, 3 mm, 100 Å), and then analyzed using LTQ Orbitrap (Thermo Finnigan, Bremen, Germany) with a nanospray configuration. The precursor ion scan MS spectra (*m/z* 300–1600) were acquired in the orbitrap with the resolution R = 60 000 at *m/z* 400 with the number of accumulated ions being 16106. The five most intense ions were isolated and fragmented in linear IT (number of accumulated ions: 36104). The resulting fragment ions were recorded with the resolution R = 15 000 at *m/z* 400.

Data processing	*Description*
Processing/normalization methods/software	Software: BioWorks 3.2, Xcalibur The extract_msn of the BioWorks program V3.2 (Thermo Electron, Waltham, MA) was used to generate the MS peak list with the default parameters. The ICIS peak-detection algorithm peaks of the Xcalibur (Thermo Electron) was used for automated detection of mass spectrum.
Sequence/annotation database	ipi.-HUMAN.v3.29.fasta protein database
ID method/software	SEQUEST, custom perl script, ProteinProphet
ID/expression measures	In the TurboSEQUEST search parameter setting, the threshold for Dta generation was 10 000, and precursor mass tolerance for Dta generation was 1.4. For the SEQUEST search, peptide tolerance was set at 3 Da and fragment ions tolerance was set at 0.01 Da. PeptideProphet™ was used to assess the MS/MS spectra quality and a threshold score for accepting individual MS/MS spectra was set at p value of 0.9, which corresponds to a 0.5% error rate in our dataset. One missed tryptic cleavage was permitted. Carboxyamidomethyl cysteine (Cys_CAM) was included as a fixed modification for iodoacetamide reduction and alkylation. As the proteins were prepared by PAGE, the cysteines might react with free acrylamide monomers to form propionamide cysteine (Cys_PAM). We included an optional 14 Da in the search to account for potential propionamide cysteine (the mass difference between Cys-PAM and Cys-CAM is 14). Methionine oxidation (116 Da) was chosen as another optional modification for the database search. Proteins with ProteinProphet p value greater than 0.9 and with more than two unique peptide hits were considered as true hits. A randomized database of the ipi.HUMAN. v3.29.fasta was used as a decoy database to calculate the false discovery rate of protein identification. The perl script used for randomization was from www.matrixscience.com/ downloads/decoy.pl.gz. The false discovery rate (FDR) was calculated by the ratio of the number of matches to the randomized database to that to the ipi.HUMAN.v3.29.fasta database.
Data analysis method/ software	We summed up the total spectrum numbers in the HCC group (12 samples) and the control group (12 samples). The spectrum numbers were normalized to the total number of spectra of all proteins identified. Ratios of spectrum numbers between the HCC and the control group were calculated. GoMiner was used to find statistically represented Gene Ontology (GO) categories. The 1528 proteins with more than two hits were used as the total input and the differentially expressed genes were analyzed using evidence level 3. *Western blot analysis* Proteins from the HCC and the uninvolved surrounding tissues were separated on 12% polyacrylamide gels and transferred to PVDFmembranes (Amersham Pharmacia. Biotech, Uppsala, Sweden). These blots were incubated for 2 h at room temperature in the TBST buffer (20 mM Tris-HCl, 140 mM NaCl, pH 7.5, 0.05% Tween-20) containing 5% skim milk, then incubated with the primary antibody anti-AAT (Alpha-1-antitrypsin, IPI00553177) (Santa Cruz Biotechnology, CA) overnight at 47°C. After washing three times in TBST, blots were incubated with HRP-conjugated secondary antibody (diluted 1:10 000, Santa Cruz Biotechnology) for 1 h at room temperature. ECL reagents were used for visualization (Pierce Biotechnology). *ELISA assay* The ELISA kit for CHI3L1 (YKL40) and MASP2 were purchased from Quidel Corp. (San Diego, CA) and Hycult biotechnology bv (Uden, The Netherlands) and ELISA was performed according to the manufacturer's instruction. Serum samples were diluted three times with PBS buffer before analysis. ROC (receiver operating characteristic) curve analysis was performed using GB STAT V10.0 (Dynamic Microsystems, Silver Spring, MD).
I/O data file formats	
Reserved/dxtension	

*ID, Identification.

References

Kolker E, Ozdemir V, Martens L, et al. (2014). Toward more transparent and reproducible omics studies through a common metadata checklist and data publications. OMICS 18, 10–14.

Montague E, Stanberry L, Higdon R, et al. (2014). MOPED 2.5—An integrated multi-omics resource: Multi-omics profiling expression database now includes transcriptomics data. OMICS 18, 335–343.

Ozdemir V, Kolker E, Hotez PJ, et al. (2014). Ready to put metadata on the post-2015 development agenda? Linking data publications to responsible innovation and science diplomacy. OMICS 18, 1–9.

Wang J, Gao F, Mo F, et al. (2009). Identification of CHI3L1 and MASP2 as a biomarker pair for liver cancer through integrative secretome and transcriptome analysis. Proteomics Clin Appl 3, 541–551.

Address correspondence to:
Biaoyang Lin, PhD
Zhejiang-California International
Nanosystems Institute (ZCNI)
Zhejiang University
866 Yu Hang Tang Road
Zi Jin Gang Campus Zhejiang University
Hangzhou
Zhejiang 310058
China

E-mail: Biaoylin@gmail.com