

“All and None” Refining Strategy; Fishing Your Correct Protein from Proteomics Ocean

Sameh Magdeldin^{1,2*}, Yutaka Yoshida¹, Ying Zhang¹, Bo Xu¹, Eishin Yaoita¹ and Tadashi Yamamoto¹

¹Department of Structural Pathology, Institute of Nephrology, Graduate School of Medical and Dental Sciences, Niigata University, Japan

²Department of Physiology, Faculty of Veterinary Medicine, Suez Canal University, Ismailia, Egypt

Abstract

We have developed an easy-to-use methodology for refining large extensible markup language (XML) - based proteomics dataset with a high stringent and simple approach using VBA- coded plug-in. A methodology we term it (All and None). Selections of targeted candidates differentially significant between compared groups were selected based on its appearance or absence followed by peptide screening with a novel and simple approach. By testing the reliability and efficiency of this method, All and None was confirmed to be an applicable process for initial screening of biological biomarkers in complex specimens and tissue extract.

Keywords: Refining dataset; Proteomics; VBA code

How many times we became confused when looking for the correct significant protein hits within those dozens of proteins list? Did you try “All and None”?

With the current revolution of proteomics machinery, high performance and sensitivity of mass spectrometry, it became usual trend to obtain results with increasing number of protein hits day after a day. On the same time, “junk” proteins are also added in parallel. While one of the most important challenges has been, and still remains, the difficulty and doubting task of finding out the correct protein hit of interest within these huge data outputs, the demand of high stringent filtering process might be necessary “sometimes” in order to rapid screen the highly significant proteins we might be interested in. Aside from the quantitative labeling methods that aim to find out differentially significant candidates (i.e. ICAT [1], iTRAQ[2], and SILAC [3]) or other label-free methods that use certain algorithms depending on peptide counts in relation to protein length or molecular weight (i.e. spectral count [4], emPAI [5], PAF [6], and APEX [7]) for the same purpose, we created a simple method for initial refining large scale proteomic data and we term it “All and None” [8]. This refining strategy relays mainly on selecting protein candidates with high confidence under stringent condition using a visual basic macro plug-in followed by manual peptide filtering.

Principle of “All and None” comparative strategy “How it works?”

Simply, as shown in Figure 1, by initial searching for the candidates shared in a replicate runs of one experimental group [All] and, at the same time, absent elsewhere in the counterpart group [None]. Assuming unbiased mass analysis, selection of shared protein candidates in one group will ascertain the consistency and relative high abundance (when compared to counterpart group) of a given protein. Meanwhile, the absence of the same candidate in the counterpart group (not necessary shared but even in a single run as well) indicates its low abundance or lack of translation. Practically, search can be done in a simple and robust way by activating our pre-prepared visual basic code (VBA) within excel spreadsheet (supplement 1), pasting VBA code and start comparison. Any desired database identifier could be used for comparison (IPI, Uniprot, GeneID, etc.). In the next step, significantly abundant proteins (based on its visualization in a run) will be searched for its quality based on peptide scoring and Ms/Ms spectra which is expressed finally as protein score. To this point, the researcher can set his desired criteria for accepting or rejecting those

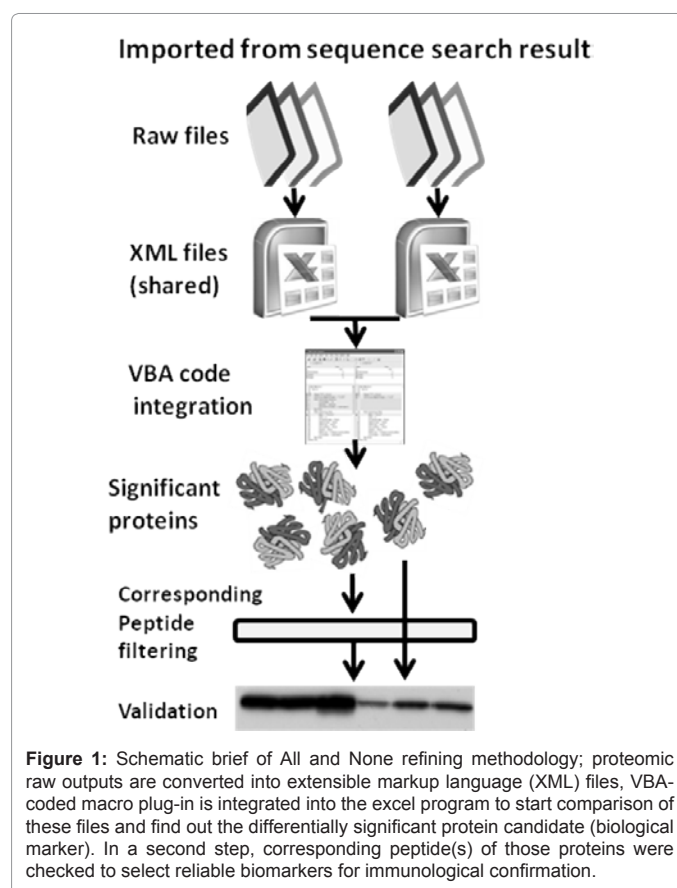


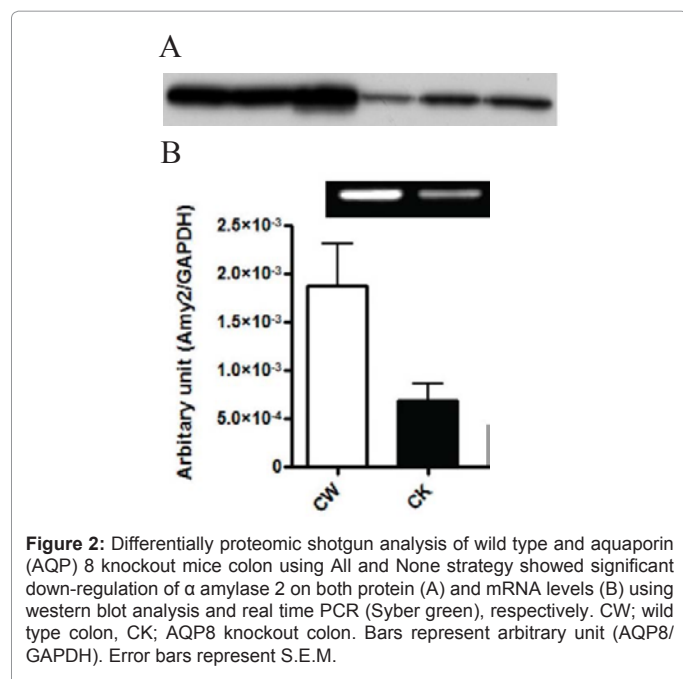
Figure 1: Schematic brief of All and None refining methodology; proteomic raw outputs are converted into extensible markup language (XML) files, VBA-coded macro plug-in is integrated into the excel program to start comparison of these files and find out the differentially significant protein candidate (biological marker). In a second step, corresponding peptide(s) of those proteins were checked to select reliable biomarkers for immunological confirmation.

*Corresponding author: Sameh Magdeldin, Department of Structural Pathology, Institute of Nephrology, Graduate School of Medical and Dental Sciences, Niigata University, Japan, Tel. +81 25 227 2152; Fax: +81 25 227 0768; E-mail: samehmagd@med.niigata-u.ac.jp, samehmagd@yahoo.com

Received May 09, 2011; Accepted June 17, 2011; Published June 25, 2011

Citation: Magdeldin S, Yoshida Y, Zhang Y, Xu B, Yaoita E, et al. (2011) “All and None” Refining Strategy; Fishing Your Correct Protein from Proteomics Ocean. J Proteomics Bioinform 4: 123-124. doi:10.4172/jpb.1000178

Copyright: © 2011 Magdeldin S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



proteins. In our experience, we confirmed the high confidence, high quality of a given protein when its corresponding peptides mostly over 2 and its scores were above identity and homology threshold of MOWSE algorithm [8,9] used in Mascot search [10]. High through put and accurate orbitrap mass spectrometry allows nowadays accepting protein with single peptide identification. These above mentioned two filters assure selecting highly significant and confidence protein candidates. In a previous experiment to validate efficiency and accuracy of this method, a comprehensive comparison between both wild type and aquaporin 8 knockout mice proteomes using All and None showed significantly down regulation of α amylase 2 on both transcriptional and translational levels [11].

Pros and cons

The usefulness of "All and None" search strategy lies in its easy-to-use methodology. Added that, to our knowledge, high stringency used in this approach is usually accompanied by successful confirmation when validated immunologically (Figure 2). Flexibility in setting up criteria for accepting a significant protein based on its peptides is also applicable according to researcher's vision. Finally, rapid selection of top differentially significant protein which is more liable to be a biomarker candidate is the most important feature in this approach. While "All and None" provides a shortcut way to find out differentially targeted proteins-based biomarker, one should be aware on the drawbacks of using such approach. For instance, the simplicity of the approach makes it basically a tool for initial screening and away from a comprehensive speculation. Moreover, other proteins which can be detected in both experimental groups with less significance will be ignored. Indeed, those might be important and still significant. Moreover, researchers should note that successful picking of possibly significant biomarker using All and None depend greatly on the quality of mass analysis result [as All and None starts where Mass spectrometry ends] Finally, it seems to us noteworthy to mention that "All and None" from its name do not provide any quantitative information about over/under regulated proteins; instead, it scans for possible biomarkers as initial step.

Conclusion

Applying "All and None" searching strategy enables us to select differentially significant proteins that most possibly found in one group with multiple folds difference when compared to the counterpart group. In the later, same candidate is not shown in the search due to its absence or very low abundance. In a next step, quality of those protein candidates is checked through their peptide scoring and Ms/Ms spectra to assure high confidence. This approach can be helpful in fishing candidate of interest from Proteomic Ocean in an easy way.

Financial & competing interest disclosure

This work was supported by JSPS (Japan Society for Promotion of Science) Grant-in-Aid for scientific research (B) to Dr. SM (23790933) from Ministry of Education, Culture, Sports, Science and Technology of Japan. "All and None" is an approach created and published by S.M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the work.

References

1. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, et al. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17: 994-999.
2. Zieske LR (2006) A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* 57: 1501-1508.
3. Amanchy R, Kalume DE, Iwahori A, Zhong J, Pandey A. (2005) Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J Proteome Res* 4: 1661-1671.
4. Zhao Y, Denner L, Haidacher SJ, LeJeune WS, Tilton RG (2008) Comprehensive analysis of the mouse renal cortex using two-dimensional HPLC - tandem mass spectrometry. *Proteome Sci* 6: 15.
5. Shinoda K, Tomita M, Ishihama Y (2010) emPAI Calc--for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* 26: 576-577.
6. Powell DW, Weaver CM, Jennings JL, McAfee KJ, He Y, et al. (2004) Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Mol Cell Biol* 24: 7249-7259.
7. Braisted JC, Kuntumalla S, Vogel C, Marcotte EM, Rodrigues AR, et al. (2008) The APEX Quantitative Proteomics Tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* 9: 529.
8. TM4 Software Suite's MultiExperiment Viewer.
9. Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 3: 327-332.
10. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551-3567.
11. Magdeldin S, Li H, Yoshida Y, Satokata I, Maeda Y, et al. (2010) Differential proteomic shotgun analysis elucidates involvement of water channel aquaporin 8 in presence of alpha-amylase in the colon. *J Proteome Res* 9: 6635-6646.