Journal of proteome research

Letter

pubs.acs.org/jpr

Analyzing the First Drafts of the Human Proteome

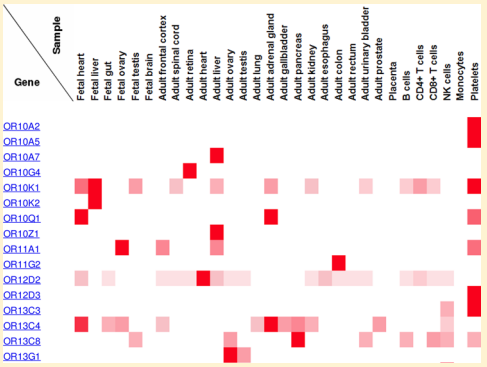
Iakes Ezkurdia,[†] Jesús Vázquez,[§] Alfonso Valencia,[‡] and Michael Tress*[,‡]

[†]Unidad de Proteómica, [§]Laboratorio de Proteómica Cardiovascular, Centro Nacional de Investigaciones Cardiovasculares, Melchor Fernández Almagro, 3, Madrid 28029, Spain

[‡]Spanish National Cancer Research Centre (CNIO), Melchor Fernandez Almagro, 3, Madrid 28029, Spain

Supporting Information

ABSTRACT: This letter analyzes two large-scale proteomics studies published in the same issue of *Nature*. At the time of the release, both studies were portrayed as draft maps of the human proteome and great advances in the field. As with the initial publication of the human genome, these papers have broad appeal and will no doubt lead to a great deal of further analysis by the scientific community. However, we were intrigued by the number of protein-coding genes detected by the two studies, numbers that far exceeded what has been reported for the multinational Human Proteome Project effort. We carried out a simple quality test on the data using the olfactory receptor family. A high-quality proteomics experiment that does not specifically analyze nasal tissues should not expect to detect many peptides for olfactory receptors. Neither of the studies carried out experiments on nasal tissues, yet we found peptide evidence for more than 100 olfactory receptors in the two studies. These results suggest that the two studies are substantially overestimating the number of protein coding genes they identify. We conclude that the experimental data from these two studies should be used with caution.



KEYWORDS: *proteomics, Nature, human proteome, protein coding genes, olfactory receptors*

We read with great interest the recent cover of *Nature* (The Human Proteome). The issue contains two large-scale proteomics analyses based around publicly available databases, ProteomicsDB and Human Proteome Map. Bernhard Kuster and coworkers[1] describe ProteomicsDB as a "mass-spectrometry-based draft of the human proteome", while the Human Proteome Map, developed by Akhilesh Pandey and colleagues,[2] offers a "draft map of the human proteome". The studies have been portrayed as a great advance in the field. As with the initial publication of the human genome, the papers are of broad appeal and will no doubt lead to a great deal of further analysis by the scientific community.

We were particularly intrigued by the number of genes detected by the two studies, numbers that far exceed what has been reported for the multinational Human Proteome Project effort.[3] These numbers were reached in part by combining spectra from multiple experiments. Although combining spectra from multiple experiments may increase coverage, the advantage of using very large data sets has been shown to come at the expense of higher false-positive protein rates.[4] Given this, we were concerned about the quality of the peptide identifications in these two studies. Data quality is especially important in large-scale proteomics experiments because researchers cannot carry out individual follow-up studies on peptides identified on a genome-wide scale.

We decided to carry out a simple quality test on the data using the olfactory receptor family. Olfactory receptors are seven transmembrane helix receptors that trigger the olfactory signal transduction pathway. These receptors first appeared in vertebrates and have duplicated to such an extent that mammalian species possess many hundreds of these genes. From the point of view of proteomics analysis, this family is highly interesting. The functional specificity of these genes indicates that expression is predominantly limited to a single tissue, although the mouse orthologue of *OR51E2* has been convincingly shown to have a function in the kidney,[5] and the Human Protein Atlas records limited RNA evidence of the expression of olfactory receptors outside of the nose (primarily in testes[6]). Olfactory receptors have very little transcript expression and should be particularly difficult to detect in proteomics experiments because they are transmembrane proteins.

A high-quality proteomics experiment that does not include a specific analysis of nasal tissues should not expect to detect much evidence of peptide expression for these genes. For example, PeptideAtlas,[7] known for having high stringency criteria, identifies just two discriminating olfactory receptor peptides. As far as we know, neither of the studies carried out experiments on nasal tissues. We found peptide evidence of 108 of these olfactory receptors in the Human Proteome Map database, and another 200 olfactory receptors are recorded in ProteomicsDB.

There are at least three reasons for the high numbers of olfactory receptors in the two studies. First, neither experiment

Received: June 10, 2014
Published: July 11, 2014

   dx.doi.org/10.1021/pr500572z | *J. Proteome Res.* 2014, 13, 3854−3855

properly distinguishes between discriminating and nondiscriminating peptides, so olfactory receptors are identified by peptides that map to more than one gene. (40 of the olfactory receptors detected in the Pandey study were identified solely by nondiscriminatory peptides.) Second, a number of peptides were wrongly identified as having a glutamine to pyroglutamic acid modification in non N-terminal positions. Third, both studies include very many low-quality spectra (Supporting Information). Most of the peptides that map to the remaining 68 olfactory receptors in the Pandey study were identified using poor spectra, and we were unable to find even one peptide that could provide unequivocal evidence of the presence of the protein. A similar in-depth study was not possible with the Kuster data, but we did look at the spectra for many olfactory receptors and found the same pattern. For example, the olfactory receptor with the most evidence in the Kuster study was *OR6J1* with eight peptides. Despite what should be overwhelming evidence, the spectral evidence of the existence of each one of these peptides was inconclusive.

The results of our analysis show that both studies are substantially overestimating the number of protein coding and noncoding genes they find. We suggest that the experimental data from these two should be used with great caution, and we feel that these two unique draft maps of the human proteome should be put on hold until they can be carefully analyzed.

■ ASSOCIATED CONTENT

Supporting Information

Unique, prototypic spectra from the datasets that were highlighted in the letter. Found spectra fall into the categories of ambiguous short peptides, peptides with a high ratio of missed cleavages, and very low-quality spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mtress@cnio.es.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Wilhelm, M.; Schlegl, J.; Hahne, H.; Moghaddas Gholami, A.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-Spectrometry-Based Draft of the Human Proteome. *Nature* **2014**, *509*, 582−587.

(2) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; et al. A Draft Map of the Human Proteome. *Nature* **2014**, *509*, 575−581.

(3) Lane, L.; Bairoch, A.; Beavis, R. C.; Deutsch, E. W.; Gaudet, P.; Lundberg, E.; Omenn, G. S. Metrics for the Human Proteome Project 2013−2014 and Strategies for Finding Missing Proteins. *J. Proteome Res.* **2014**, *13*, 15−20.

(4) Reiter, L.; Claassen, M.; Schrimpf, S. P.; Jovanovic, M.; Schmidt, A.; Buhmann, J. M.; Hengartner, M. O.; Aebersold, R. Protein Identification False Discovery Rates for Very Large Proteomics Data Sets Generated by Tandem Mass Spectrometry. *Mol. Cell. Proteomics.* **2009**, *8*, 2405−2417.

(5) Pluznick, J. L.; Protzko, R. J.; Gevorgyan, H.; Peterlin, Z.; Sipos, A.; Han, J.; Brunet, I.; Wan, L. X.; Rey, F.; Wang, T.; et al. Olfactory Receptor Responding to Gut Microbiota-Derived Signals Plays a Role in Renin Secretion and Blood Pressure Regulation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 4410−4415.

(6) Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; et al. Towards a Knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28*, 1248−1250.

(7) Aebersold, R.; Desiere, F.; Deutsch, E.; King, N.; Nesvizhskii, A.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655−D658.