

Opportunities and Challenges in Omics

MingMing Ning · Eng H. Lo

Received: 21 October 2010 / Accepted: 21 October 2010 / Published online: 12 November 2010
© Springer Science+Business Media, LLC 2010

Keywords Proteomics · Proteome · Genome · Reproducibility · Noise · Continued convergence · Discovery proteomics · Targeted proteomics · Multiple reaction monitoring · Translational stroke research · Neurovascular research

Omics and Translational Stroke Research

Recent advances in “omics” technology have provided a powerful set of tools and concepts that allow us to dissect the entire phenotypic and functional network of genes and proteins present in a cell or organism. Decades of technological development has finally allowed biology and technology to meet each other half way as “omics” are being utilized to probe systems in a wide spectrum of biology and medicine.

This special issue of TSR gives examples of how genomics and proteomics are used to study mechanisms and disease states—demonstrating advances and challenges of “omics” in translational neurovascular disease. Specifically, the authors present both bench and bedside examples of mechanism- and disease-driven inquiry. Studies probe cell–cell signaling of extracellular secretomes [1], intracellular mechanisms [2], and central vs. peripheral responses after

stroke [3]. Efforts are made to understand human disease-specific states by finding biomarkers for the diagnosis and therapeutic efficacy of hemorrhagic [4] and ischemic stroke [5–8]. Taken together, this issue showcases omics inquiries utilizing target-driven [4, 6], discovery-driven [1, 2, 5, 7], and combination approaches [8]. Dr. Sharp's accompanying commentary provides an insightful and detailed summary of these articles and their interpretations [9].

Opportunities and Challenges

Over the last decade, “omics” technology has advanced from cataloging lists of genes, proteins, and SNPs to disease-specific in-depth analyses of meta-genetics, protein–protein interactions, modifications, and pathway mapping. Large-scale genome-wide association studies and high-throughput techniques have become more efficient and productive in exploring vastly uncharted territory. “Discovery” approaches to find unknown factors of a particular disease state are made possible by multiplex technology such as rapid sequencing and better mass spectrometry instrumentation with better resolution to detect complex protein mixtures. “Targeted” approaches utilizing mass spectrometry multiple reaction monitoring technology afford rapid quantitation of multiple peptides without antibodies. [10]. Despite its complexity, “omics” has brought us one step closer to the ultimate clinical phenotype in complex diseases such as stroke that likely involve multiple gene, multiple organ, and environmental influences. Translational stroke research should greatly benefit from these approaches, as “omics” will eventually map entire cellular–molecular pathways to guide in vivo studies and allow bedside exploration to be validated at the bench. Indeed, exciting

M. Ning (✉) · E. H. Lo (✉)

Clinical Proteomics Research Center and Neuroprotection Research Laboratory, Departments of Neurology and Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

e-mail: Ning@hms.harvard.edu

e-mail: Lo@helix.mgh.harvard.edu

work from Sharp and colleagues has shown that genomic signatures in peripheral blood may help detect and identify stroke etiologies [11].

However, the promise offered by the still-new fields of “omics” has often been met with some skepticism from the larger scientific community—including ourselves. In part, this may be because of the constant flux in a rapidly advancing field, because its potential has sometimes been over-hyped for financial interest [12, 13], and because some of its premises are fundamentally different in important ways from more traditional approaches. Such skepticism is healthy; it should be taken seriously and addressed directly for the field to move forward.

There are three principal criticisms commonly leveled against “omic” studies:

- **Reproducibility:** why cannot one group's findings be reproduced in another individual or in another cohort [14–16]?
- **Noise:** “omics” can generate so much data that noise overwhelms signal. Why is it that we find fewer clinically relevant biomarkers as our instruments and methodologies advance [17]?
- **Fishing:** the search for novel markers in complex samples, i.e., discovery-mode “omics”, has been frequently criticized as a “fishing expedition” in which one blindly hopes to find something interesting and yet often falls short.

These issues are closely inter-related. The problem of reproducibility follows from the problem of noise, and the existence of noisy data sets that cannot be reproduced leads to the suspicion that “omic” investigators are “just fishing.” Perhaps, in part, the answer in each case lies in the careful design of “omic” experiments for a focused biologically relevant question and the appropriate interpretation and validation of their results. In particular, translational stroke research can uniquely benefit from omic methodology, utilizing cellular screen to guide in vivo studies to minimize noise, and conversely, conducting bedside explorations to provide candidates for further validation at the bench to focus and insure reproducibility.

Dr Sharp's editorial focused on genomics. Here, we will look at proteomics as an example to address the three issues above [14–19]. While genomics and proteomics share much similarity and contemporary proteomics has become fruitful as a result of the genomics revolution, there exist intrinsic differences that will require unique solutions. In particular, the proteome is inherently more diverse and less stable than the genome—the source of both the promise and challenge.

New Perspectives?

Reproducibility

The more precisely the position is determined, the less precisely the momentum is known in this instant, and vice versa.

–Heisenberg

The heart of the issue for “omics” methodology has been reproducibility—healthy skepticism of the reproducibility issue has advanced genomic exploration to derive the training set from well-characterized cohorts and to validate in separate cohorts [20]. The proteome is somewhat different, but the simplest solution of utilizing well-characterized samples with standard operating procedures enhances reproducibility greatly.

Nevertheless, unlike a relatively stable genome, the proteins present in a given system will necessarily differ not only from one individual to the next, but also from one moment to the next. Proteomic studies are extraordinarily sensitive to such changes as well as to differences in sample preparation protocols and instrumentation. As confounders, such variables can quickly overwhelm experimental data with extrinsic noise. But, in the context of well-designed and well-controlled experiments, the same sensitivity to individual and temporal variation can augment the wealth of biologically useful information. For example, the study of a well-phenotyped individual organism over time may offer more reliable hypothesis-generating targets than pooled heterogeneous samples of large sizes.

Accordingly, it can be equally argued that sensitive measurement of rapidly changing biological variability potentially leaves no room for a traditional sense of “reproducibility.” Is it actually possible to reproduce/recapture both quantity and temporal profile from experiment to experiment or from one individual to another? Heisenberg's principle of uncertainty states that it is impossible to determine simultaneously both position and momentum of a particle—the foundation of quantum mechanics' “wave-particle duality” to address the inadequacy of classical debate of the one-dimensional theories of particle versus wave. In acknowledging the “uncertainty” and complexity of matter, the theory laid foundations for quantum mechanics, calling for alternative ways to describe matter. Perhaps, an analogous principle should be applied to proteomics—where the unique biological signature must also be described in relation to both space and time? That is, in a biological entity, the precise quantity of any protein is expressed as a function of many “phenotypic dimensions” over time. In order to be fully

descriptive, proteomics should not be simply “reproducible” in one dimension, but proteomic patterns will have to be “continuously convergent.”

Innovative bioinformatics and mathematical tools can help us to store, visualize, sort, and reconstruct, such that similar conclusions can be drawn from multiple datasets. But, this may not mean the exact same proteins are elevated. Instead, global pathways or interactions may be preserved through phenotypically different ways. Individual proteomic profiles may never be identical (“reproducible”), but by making the right interpretive comparisons in experiments constructed to focus on relevant changes within each single subject, they may consistently tell us the same story and that is what matters.

By analogy, each organism is still the same or its own “replicate” as it ages. Its biological and physical features may change drastically over time as any biological entity cannot stay static or “noise-free.” But, it is still the same organism! So, perhaps a new set of criteria needs to be established to measure this “continuous convergence” of a distinct biological signature with respect to time and intrinsic biological “noise.”

Noise

*How do you make sense of your life? Signal to noise:
What's signal? What's noise?
—Neil Gaiman*

Accordingly, it may be necessary to recalibrate our understanding of what reproducibility means in the context of noise. And all biological systems have noise. For example, early protein chemistry of what is now termed “targeted” or “candidate” approaches that focus on a single or a few biomarkers have been successful. Examples of useful clinical markers in cancer (e.g., PSA for prostate cancer diagnosis), infectious disease (CD4 count to follow progression of HIV), or cardiovascular disease (CRP in cardiac risk stratification) all pre-date some of the “omics” explosion and are strongly rooted in bench research. A major introspective question within the omics/biomarker field is, “why is it that over the last decade, we are equipped with better instrumentation and omics technology, and yet we have fewer discoveries of clinically relevant biomarkers” [17]? And, even if there are new findings, they seem to be nonspecific inflammatory “noise,” and they become irreproducible and non-predictive in another individual or another cohort. Of course, many potential reasons have been discussed, such as overfitting of data and various and different sample/instrumentation quality [18, 19].

However, it is unreasonable to expect the same mass spectrometry spectra to be generated by different investigators working with different protocols on different machines, using pooled samples taken at different times from different cohorts of multiple individuals. Though such methods were important in early efforts to chart the human proteome—understood as an abstract entity analogous to the genome—they may not be appropriate for translational research directed at clinically relevant variations in individual patients.

Pragmatically and methodologically, if omics technology can afford us the sensitivity of studying a multitude of markers at the same time, can we leverage this power to study a smaller number of individuals over time and space, across specific pathways, interactions, phenotypes, or disease states, utilizing each individual as their own control, thereby reducing the caveats of complexity/dimensionality/confounders? While we carry our own blood type for future match, can we each have our own non-diseased proteome baseline, such that, when we are ill, a novel signal can be detected more readily? Is the “noise” also worthy of investigating and incorporating into the new perspective? Not just philosophically, but pragmatically, perhaps, we need another set of criteria for reproducibility in the context of the noise of life? Once again, expecting everything to be the same or reproducible at all times, just because they share similar disease states may not reflect our ever-changing homeostasis. And, proteomic investigations may require a search for a continued convergence of patterns.

Fishing

*Cast the net on the right side of the boat, and ye shall
find.
—John 21:6*

Proteomics is not blind fishing, nor is it entirely new, since it is built on decades of careful work in the study of proteins, which even pre-dates the discovery of DNA. It is a method, enabled by advancing technology in mass spectrometry and protein separation chemistry, to be able to study expression and interaction in real time. Though a powerful methodology, it cannot take the place of a well-designed experiment, years of well-studied targets, or thoughtful hypothesis of focused biological questions at hand. The science remains the same, but omics technology can now offer a more multi-dimensional probing of complex organisms. While it offers more details—as when perspective was first being introduced into painting to give a more realistic 3D depiction of the world—the subjects of rigorous scientific inquiry remain the same.

One important innovation from “discovery” proteomics is its ability to yield multiple novel candidates for further study. To simply label this as “fishing” is to turn our back on a great potential benefit. Of course, we should be on guard as always against poorly designed experiments and unfocused hypotheses. But, real fishermen possess a great deal of detailed skill and knowledge about fish, the water in which they live, and how/when/where to catch the fish they want. Similarly, the type of protein–protein interactions, pathways, and candidates should fit the context of the specific and focused hypothesis.

For example, well-planned discovery proteomics should aim to capture the augmented signal by asking questions with sufficiently large signals such as pre- and post-therapeutic challenge or cellular stimuli to gauge therapeutic efficacy/response. And, perhaps a combination of targeted and discovery approaches would yield more fruitful results that are specific to the mechanism under investigation. Sometimes, starting with discovery proteomics in a less complex mixture such as cell culture to guide the more complex *in vivo* exploration with targeted methodology may yield helpful results. And *vice versa*, sometimes starting with clinical exploration to take back to the animal model for validation can help to confirm biological relevance. Incorporating a longitudinal temporal component and crossover design will also minimize noise and maximize signal. Therefore, a well-designed proteomic study has the potential to answer specific scientific questions while at the same time discovering new candidates for further investigation, greatly accelerating the pace of new hypothesis generation to compliment traditional methods. If one never goes fishing, one cannot catch fish. So, happy (wise, well-planned, and productive) fishing!

Omics, Continued Convergence, and Beyond

Translational research in stroke holds enormous potential because this is a disease that manifests not only with multiple cell–cell interactions within the brain itself, but also in the context of multi-organ interaction. This is where omics technology can offer invaluable power and perspective to study complexity.

Perhaps, to be thought-provoking, do we need to look for a broader definition of reproducibility in the context of biological noise and complexity? Since intrinsic biological changes are detected by omics methodology, the interpretation requires innovative and more complex mathematical, informatics reconstruction to incorporate new dimensionality in analyzing data. While genomics share much similarity with proteomics, there are fundamental differences. Perhaps, we need to use other ways to judge “reproducibility” in proteomics data such as a *continued convergence* of signals and patterns to take into context the dynamic

equilibrium of homeostasis over time and space? Moreover, can genome and proteome intersect and give more robust information? Do omics only include the genome and the proteome? What about other omics—metabolomic profiles—the living products of the genome/proteome? Omics allows us to experience the emerging and vastly different perspectives of an ever-changing biological entity. Our current challenge to understand and streamline multi-dimensional data and to view the complete landscape of a biological entity can be equally rewarded with the knowledge of these added new perspectives.

Acknowledgments Supported in part by grants from the US National Institutes of Health/NINDS. We also thank Dr David McMullin for his thoughtful contribution in the discussion and preparation of this manuscript.

References

- Liu J, Gao BB, Feener E. Proteomic identification of novel plasma kallikrein substrates in the astrocyte secretome. *Transl Stroke Res.* 2010;24. doi:10.1007/s12975-010-0039-z.
- Lim K-Y, Chua J-H, Tan J-R, Swaminathan P, Sepramaniam S, Armugam A, Wong P T-H, Jeyaseelan K. MicroRNAs in cerebral ischemia. *Transl Stroke Res.* 2010;13. doi:10.1007/s12975-010-0035-3.
- Vartanian KB, Stenzel-Poore MP. Toll-like receptor tolerance as a mechanism for neuroprotection *Transl Stroke Res.* 2010. doi:10.1007/s12975-010-0033-5.
- Markus H. Genetic studies in ischaemic stroke *Transl Stroke Res.* 2010;29. doi:10.1007/s12975-010-0038-0.
- Montaner J, Salat D, Garcia-Berrocso T, Molina CA, Chacon P, Ribo M, Alvarez-Sabin J, Rosell A. Reperfusion therapy for acute stroke improves outcome by decreasing neuroinflammation. *Transl Stroke Res.* 2010;12. doi:10.1007/s12975-010-0038-0.
- Brunkhorst R, Pfeilschifter W, Foerch C. Astroglial proteins as diagnostic markers of acute intracerebral hemorrhage—pathophysiological background and clinical findings. *Transl Stroke Res.* 2010;25. doi:10.1007/s12975-010-0040-6.
- Sideso E, Papadakis M, Wright C, Handa A, Buchan A, Kessler B, Kennedy J. Assessing the quality and reproducibility of a proteomic platform for clinical stroke biomarker discovery. *Transl Stroke Res.* 2010;18. doi: 10.1007/s12975-010-0036-2.
- MingMing Ning and Eng H. Lo, MGH-Harvard (TRSR 34). Opportunities and challenges in Omics. *Transl Stroke Res.* 2010. doi:10.1007/s12975-010-0047-z.
- Sharp FR. Where do omics and markers go next? *Transl Stroke Res.* 2010. doi:10.1007/s12975-010-0043-3.
- Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol.* 2008;4:222.
- Stamova B, Xu H, Jickling G, Bushnell C, Tian Y, Ander BP, et al. Gene expression profiling of blood for the prediction of ischemic stroke. *Stroke.* 2010;41:2171–7.
- Marcus A. Testing for ovarian cancer is on the way. *Wall St J.* 2002; D1–D2
- Pollack A. New cancer test stirs hope and concern. *New York Times,* 2004;D1, D6
- Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet.* 2002;359:572–7.

15. Zhu W, Wang X, Ma Y, Rao M, Glimm J, Kovach JS. Detection of cancer-specific markers amid massive mass spectral data. *Proc Natl Acad Sci USA*. 2003;100:14666–71.
16. Ransohoff DF. Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst*. 2005;97:315–9.
17. Anderson NL. The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin Chem*. 2010;56:177–85.
18. Liotta LA, Lowenthal M, Conrads TP, Veenstra TD, Fishman DA, Petricoin III EF. Misinformation generated by lack of communication between producers and consumers of publicly available experimental data. *J Natl Cancer Inst*. 2005;97:310–4.
19. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomics tests for ovarian cancer. *J Natl Cancer Inst*. 2005;97:307–9.
20. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346:1937–47.