GuestEditorial

BY ATUL BUTTE, PhD, ASSISTANT PROFESSOR OF PEDIATRICS
AT STANFORD UNIVERSITY SCHOOL OF MEDICINE

Democratizing Integrative Biology

The word Om (or Aum) has many meanings in ancient Hindu philosophy, one of which is "that which contains all other sounds." The meaning has relevance to the now commonly used suffix "-ome", used to describe the nearly-comprehensive cataloging of discrete or countable items from a single vantage point (e.g. genome, proteome, envirome, and others). Incredible discoveries in life science and medicine have certainly come about from the broadening of thinking of translational scientists, from single molecules to nearly-comprehensive sets of molecules, such as the discovery of molecular subtypes of cancers through gene expression microarrays. But there has also been some disappointment, as some aspects of disease remain resistant to understanding through the measurements we intuitively use, like the genetic architecture of complex diseases still hidden from genome-wide association studies. It is for this reason that integration across measurements made from several vantage points may grant us the missing clues towards deciphering still unsolved mysteries in life science and medicine.

Much has already been written about the potential of integrating the results of cross-modality experiments (*1-4*). Vidal and others have noted that integration of multiple functional maps can lead to novel informatics algorithms and findings (*2, 3*). And a number of research groups have integrated some combination of gene expression data, protein interactions, phenotypes, linkage data, and/or RNAi studies to generate results that are relevant to diseases including obesity, cancer, Leigh syndrome, and cardiovascular disease (*5-10*). I mention these examples to illustrate that exploration into the integration of multiple modalities is well underway, and can yield high-impact results even for translational medical research.

However, the challenge is that methods for studying biology in an integrative manner are not yet easily accessible to most clinician scientists interested in discovering disease mechanisms or disease biomarkers. Integrating results across measurement modalities (e.g. RNA and proteins, genotype and RNA, etc.) requires a level of computational sophistication and biological knowledge that is difficult to operationalize today. This lack of tools has its greatest impact on translational research. Though I acknowledge that clinical scientists have a number of other hurdles to overcome in biomedical institutions (e.g., getting research resources and protected time), I believe that deploying web-based integrative biology tools to clinician scientists could enable them to start hypothesis generation and discovery of candidate markers for the conditions they treat.

For example, an interventional cardiologist empowered in this way might be in the best position to ask a novel biomedical question looking for candidate serum markers for coronary artery stent restenosis across diverse biomedical data. Democratizing the process of integrative biology to the clinician scientist, and providing web-based tools operating in the cloud for them to conduct integrative biology experiments using their own data as well as public data, could eliminate one of the remaining bottlenecks in the translational lifecycle.

I encourage computational scientists to consider developing and deploying tools for the quantitative clinician scientist. As another definition of Om is "the essence of the universe," there are still many –omes remaining for translational scientists to explore, integrate, and harness for the improvement of human health.

Democratizing the process of integrative biology to the clinician scientist, and providing web-based tools operating in the cloud for them to conduct integrative biology experiments using their own data as well as public data, could eliminate one of the remaining bottlenecks in the translational lifecycle.

Footnotes

1. R. Paylor, *Nat Methods* **6**, 253 (Apr, 2009).
2. H. Ge, A. J. Walhout, M. Vidal, *Trends Genet* **19**, 551 (Oct, 2003).
3. M. Vidal, *Cell* **104**, 333 (Feb 9, 2001).
4. A. J. Walhout *et al.*, *Curr Biol* **12**, 1952 (Nov 19, 2002).
5. E. E. Schadt, S. A. Monks, S. H. Friend, *Biochem Soc Trans* **31**, 437 (Apr, 2003).
6. E. E. Schadt *et al.*, *Nature* **422**, 297 (Mar 20, 2003).
7. V. K. Mootha *et al.*, *Proc Natl Acad Sci USA* **100**, 605 (Jan 21, 2003).
8. M. Stoll *et al.*, *Science* **294**, 1723 (Nov 23, 2001).
9. D. R. Rhodes *et al.*, *Proc Natl Acad Sci USA* **101**, 9309 (Jun 22, 2004).
10. S. B. English, A. J. Butte, *Bioinformatics* **23**, 2910 (Oct 5, 2007). □