



"Omic" Risk Assessment

Michael B. Major and Randall T. Moon (26 May 2009)

Science Signaling 2 (72), eg7. [DOI: 10.1126/scisignal.272eg7]

The following resources related to this article are available online at <http://stke.sciencemag.org>.
This information is current as of 23 August 2012.

Article Tools	Visit the online version of this article to access the personalization and article tools: http://stke.sciencemag.org/cgi/content/full/sigtrans;2/72/eg7
Related Content	The editors suggest related resources on <i>Science's</i> sites: http://stke.sciencemag.org/cgi/content/abstract/sigtrans;3/103/pc1 http://stke.sciencemag.org/cgi/content/abstract/sigtrans;2/72/ra25 http://stke.sciencemag.org/cgi/content/abstract/sigtrans;2/70/pt4 http://stke.sciencemag.org/cgi/content/abstract/sigtrans;2/68/eg6 http://stke.sciencemag.org/cgi/content/abstract/sigtrans;1/47/eg9
References	This article has been cited by 1 article(s) hosted by HighWire Press; see: http://stke.sciencemag.org/cgi/content/full/sigtrans;2/72/eg7#BIBL
Glossary	Look up definitions for abbreviations and terms found in this article: http://stke.sciencemag.org/glossary/
Permissions	Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl



Michael B. Major is an Assistant Professor at the Lineberger Comprehensive Cancer Center, Department of Cell and Developmental Biology, University of North Carolina at Chapel Hill.



Randall T. Moon is a member of the *Science Signaling* Editorial Board and the William and Marilyn Conner Chair and Director of the Institute for Stem Cell and Regenerative Medicine, University of Washington School of Medicine, Seattle, and Investigator at Howard Hughes Medical Institute in the Department of Pharmacology.

Cite as:

M. B. Major, R. T. Moon, "Omic" risk assessment. *Sci. Signal.* **2**, eg7 (2009).

10.1126/scisignal.272eg7

"Omic" Risk Assessment

WITH RELATIVE EASE AND DECREASING EXPENSE, WE CAN NOW SEQUENCE ENTIRE GENOMES, simultaneously quantitate the expression of tens of thousands of mRNAs, physically define protein-protein interaction networks, and connect genotype with phenotype through functional genomics. We are no doubt in the scientific age of "omics," which has as a defining characteristic—a lot of data. And therein lies a current and pressing challenge: How do we judge the integrity of the data these technologies generate, and once accepted, how do we best use the data in the construction of predictive models—models that would both help us understand the underlying biology and illuminate new therapeutic strategies?

RNAi-based functional genomics and protein-interaction screens are gaining popularity; they provide unbiased and comprehensive surveys of protein functionality and physical interconnectivity. That said, the imperfections in these technologies are substantial, thus preventing unabridged positivity. At the top of the list is the all too familiar observation that not all primary screen discoveries validate in secondary assays; the data are littered to varying degrees with false positives and false negatives. We present some ideas as to how, when, and if we should deal with the noise.

For both functional genomic and protein interaction screens, the false negative discoveries are easily handled—ignore them, at least for the time being—for as Carl Sagan wrote, "the absence of evidence is not evidence of absence." In time, we will have amassed sufficient RNAi and proteomic data—data derived from diverse biological contexts and from multiple laboratories using distinct tools—that we will become more confident in the validity of negative data.

The false positives, on the other hand, pose a real threat to scientific progress, insofar as they compromise the validity of constructed modeling networks, and become an expensive waste of time and effort if pursued. In our opinion, there is a minimum amount of validation that must be done for siRNA-based discoveries. First, multiple nonoverlapping siRNAs must yield the same phenotype. Ideally, the number of siRNAs tested would exceed three, and a minimum of two should yield a phenotype statistically outside of chance, as indicated with a z-score. Second, the siRNAs targets positive in this primary screen must also be positive in a secondary screen that employs a different phenotypic metric. We consider siRNA screen hits that pass these thresholds to be truly validated and worth the time and energy of pursuit should interest exist. We rely upon a similar but less defined risk assessment strategy for evaluating mass spectrometry (MS)-based protein-interaction networks. At a minimum, the identification of the peptide based on bioinformatics must be trustworthy and the identified interacting protein must not be present in control purifications or across many unrelated preparations. Establishment of the affinity resin-associated background is critical. Beyond these criteria, confidence obviously increases with increasing data, such as affinity purification/MS in the reverse direction, immunoprecipitation (IP)/MS, IP/Western blot, and in vitro binding studies.

So, should we exclude from further consideration siRNA screen hits with a low z-score or a protein-protein interaction based upon a single identified peptide? Absolutely not! For example, exciting insights into Wnt/ β -catenin signaling were obtained not through a focused siRNA screen for regulators of this pathway in a colorectal cancer cell line, but through the discovery that one of the relatively weak hits physically interacted with a protein that is present in the developing pancreas and is mutated in 20% of all pancreatic cancers. To see what is important and real, integrate everything. The integration of data generated through disparate technologies will simultaneously facilitate the elimination of false discoveries and provide insights into mechanisms and disease relevance. In our experience, integrating data from siRNA screens with the complete data from either proteomic screens or from small molecule screens reduces the number of candidate genes or proteins of interest to a number such that investigators can readily pursue further validation assays with all of them. It is, of course, easier to apply a cut-off value to a single screen than to conduct and integrate two technologically distinct screens. However, the revelations that are emerging from integrating distinct screens are worth the added effort.

—Michael B. Major and Randall T. Moon