

Minireview

## Signaling netwErks get the global treatment

Michael B Yaffe<sup>\*†‡</sup> and Forest M White<sup>\*‡</sup>

Addresses: <sup>\*</sup>Center for Cancer Research, <sup>†</sup>Department of Biology and <sup>‡</sup>Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Correspondence: Michael B Yaffe. Email: myaffe@mit.edu

Published: 1 February 2007

*Genome Biology* 2007, **8**:202 (doi:10.1186/gb-2007-8-1-202)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/1/202>

© 2007 BioMed Central Ltd

### Abstract

Two landmark studies of cell signaling, by RNA interference and phosphoproteomics, provide complementary global views of the pathways downstream of receptor kinases, including those regulated by Erks.

In the prelapsarian days of signal transduction, pathways were simple, everything was linear, and life was good. Like the misfortunes that befell Voltaire's *Candide*, however, the unfortunate intervention of reality has clouded that wonderfully innocent view by adding layer upon layer of complexity to what was once neat, clean and relatively easy to think about.

Consider the signaling pathways that regulate protein phosphorylation, the most common post-translational modification of proteins. These pathways and networks regulate almost all aspects of cell biology and, when dysregulated, have been implicated in the pathology of a wide range of human disorders ranging from cancer to neurodegeneration. To understand how information flows through these networks, and to identify critical nodes within protein-kinase signaling pathways that should be targeted for therapeutic intervention, it becomes necessary to understand the molecular mechanisms that underlie network regulation in all their gory detail. This presents us with two fundamental challenges: first, all of the components that touch the network need to be identified, along with their dynamic regulatory points; second, the physical and functional connections within the network must be mapped.

Until now both these challenges have been limited primarily by the lack of appropriate tools for investigating biological systems at the genome- or proteome-wide level. Two recent publications from Friedman and Perrimon [1] and Mann and his colleagues [2] address these challenges. Friedman and Perrimon [1] used genome-wide RNA interference

(RNAi) screening to functionally annotate proteins regulating activation of extracellular signal-regulated kinases (ERKs), whereas Olsen *et al.* [2] carried out a proteome-wide quantitative analysis of protein-phosphorylation dynamics in the epidermal growth factor receptor (EGFR) signaling network. Although neither method by itself provides all the data necessary to understand the signaling networks, together, they reveal, like peeling the layers from an onion, many potential molecular mechanisms that underlie complex biological processes.

### The receptor tyrosine kinase-ERK activation pathway investigated by RNAi

To investigate the proteins that regulate signaling between receptor tyrosine kinases (RTKs) and ERK activation on a global scale, Friedman and Perrimon [1] carried out an unbiased functional screen on engineered *Drosophila* S2 cells using a collection of double-stranded RNAs (dsRNAs) covering more than 95% of the fly genome. In the primary screen, S2R<sup>+</sup> cells stably expressing yellow fluorescent protein (YFP)-tagged Rolled, the *Drosophila* homolog of ERK (dErk), were stimulated with insulin, and the resulting levels of phosphorylated dErk (pdErk) were measured 10 minutes later by immunohistochemistry and normalized to the total amount of YFP-tagged Erk. All measurements were performed in duplicate, using a total of 92,000 dsRNAs against 20,420 genes. The data were then analyzed by computing a Z-score that measured how many standard deviations the pdErk/dErk ratio deviated

from the mean. Surprisingly, more than 5% of all the genes examined (1,168 in total), which included all the core components of the RTK/dErk pathway, had some effect on the extent of dErk activation. Follow-up secondary screens were performed on 362 of these genes using two different cell lines (S2R<sup>+</sup> and Kc167 cells expressing the *Drosophila* EGF receptor type II) together with specific treatments that preferentially activated different RTKs: the insulin receptor, the EGF receptor, and the PVR receptor for platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF). In the end, 331 genes were identified as modulators of the *Drosophila* RTK/dErk pathway, about two-thirds of which have known human homologs. Interestingly, a large number of the dsRNAs had effects that depended on the cell type, whereas others showed alterations in dErk activation that were stimulus-specific.

What 'bottom-line' lessons can we take away from the analysis and characterization of these genes? First, well over 50% of the genes affecting dErk activation corresponded either to unknown gene products or to proteins with no known function, at least as categorized by Gene Ontology (GO) terms [3]. Second, proteins functioning in a wide range of physiological processes, including metabolism, cell-cycle control and mitosis, transcription, translation, RNA binding and splicing, organogenesis, cell migration and apoptosis, impact on the acute 'activatability' of the RTK/dErk pathway, perhaps through tonic negative or positive feedback. Many of these Erk-modulating proteins are themselves known to be regulated by RTK stimulation and ERK phosphorylation. Third, the effects of RNAi-mediated downregulation of these proteins on dErk activation could be subtle - net changes of only 15-30% in dErk activation were observed following downregulation of many of them - attesting to the statistical robustness of the RNAi screening analysis. Fourth, while some families of gene products consistently enhance dErk activation (chaperones, GTPases, trafficking proteins, and proteasome components) or suppress it (ion channels) under all conditions tested, a surprising number of gene products seemed to affect baseline and RTK-stimulated dErk activity in opposite directions.

For example, downregulation of cytoskeletal components and phosphatases by and large seemed to enhance basal Erk activation while suppressing insulin-stimulated activation; downregulation of general transcription factors and splicing components had exactly the opposite effects. In contrast to insulin signaling, basal Erk activation is thought to be due largely to signaling via PVRs responding to endogenous PDGF- and VEGF-related proteins present in or secreted into the medium. Thus, one interpretation of the RNAi data on phosphatases is that activating phosphorylation events are rate-limiting for the PVR pathway, with phosphatases providing tonic inhibition. In contrast, one or more inhibitory phosphorylation sites appear to dominate the

insulin RTK pathway, so that phosphatase activity is required for maximal activation. What are the critical phosphorylated substrates that are the targets of these phosphatases? Answering this question is where the potential of the work of Olsen *et al.* [2] lies, as we will see later.

### Putting genes in order

How else could one organize the Friedman and Perrimon RNAi 'hits'? In a classic genetic screen, one would be able to dissect out the order in which these 331 proteins are acting in the signaling pathways through epistasis analysis (which analyzes whether a mutation in one gene masks the effects of a mutation in a second gene). Unfortunately, as RNAi gives rise to hypomorphic alleles rather than genetic nulls, combining multiple RNAi treatments in such an analysis can give misleading results. To get round this problem, Friedman and Perrimon used a clever trick. They investigated which dsRNAs could suppress dpErk production following induction of a constitutively active allele of the small GTPase Ras (Ras-v12), an intermediate component of the signaling pathway that connects RTKs to Erks. This revealed that 85 of the 331 identified genes were directly affecting the Ras-Raf-Mek-Erk signaling module, while the remaining genes presumably either act upstream of Ras or serve to modulate the linkage between RTK stimulation by its ligand and Ras activation.

A few specific RNAi hits merit comment. Intriguingly, some components of the 'target of rapamycin' (TOR) and AKT (protein kinase B) kinase pathways - including the GTPase Rheb, the TOR substrate-binding protein Raptor, and S6 kinase along with TOR and AKT themselves - seem to inhibit dErk activation. In contrast, the TOR/AKT pathway antagonists TSC2, a component of the Rheb GTPase activator complex, and PTEN, a phosphatase, seem to facilitate Erk activation, probably through indirect effects on the insulin receptor itself. Similar effects of AKT on ERK activation have been seen in mammalian cells [4]. Finally, Friedman and Perrimon [1] studied two novel RTK/dErk-regulating gene products in detail - the Ste20 kinase dCGKIII, and dPPM1, a putative T-loop phosphatase that binds directly to dErk. Friedman and Perrimon convincingly show direct effects of both dCGKIII and dPPM1 knockdowns on Erk-regulated pathways *in vivo* using flies with appropriately sensitized genetic backgrounds, and go on to show that the human homologs of these genes (*MST3/4* and *PPM1 $\alpha$* , respectively) show related effects on mammalian ERK activation in human prostate DU145 and LNCaP cells. Thus, the extensive interconnectedness of the RTK/dErk pathway observed in flies seems to be conserved across wide evolutionary divides.

### A proteomic approach to signaling pathways

In an alternative technical approach to defining a global signaling network resulting from stimulation of the EGFR, Olsen *et al.* [2] used mass spectrometry to identify and

quantify 6,600 protein phosphorylation sites in HeLa cells across 5 time points following EGF stimulation. In this study, HeLa cells were stable-isotope labeled in culture, then stimulated with EGF for either 0, 5 and 10 minutes or 1, 5 and 20 minutes. Following cell lysis and subcellular fractionation into nuclear and cytosolic fractions, proteins were enzymatically digested to peptides. The resulting samples were further fractionated by strong cation exchange, and phosphorylated peptides from each fraction were enriched by titanium dioxide. Finally, a total of 116 liquid chromatography tandem mass spectrometry (LC-MS/MS) analyses were performed to identify phosphorylation sites and measure their level of phosphorylation relative to the 5 minute stimulation point.

The data generated by Mann's group [2] represent a quantum leap forward in efforts to map the global phosphoproteome. They have identified 6,600 phosphorylation sites on 2,244 proteins, quantified these sites across 5 time points of EGF stimulation, and also estimated their subcellular localization, although the efficiency and accuracy of this fractionation is not indicated in the paper. Coverage extends from the autophosphorylation of EGFR that initiates the pathway through to phosphorylation of terminal effectors such as transcription factors, and covers a broad dynamic range of signal intensity. However, even this heroic effort and the massive dataset are still far from comprehensive, as many well-characterized phosphorylation sites are missing from this analysis. Only 103 tyrosine phosphorylation sites, for example, are reported, although others have detected over 300 tyrosine sites in the ErbB signaling network alone [5].

Olsen *et al.* [2] found that tyrosine-phosphorylated peptides occur at a much higher frequency than expected from the abundance ratios of serine, threonine, and tyrosine phosphorylation measured by Hunter and Sefton in the 1980s [6]. The Hunter and Sefton study, however, used phospho-amino acid analysis to reveal only those phosphorylation sites that had incorporated radiolabeled phosphate during the 15-18-hour course of their experiment. In comparison, the ratios determined by Olsen *et al.* [2] reflect the relative frequency of identification of all phosphorylation sites identified in the mass spectrometry study. To compare these studies correctly, it would be necessary to include the absolute abundance of each phosphorylation site in the Olsen *et al.* data.

Their quantification of temporal phosphorylation profiles distinguishes the EGF-responsive phosphorylation sites and significantly enhances the study by Olsen *et al.* [2], as it enables the partial classification of phosphorylation sites through clustering of sites with similar profiles. These data may indicate connectivity in the signaling network but, as with any large-scale dataset, it is difficult to do more than speculate as to the potential pathways involved, and

additional functional validation through biochemical manipulation of the system is required. Regardless of this, the dataset collected by Olsen *et al.* [2] is a rich resource likely to be heavily mined by investigators in the signaling community who wish to examine phosphorylation sites affected by EGF stimulation.

In essence, the beauty of the complementary studies of Friedman and Perrimon [1] and Olsen *et al.* [2] is that the former is essentially a study of function without 'form', whereas the latter concentrates primarily on 'form' in the absence of detailed function. The naysayers among us will conclude from the first study that 'everything is connected to everything' and from the second study that 'everything is also phosphorylated everywhere'. In fact, the data tell a much richer and more subtle story, one that is likely to take the next decade or two to unravel. The shortest path capable of connecting these two datasets remains uncertain, however - how can one use these two compendia to link biological consequence to phosphorylation-site mapping? One answer may lie in extending the analyses to include additional 'orthogonal' quantitative datasets for these same cells under the same conditions. For example, these complementary datasets should probably include some measure of cellular outcomes after RTK stimulation, either by measurements of phenotype (for example, cell proliferation, migration, glucose metabolism, and apoptosis), or through additional 'omics' data, including gene-expression profiles. Mathematical and computational approaches could then be used to identify which pathways, and which particular phosphorylation events, best correlate with a particular phenotype [5,7-9].

Ultimately, if we are to comprehensively ascribe a function to all phosphorylation sites mapped by mass spectrometry, it will probably be necessary to acquire datasets similar to those obtained by Olsen *et al.* [2] across a variety of conditions and in multiple cell lines. Collecting these data for the global phosphoproteome is probably not the best way to tackle this problem, however; such an experiment would generate a glut of data and would require dramatic improvements in both the experimental and data-analysis workflows (Olsen *et al.* carried out 116 LC-MS/MS analyses to decode phosphorylation events at five time points for just one stimulation type in a single cell line).

Even if we assume that such datasets could be obtained, it is not clear how the resulting phosphorylation-site data could best be mapped onto the frizzled 'hairballs' of protein-protein interaction maps [10] to reveal things of biological importance. Instead, we might begin to unite form with function by focusing on insights obtained by the Friedman and Perrimon approach [1], that is, starting with functional screens capable of identifying genes, and their associated proteins, that regulate selected biological processes. One could then proceed to a comprehensive mass-spectrometric

analysis (still a massive undertaking, but less so than determining the global phosphoproteome) aimed at identifying and quantifying phosphorylation on these selected proteins, providing potential molecular mechanisms (form) underlying the characterization provided by the genomic screen (function). Focusing on one type of post-translational modification for a subset of selected proteins known to regulate specific cellular responses could reduce the layers upon layers of complexity to a manageable size, and finally allow us to peel the onion without tears. But perhaps this return to simplicity is just Candide's revenge.

## References

1. Friedman A, Perrimon N: **A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling.** *Nature* 2006, **444**:230-234.
2. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M: **Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.** *Cell* 2006, **127**:635-648.
3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
4. Zimmermann S, Moelling K: **Phosphorylation and regulation of Raf by AKT (protein kinase B).** *Science* 1999, **286**:1741-1744.
5. Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim HD, Grantcharova V, Lauffenburger DA, White FM: **Effects of HER2 overexpression on cell signaling networks governing proliferation and migration.** *Mol Syst Biol* 2006, **2**:54.
6. Hunter T, Sefton BM: **Transforming gene product of Rous sarcoma virus phosphorylates tyrosine.** *Proc Natl Acad Sci USA* 1980, **77**:1311-1315.
7. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK: **Physico-chemical modelling of cell signalling pathways.** *Nat Cell Biol* 2006, **8**:1195-1203.
8. Janes KA, Yaffe MB: **Data-driven modelling of signal-transduction networks.** *Nat Rev Mol Cell Biol* 2006, **7**:820-828.
9. Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M: **Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation.** *Science* 2005, **308**:1472-1477.
10. Han JD, Dupuy D, Bertin N, Cusick ME, Vidal M: **Effect of sampling on topology predictions of protein-protein interaction networks.** *Nat Biotechnol* 2005, **23**:839-844.