Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Cell system perturbation for time-resolved quantification of tyrosine phosphorylation in complex samples

Forest M. White, Biological Engineering, MIT

Signal transduction mediated by protein phosphorylation regulates many cellular biological processes. Aberrations in protein phosphorylation due to kinase (or phosphatase) mutation or overexpression leads to dysregulation of cellular signaling and has been linked to a variety of pathologies, including cancer, autoimmune, and metabolic disorders [1]. Quantification of specific phosphorylation sites regulating signaling pathways involved in these pathological disorders should enable a better understanding of the genesis and progression of the disease state, potentially providing targets for more effective therapeutic intervention.

In order to effectively monitor protein phosphorylation events governing signaling cascades, we have developed a methodology enabling the simultaneous quantification of tyrosine phosphorylation of specific residues on dozens of key proteins in a time-resolved manner. In our initial implementation, we have applied this technology to the analysis of cellular signaling pathways initiated by epidermal growth factor (EGF) stimulation of the epidermal growth factor receptor (EGFR), identifying and quantifying temporal phosphorylation profiles of tyrosine phosphorylation sites on 77 proteins. Application of bioinformatics tools, such as self-organizing maps, has resulted in identification of several cohorts of tyrosine residues exhibiting self-similar temporal phosphorylation profiles, operationally defining dynamic modules in the EGFR signaling network [2].

We have recently extended our analysis of the EGFR signaling network to interrogate the effects of increased expression of HER2, an EGFR family member whose over-expression has been correlated with poor prognosis in several cancer sub-types [3]. In order to better understand the function of HER2 in regulating cellular biology, we used quantitative mass spectrometry to analyze dynamic effects of HER2 over-expression on phosphotyrosine signaling in human mammary epithelial cells stimulated by EGF or HRG.

In order to identify clusters of tyrosine-phosphorylated peptides exhibiting similar temporal dynamics, as well as to globally visualize the high-dimensional information we have obtained, we used the self-organizing map (SOM) algorithm [4]. The SOM is a versatile clustering algorithm that transforms high-dimensional data into lower dimensional display, in a non-linear manner. Here we use the unified-distance matrix (or, U-matrix) approach, which allows robust identification of clusters, and the component plane representation which facilitates comparison of peptide phosphorylation response to exogenous stimuli [4,5]. Self-organizing map analysis of the resulting phosphoproteomic dataset permitted elucidation of network modules differentially regulated in HER2-overexpressing cells in comparison with parental cells for EGF and HRG treatment.

In order to identify key signaling components which regulate downstream biological response, we have constructed a model using partial least-squares regression (PLSR). Information obtained through our proteomic studies was represented in an $M \times N$ matrix (the X-block), where M is the number of conditions investigated, and N is the number of

1-4244-0033-3/06/$20.00 ©2006 IEEE.

peptide metrics measured. An entry in the matrix with coordinates (*i,j*) represents the column *j* metric (i.e. ERK Y187 phosphorylation at 5 minutes) measured under the row *i* condition (i.e. parental cell line treated with EGF). For each condition, the metrics included in the model were phosphorylation measurements at 5, 10, and 30 minutes in addition to the integral of this time course (with integrals being used as a measurement for the 'net' phosphorylation over the 30 minute time course). Cell behavior measurements comprised an *M x P* matrix (the Y-block), where M is again the number of conditions and P is the number of behavior measurements obtained. Partial least-squares regression analysis produced a vector of coefficients indicating the importance of each signaling metric with respect to cellular migration and proliferation.

By combining these modeling approaches, it was possible to associate EGFR-family dimerization at the cell surface to activation of specific phosphorylation sites which appear to most critically regulate proliferation and/or migration. Thus, we have characterized quantitative relationships between diverse ligand stimulation, network-wide signaling, and cell functional responses that may help explain how HER2 over-expression dysregulates network signals governing important downstream cell behaviors.

References

1. Hunter T (2000) Signaling--2000 and beyond. *Cell* **100**:113-127.

2. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* **4**:1240-1250.

3. Hynes NE, Lane HA (2005) ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* **5:** 341-354.

4. Kohonen T (2001) *Self-Organizing Maps*. Springer, Berlin, Germany

5. Ultsch A, Siemon H (1989) Exploratory data analysis: Using Kohonen networks on transputers.