

# High-throughput Phosphotyrosine Proteomics Analysis in EGFR Pathway using Self-Organizing Maps

Sampsa Hautaniemi<sup>1</sup>, Alejandro Wolf-Yadlin<sup>1</sup>, Yi Zhang<sup>1</sup>, Forest M. White<sup>1</sup>, Douglas A. Lauffenburger<sup>1</sup>

**Keywords:** Systems biology, computational proteomics, clustering, self-organizing map

## 1 Introduction.

Most biological responses of higher organisms to drugs and other extracellular cues are cellular receptor related. Activation of receptors, e.g. due to ligands, leads to activations of several signaling protein pathways, which in turn govern physiological cell behavioral functions, such as proliferation, death, differentiation, and migration. An important class of receptors is that of receptor tyrosine kinases (RTK) because different tyrosine phosphorylation sites can act as docking sites for a variety of proteins and thereby can activate several signaling protein pathways. One of the most prominent RTK systems for therapeutical intervention in human cancers is epidermal growth factor receptor (EGFR). The EGFR system has several tyrosine kinase docking sites that are able to initiate several signaling pathways, such as the mitogen-activated protein kinase (MAPK) and the phosphoinositide-3 (PI3) kinase pathways, which play crucial roles in cancer development.

As the phosphotyrosine sites of EGFR are starting points for several signaling pathways whose dysfunction has been linked to human cancers, successful *in silico* modeling requires further elaboration on what pathways are activated by a set of extracellular cues for EGFR. To this end, we quantitatively measured phosphorylation level of phosphotyrosine sites using high-throughput LC/MS/MS [3]. However, high-throughput measurement methods call for mathematical methods that are able to deal with the deluge of data. Here, we answer this need with Self-Organizing Maps (SOMs) and demonstrate how the SOM can be used in clustering and analyzing temporal activation of phosphotyrosine sites.

## 2 Methods.

We analyzed phosphotyrosine sites in the EGFR signaling network from human mammary epithelial cells (HMEC) across four times points (0, 5, 10, 30 minutes) and identified >70 phosphorylation sites in >50 different proteins [3]. Here, we normalized the data with respect to the five minute time point, clustered the data using the SOM and visualized using the component plane presentation [2]. The SOM plot for the data across three time-points is given in Figure 1. The three component planes show temporal behavior of the phosphotyrosine sites and from the U-matrix two clusters are evident.

The two clusters consist of 18 and 38 phosphorylation sites, and the phosphorylation sites contained within these clusters can be explained biologically. For example, in Cluster 1, nine phosphorylation sites (50%) are ephrins, which in turn are strongly linked with cancer development; in one of our earlier studies we found that the receptor tyrosine gene *EphB2* is mutated in human prostate cancer [1]. In addition to identifying sets of similarly activated phosphotyrosine sites, the results provide insights on the dynamics of individual

---

<sup>1</sup>Biological Engineering Division, MIT, E-mails: {sampsah, alewolf, yzh, fwhite, lauffen}@mit.edu

phosphotyrosine sites. As such, our results can be used in directing future experiments that aim at elucidating cellular signaling pathways. In summary, our results demonstrate that the SOM provides an excellent format for visualization and analysis of high-throughput proteomics data.

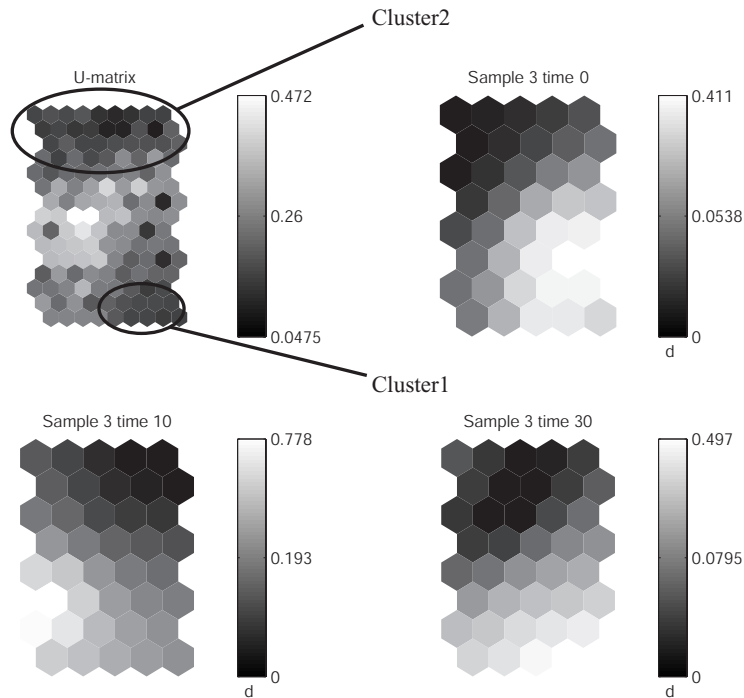


Figure 1: The SOM display after clustering the 81 phosphorylation sites.

## References

- [1] Huusko, P., Ponciano-Jackson, D., Wolf, M., Kiefer, J., Azorsa, D., Tuzmen, S., Weaver, D., Robbins, C., Moses, T., Allinen, M., Hautaniemi, S., Chen, Y., Elkahoulou, A., Basik, M., Bova, G., Bubendorf, L., Lugli, A., Sauter, G., Schleutker, J., Ozcelik, H., Elowe, S., Pawson, T., Trent, J., Carpten, J., Kallioniemi OP. and Mousses, S. 2004. NMD microarray analysis reveals mutations of the *EphB2* gene in human prostate cancer. *Nature Genetics* 36:979–983.
- [2] Vesanto, J., Himberg, J., Alhoniemi, E. and Parhankangas, J. 2000. SOM toolbox for Matlab 5. Technical Report A57, *Helsinki University of Technology*, Finland.
- [3] Zhang, Y., Wolf-Yadlin, A., Pappin, D.J., Rush, J., Lauffenburger, D.A. and White, F.M. 2005. Time-resolved phosphotyrosine analysis reveals dynamics modules in EGFR signaling. Submitted.