

## Omic-Sig: Utilizing Omics Data to Explore and Visualize Kinase-Substrate Interactions

**T. Mamie Lih <sup>1,\*</sup>, David J. Clark <sup>1</sup> and Hui Zhang <sup>1</sup>**

<sup>1</sup> Johns Hopkins University, School of Medicine, Baltimore, MD 21287, USA.

\* Correspondence: [tlih1@jhmi.edu](mailto:tlih1@jhmi.edu)

**Abstract:** Protein phosphorylation is one of the most prevalent post-translational modifications, resulting from the activity of protein kinases phosphorylating specific substrates. Multiple cellular processes are regulated via protein phosphorylation, with aberrant signaling driven by dysregulation of phosphorylation events and associating with disease progression (e.g., cancer). Mass spectrometry-based phosphoproteomics approaches can be leveraged for studying alterations of kinase-substrate activity in clinical cohorts. However, the information gained via interrogation of global proteomes and transcriptomes can offer additional insight into the interaction of kinases and their respective substrates. Therefore, we have developed the bioinformatics, data visualization software tool, Omic-Sig, which can stratify prominent phospho-substrates and their associated kinases based on differential abundances between case and control samples (e.g., tumors and their normal adjacent tissues from a cancer cohort) in a multi-omics fashion. Omic-Sig is available at <https://github.com/hzhangjhu/Omic-Sig>.

## Introduction

Protein phosphorylation is one of the more prominent post-translational modifications, serving an important role in regulating multiple cellular processes, such as cell growth, cell differentiation, and metabolism.<sup>1,2</sup> Phosphorylation occurs when protein kinases phosphorylate their selected substrates, leading to a cascade of signaling events. Multiple reports have indicated that aberrant phosphorylation-based signaling can drive various diseases, including cancer.<sup>1,3</sup>

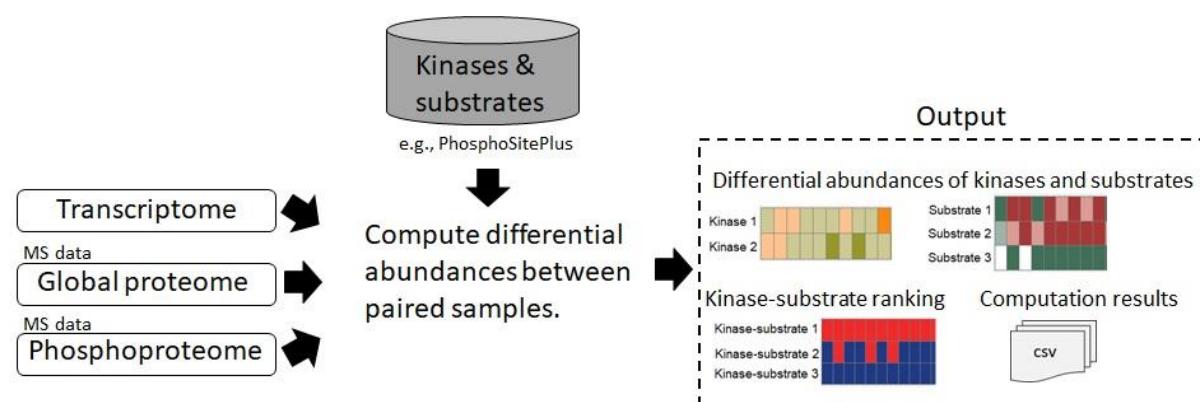
Mass spectrometry (MS) is a wide-spread analytical tool that can be applied to comprehensively investigate the phosphoproteome. By conducting comparative analysis using phospho-data of paired samples (e.g., tumors and their normal adjacent tissues (NAT) in a cancer cohort), a deeper understanding of disease can be gained and lead to the identification of potential therapeutic targets.<sup>4</sup> Furthermore, coupling phosphoproteomics with global proteomic and transcriptomic profiles can provide more insight into the altered expression profiles of kinases and their respective substrates.<sup>3</sup>

In this study, we have developed a software tool, Omic-Sig, to explore and visualize the kinase-substrate interactions by computing and ranking the differential abundances between case and control samples. Omic-Sig is constructed using C# (Microsoft Visual Studio 2017) and R (version 3.5) to provide a user-friendly interface and comprehensive data visualization.

## Methods

### Workflow of Omic-Sig

As shown in Figure 1, Omic-Sig can accept MS-based global proteomic and phosphoproteomic data as well as transcriptomic data generated from RNA-Seq as input. For each uploaded omic data set, the abundances between case and control samples (e.g., Tumor/NAT pair) are computed. By importing information on kinases and substrates, Omic-Sig will stratify kinase-substrate interactions by ranking the cases based on the calculated fold changes among different substrates of a kinase to obtain the highest ranked phosphorylation events in the majority of cases. Only phosphorylation events have more than 60% (user-adjustable) of cases with  $> 1.5$  fold change (user-adjustable) are considered for the stratification.



**Figure 1.** Workflow of Omic-Sig.

### *Input Data*

Omic-Sig requires expression matrices of global proteins and phosphosites as well as information on kinase-substrate interactions (e.g., PhosphoSitePlus<sup>5</sup>). It is optional to upload

an mRNA expression matrix. For the expression matrices, the columns represent case or control samples, whereas each row may represent a protein, gene, or phosphosite, depending on the type of the omic data. To correctly infer the sample pairs, users need to provide a sample information table, where each row represents a sample and columns show the corresponding sample type (e.g., Tumor or NAT) and pair tag (i.e., a pair of samples is assigned the same identification tag). The expression matrix should be normalized and log-transformed prior to uploading into Omic-Sig.

### *Output*

We utilized a CPTAC (Clinical Proteomic Tumor Analysis Consortium) colon cancer dataset,<sup>4</sup> downloaded from LinkedOmics,<sup>6</sup> to exemplify the Omic-Sig pipeline. Following Omic-Sig analysis, a table of stratified kinase-substrate interactions is displayed (Figure 2) on the user-interface. Separate tables showing differential abundances between paired samples from each omic data for kinases and substrates are displayed as well (data not shown). All the computational results are automatically exported as csv files in user-specified output directory. Users can visualize the results by selecting the kinases-substrate interactions of interest (Figure 3).

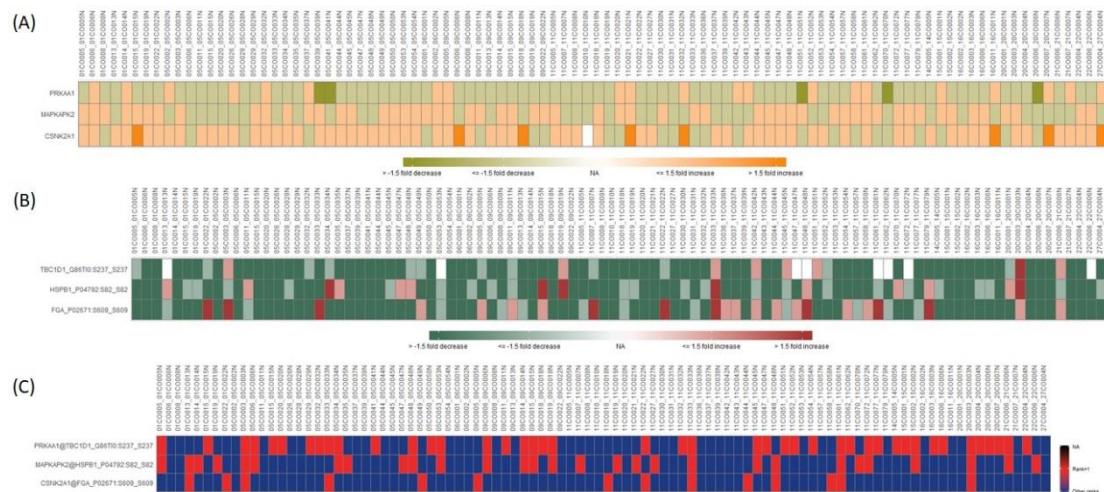
Results

Plot

Kinases and substrates Differential expression of phospho Differential expression of kinases

Select	Kinase	Substrate	Protein	Substrate.Seq	Substrate.Sites	# of Rank 1	01CO005_01CO00	01CO006_01CO00
<input type="checkbox"/>	PRKCD	RPS3		P23396:T221	T221	96	-0.146	0.341
<input type="checkbox"/>	CSNK2A1	MDC1		Q14676:S453	S453	75	1.39	1.022
<input type="checkbox"/>	CSNK2A1	CAV2		P51636:S23	S23	5	-1.008	-1.961
<input checked="" type="checkbox"/>	CSNK2A1	FGA		P02671:S609	S609	14	-1.509	-3.173
<input type="checkbox"/>	CSNK2A1	L1CAM		P32004:S1181	S1181	2	-2.368	-3.848
<input checked="" type="checkbox"/>	PRKAA1	TBC1D1		Q86T10:S237	S237	43	-0.331	-1.76
<input type="checkbox"/>	PRKAA1	LIPE		Q05469:S855	S855	53	-2.729	-0.631
<input type="checkbox"/>	CAMK1	LIPE		Q05469:S855	S855	96	-2.729	-0.631
<input type="checkbox"/>	PRKCB	PRKCB		P05771-2:S660	S660	96	-0.075	-0.536
<input checked="" type="checkbox"/>	MAPKAPK2	HSPB1		P04792:S82	S82	29	-1.491	-2.042
<input type="checkbox"/>	MAPKAPK2	HSPB1		P04792:S15	S15	27	-2.059	-2.252

**Figure 2.** Kinase-substrate table shows kinases and their respective substrates as well as differential abundances between case and control samples for the substrates. Highlighted rows are the selected kinase-substrate interactions for visualization.



**Figure 3.** Visualizing kinase-substrate interactions. Omic-Sig generates the plots for selected kinase-substrate pairs. (A) and (B) shows differential abundances across case and control samples for selected kinases and phospho-substrates. (C) Ranked kinase-substrate events.

## Conclusions

In summary, Omic-Sig is a bioinformatics tool to stratify phospho-substrates and their associated kinases by utilizing the differential abundances between case and control samples in phosphoproteomics, global proteomics, and transcriptomics data. The tool provides data visualization on the kinase-substrate interactions allowing users to easily examine the similarity and differences across samples and different omics data. To the best of our knowledge, Omic-Sig is the first tool using paired samples to study and visualize kinase-substrate relations in multi-omics fashion.

## Funding

This work was supported by the National Institutes of Health, National Cancer Institute, the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA210985) and the Early Detection Research Network (EDRN, U01CA152813).

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- (1) Ubersax, J. A.; Ferrell, J. E., Jr. *Nat Rev Mol Cell Biol* **2007**, *8*, 530-541.
- (2) Yang, C. Y.; Chang, C. H.; Yu, Y. L.; Lin, T. C.; Lee, S. A.; Yen, C. C.; Yang, J. M.; Lai, J. M.; Hong, Y. R.; Tseng, T. L.; Chao, K. M.; Huang, C. Y. *Bioinformatics* **2008**, *24*, i14-20.
- (3) Drake, J. M.; Paull, E. O.; Graham, N. A.; Lee, J. K.; Smith, B. A.; Titz, B.; Stoyanova, T.;

Faltermeier, C. M.; Uzunangelov, V.; Carlin, D. E.; Fleming, D. T.; Wong, C. K.; Newton, Y.;

Sudha, S.; Vashisht, A. A.; Huang, J.; Wohlschlegel, J. A.; Graeber, T. G.; Witte, O. N.; Stuart,

J. M. *Cell* **2016**, *166*, 1041-1054.

(4) Vasaikar, S.; Huang, C.; Wang, X.; Petyuk, V. A.; Savage, S. R.; Wen, B.; Dou, Y.; Zhang,

Y.; Shi, Z.; Arshad, O. A.; Gritsenko, M. A.; Zimmerman, L. J.; McDermott, J. E.; Clauss, T.

R.; Moore, R. J.; Zhao, R.; Monroe, M. E.; Wang, Y. T.; Chambers, M. C.; Slebos, R. J. C.,

et al. *Cell* **2019**, *177*, 1035-1049 e1019.

(5) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E.

*Nucleic Acids Res* **2015**, *43*, D512-520.

(6) Vasaikar, S. V.; Straub, P.; Wang, J.; Zhang, B. *Nucleic Acids Res* **2018**, *46*, D956-D963.