Robust Inference of Kinase Activity using Functional Networks

Serhan Yılmaz[1], Marzieh Ayati[2], Daniela Schlatzer[1], A. Ercüment Çiçek[3,4], Mark R. Chance[1], Mehmet Koyutürk[1]

1 2 3 4

TALK: Monday, July 13, 2020 4:00 PM-4:20 PM @Virtual-Conference

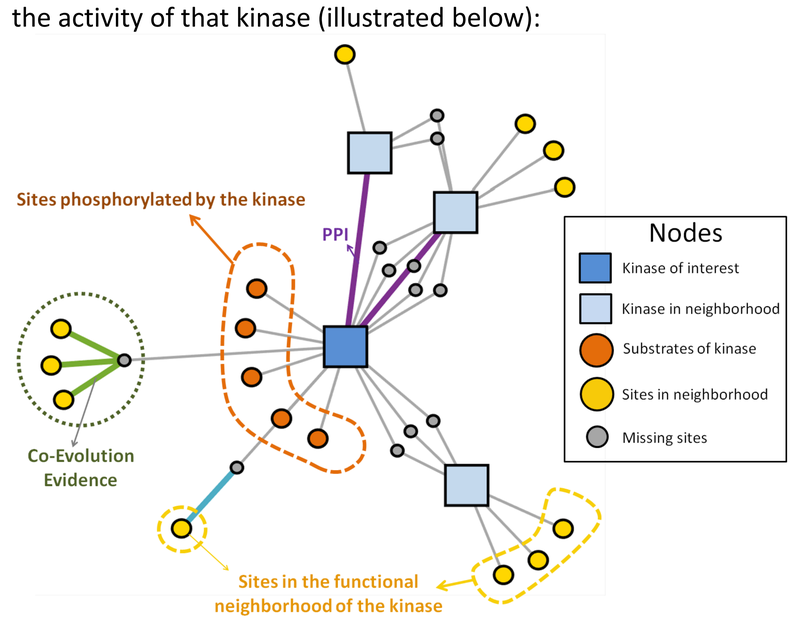
Web Application: http://rokai.io **Contact:** serhan.yilmaz@case.edu

Summary

Mass spectrometry enables high-throughput screening of phospho-proteins across a broad range of biological contexts. When complemented by computational algorithms, phospho-proteomic data allows the inference of kinase activity, facilitating the identification of dysregulated kinases in various diseases including cancer, Alzheimer's disease and Parkinson's disease. To enhance the reliability of kinase activity inference, we present a network-based framework, RoKAI, that integrates various sources of functional information to capture coordinated changes in signaling. Through computational experiments, we show that phosphorylation of sites in the functional neighborhood of a kinase are significantly predictive of its activity. The incorporation of this knowledge in RoKAI consistently enhances the accuracy of kinase activity inference methods while making them more robust to missing annotations and quantifications. This enables the identification of understudied kinases and will likely lead to the development of novel kinase inhibitors for targeted therapy of many diseases. RoKAI is available as web-based tool at http://rokai.io.

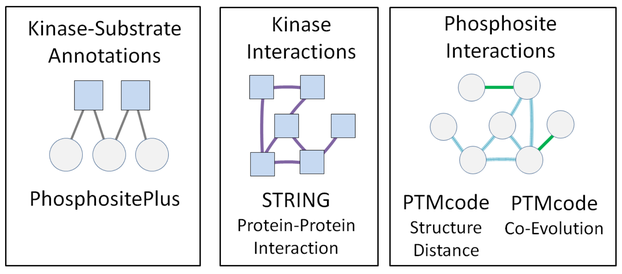
Our Method — RoKAI

To improve the reliability of the kinase activity inference methods, we present RoKAI: A network-based framework that integrates various sources of functional information to capture coordinated changes in signaling. We hypothesize that biologically significant changes in signaling manifest as hyper-phosphorylation or de-phosphorylation of multiple functionally related sites. Therefore, having consistently hyper-phosphorylated (or de-phosphorylated) sites in the functional neighborhood of a kinase can provide further evidence about the activity of that kinase (illustrated below):



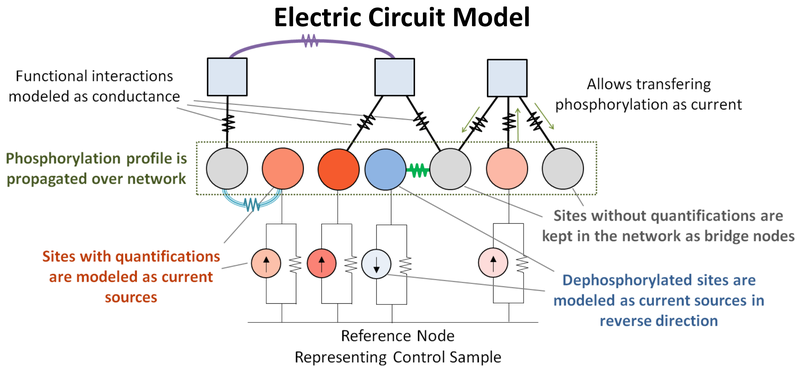
Heterogeneous Network Model

Our framework, RoKAI, uses a heterogeneous network model having kinases and phosphosites as nodes to integrate relevant sources of functional information, including:



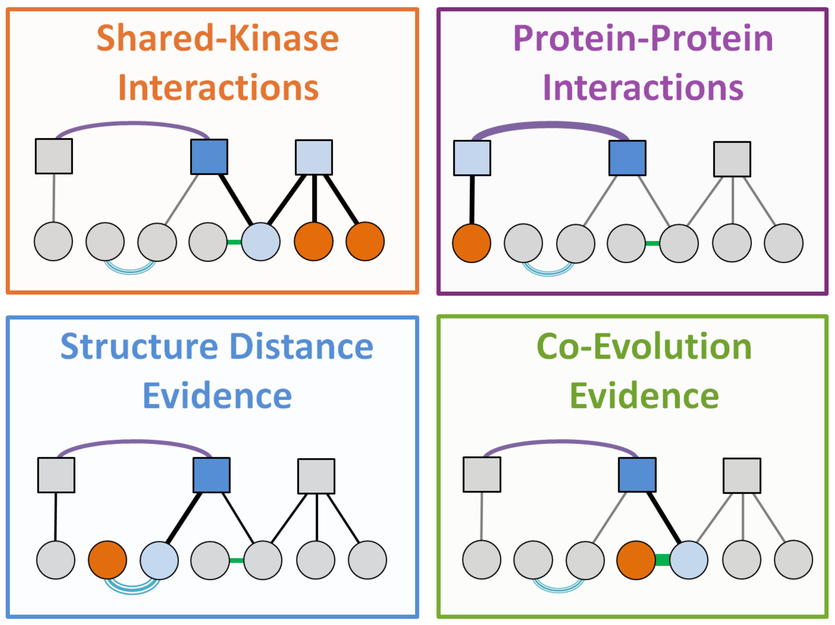
Network Propagation in RoKAI

On this heterogeneous network, we propagate phosphorylation profile to obtain representative phosphorylation levels capturing coordinated changes in signaling. We develop an electric-circuit based network propagation algorithm (illustrated below) that is specifically designed to accommodate missing sites not identified by MS. To predict the kinase activities, we use the resulting representative phosphorylation levels in combination with existing kinase activity inference methods.



Predictivity of sites in functional neighborhood

We hypothesize that phosphorylation of sites in the functional neighborhood of a kinase would be predictivity of its activity. Overall, we consider four types of information sources for functional neighboorhood. To test their utility, for each kinase, we compute a representative score equal to the mean phosphorylation of sites in the neighborhood of the kinase (but not a known substrate of that kinase). See below for an illustration of sites in the neighborhood for each type:



To assess the utility of these potential information sources for inferring kinase activities, we make use of four summary statistics:

- Predictivity: Difference in the distributions of 'true' annotated kinases and other kinases in terms of the phosphorylation of sites in the functional neighborhood for the corresponding information source.
- Coverage: Fraction of kinases with at least one site in the neighborhood for the corresponding information source.
- Complementarity: Measures whether the phosphorylation of sites in the neighborhood of a kinase would provide complementary information, in addition to the phosphorylation of known substrates.
- Overall Effect: An estimation for overall utility of information sources.

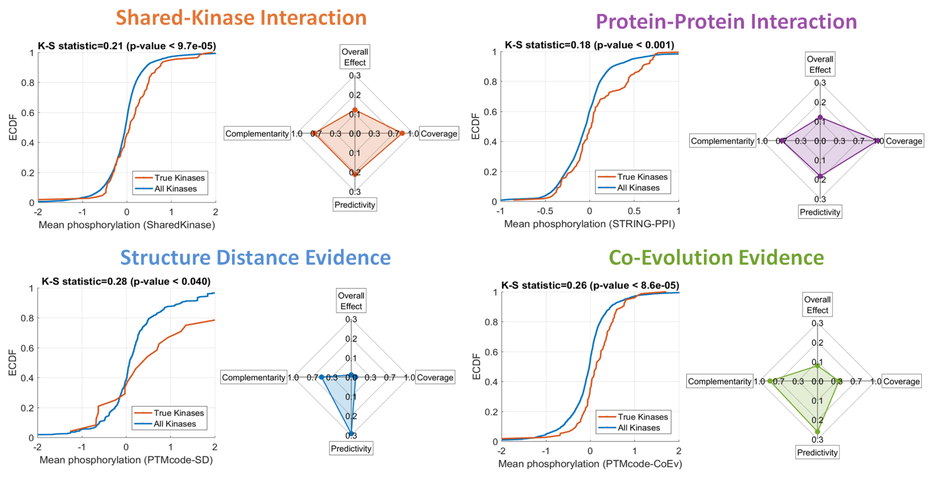
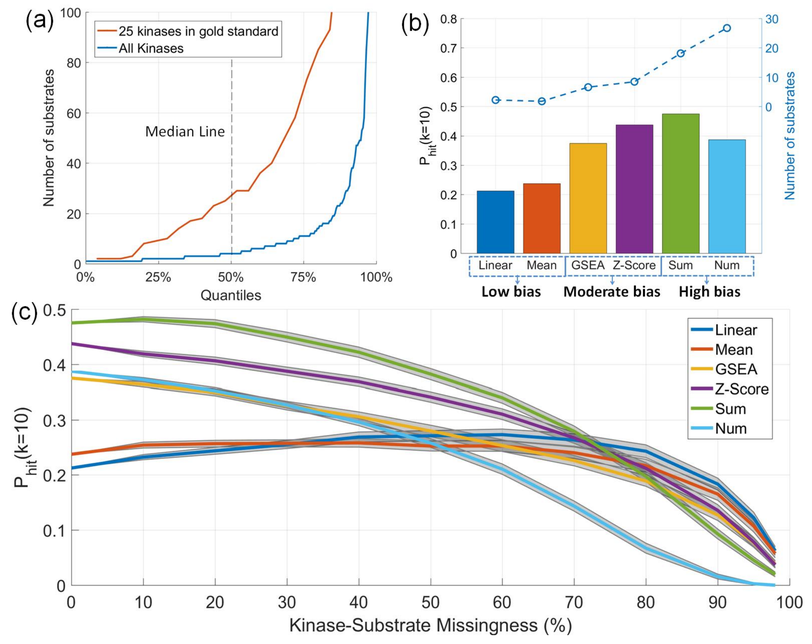


Fig. 1. In each panel, the left plot compares the empirical cumulative distribution (ECDF) of the phosphorylation levels of the "information-providing" sites for "true" perturbed kinases in the benchmark data against all other kinases. Whereas, the right plot shows the summary statistics for the utility of the information sources.

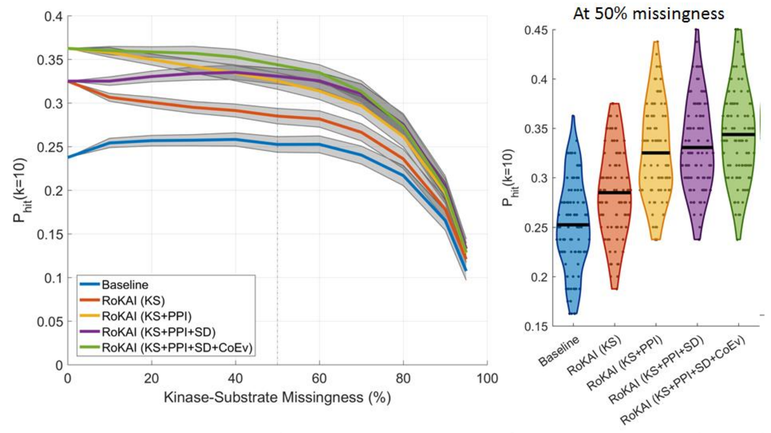
Bias in benchmarking data

For benchmarking, we use the gold standard annotations (curated by Hernandez-Armenta et al., 2017). We observe that there is a substantial bias in these annotations toward rich kinases with many known substrates (panel a). This causes an issue for evaluation as methods biased towards rich kinases have higher performance estimations due to bias (panel b). To overcome this issue, we perform a robustness analysis (panel c) and show that methods with high bias are not robust for missing kinase-substrate annotations.

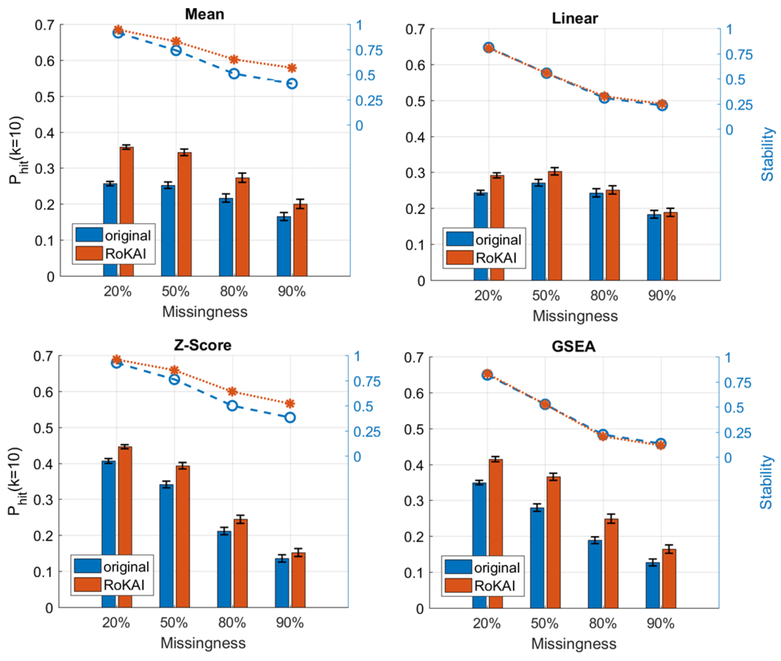


Benchmarking RoKAI

First, we gradually explore a set of functional networks with RoKAI. To benchmark the performance of RoKAI with these networks, we use mean substrate phosphorylation (baseline method) to infer kinase activities based on RoKAI-enhanced phosphorylation profiles and perform a robustness analysis. As seen below, RoKAI improves the prediction accuracy in a robust manner (regardless of missingness) and the addition of each network provides further improvements.



Next, we investigate the improvement of RoKAI over a broad range of kinase activity inference methods by performing a robustness analysis. As it can be seen below, for all methods and missingness values tested, RoKAI consistently enhances the accuracy of the inference methods and makes them more robust to missing kinase-substrate annotations:



Incorporating Missing Phosphosites

RoKAI can further improve the kinase activity inference by utilizing of the unidentified sites (without quantifications) as bridge nodes connecting other phosphosites.

Fig. 2. In type I (illustrated in top left), the network used by RoKAI consists only of sites with quantifications. Whereas, in type II (illustrated in top right), the network additionally includes sites without quantifications to utilize them as bridge nodes.

How to use RoKAI in your research?

RoKAI is available as web-based tool at http://rokai.io

You can run RoKAI on your data by uploading a single file having:

Protein: Ensembl protein identifier

Position: Position of the site

Quantification: Log2 fold change for the phosphorylation of the site

Protein	Position	Quantification
ENSP00000005226	283	0.37
ENSP00000005226	287	1.4
ENSP00000005260	261	0.54
ENSP00000005260	293	-1.83
ENSP00000005260	295	0.05

RoKAI App

Robust Inference of Kinase Activity

using network propagation on functional networks