

# Robust Inference of Kinase Activity Using Functional Networks

## Abstract

**Motivation:** Recent developments in mass spectrometry (MS) enable high-throughput screening of phospho-proteins across a broad range of biological contexts. Phospho-proteomic data complemented by computational algorithms enable the inference of kinase activities based on the phosphorylation levels of their substrates. This leads to the development of kinase inhibitors for targeted therapy of various diseases, including cancer, Alzheimer’s disease and Parkinson’s disease, among others. However, the inadequacy of known kinase-substrate associations and the incompleteness of MS-based phosphorylation data pose important limitations on inference of kinase activity in a robust manner.

**Limitations of existing inference methods:** The atlas of human kinase regulation provides phospho-proteomic data for more than eighty perturbation studies, providing a useful resource for benchmarking kinase activity inference. In our computational experiments, we observe that the benchmark data is substantially biased in favor of “rich kinases” with many known substrates. To elucidate the effect of this bias, we use Monte Carlo simulations to perform a robustness analysis with varying levels of missingness. The results of this analysis shows that some of the most commonly utilized methods for kinase activity inference are vulnerable to the incompleteness of available kinase-substrate annotations, which raises concerns about their reliability.

**Our Approach:** With a view to improving the robustness of kinase activity inference methods to missing annotations, we develop a framework that comprehensively utilizes available functional information on kinases and their substrates. Our framework, RoKAI, uses a heterogeneous network model to integrate relevant sources of functional information, including: (i) kinase-substrate associations from PhosphositePlus, (ii) co-evolution and structural distance evidence between phosphosites from PTMcode, and (iii) protein-protein interactions (PPI) from STRING for interactions between kinases. On this heterogeneous network, we propagate the quantifications of phosphosites to obtain representative phosphorylation levels capturing coordinated changes in signaling. We develop a network propagation algorithm that is specifically designed to accommodate missing sites not identified by MS. To predict changes in kinase activity, we use the resulting representative phosphorylation levels in combination with existing kinase activity inference methods.

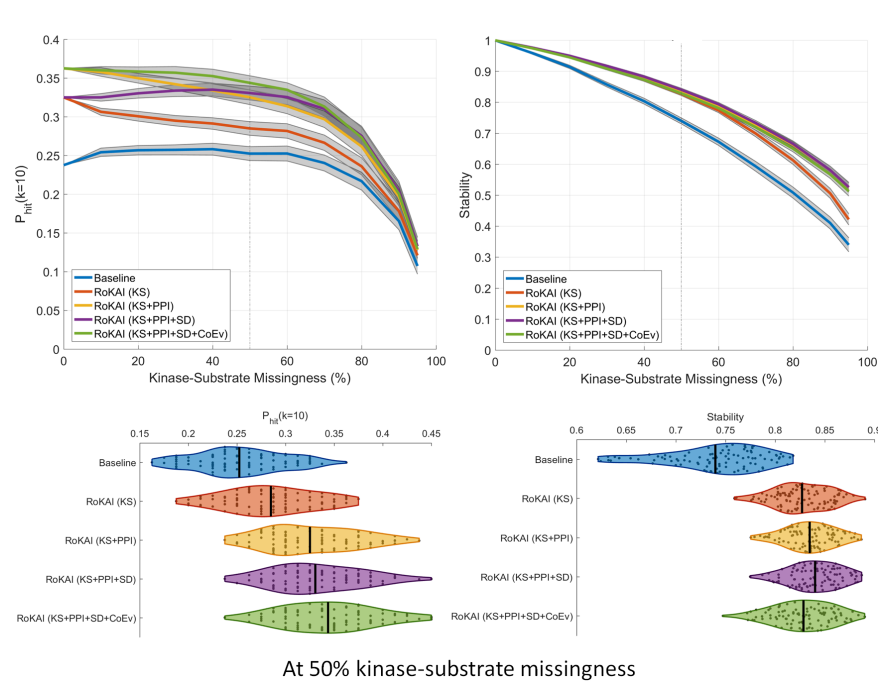
**Utility of functional information for kinase activity inference:** We first characterize the contribution of each source of functional information on enhancing kinase-activity inference. Our results show that incorporation of “shared kinase associations” (i.e., transferring information between sites targeted by the same kinase) significantly improves kinase activity inference. We observe that, other sources of functional information considered (PPI, co-evolution and structure distance evidence) also provide statistically significant information for kinase activity inference. However, their contribution is smaller in comparison due to either (i) limited coverage or (ii) low complementarity with existing kinase-substrate annotations.

**Performance of RoKAI-enhanced inference methods:** Finally, we systematically investigate the performance of RoKAI in improving the performance of existing kinase activity methods (Figure 1). Results of these computational experiments show that RoKAI consistently improves the accuracy, stability, and robustness of commonly used kinase activity inference methods that are benchmarked.

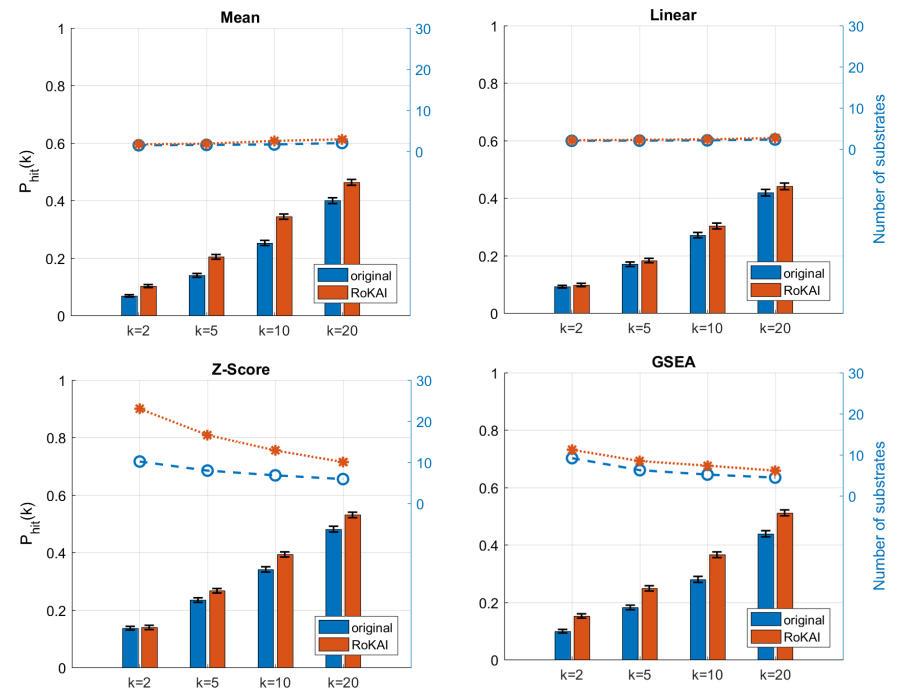
**Conclusion:** Overall, our results demonstrate the utility of functional information in expanding the scope of kinase activity inference and establish RoKAI as a useful tool that can lead to the identification of novel kinases as potential targets for various diseases.

**Availability:** RoKAI is available as a web-based tool at <http://rokai.io>. The source code in MATLAB is available at <http://compbio.case.edu/omics/software/rokai>.

**Preprint in bioRxiv:** <https://doi.org/10.1101/2020.05.01.062802>



(a) Comparison of the accuracy and stability of mean substrate phosphorylation (baseline implemented by KSEA) and its RoKAI-enhanced versions using various functional or structural networks. (Top-Left) The hit-10 performance (the probability of ranking a true perturbed kinase in the top ten), as a function of missingness (the fraction of kinase-substrate associations that are hidden). (Top-Right) Stability of the inferred activities (measured by the average squared correlation between inferred activities when different portions of kinase-substrate associations are hidden from the inference methods), as a function of missingness. (Bottom-Left) The distribution of hit-10 probabilities for 100 instances at 50% missingness. (Bottom-Right) The distribution of stability for 100 instances at 50% missingness.



(b) Contribution of RoKAI (combined network) in improving the performance of different kinase activity inference methods for predicting the true (annotated) kinase in the top  $k$  kinase predictions for various  $k$ . The bars show the mean probability of predicting a true kinase among the top  $k$  kinases at 50% kinase-substrate missingness. The blue bars indicate the prediction performance using the original (unmodified) phosphorylation profiles and red bars indicate the performance of using RoKAI-enhanced profiles for inferring kinase activity. The dashed lines indicate the average number of substrates of the top  $k$  kinases predicted by the corresponding inference method. The black error bars indicate the 95% confidence intervals for the mean performance across 100 runs.

Figure 1. Performance of RoKAI-enhanced inference methods.