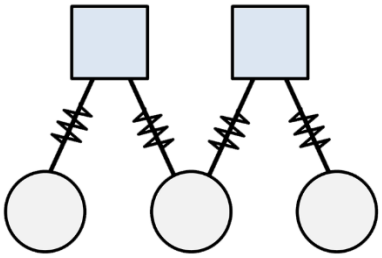


RoKAI App

Robust Inference of Kinase Activity

using network propagation on functional networks

User Manual

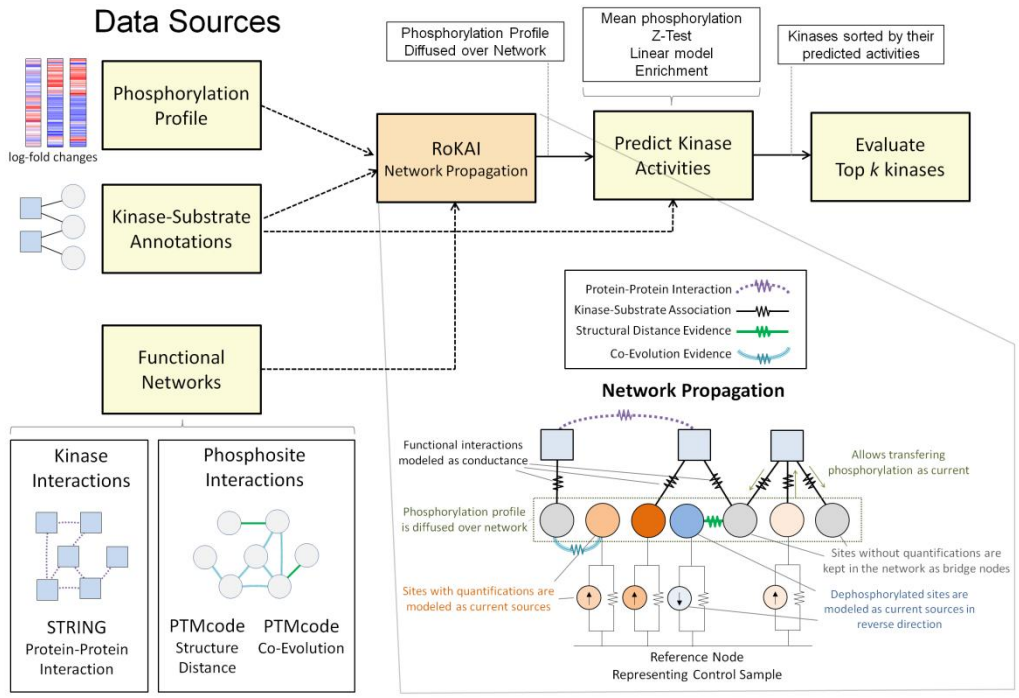
Table of Contents

Case Western Reserve University

Email: serhan.yilmaz@case.edu

Chapter A

Introduction



RoKAI is a computation tool for inferring kinase activity in a robust manner using functional networks.

RoKAI operates on a heterogeneous network having kinases and phosphosites as nodes and available functional associations as edges, including protein-protein interactions, kinase-substrate annotations, co-evolution and structure distance evidence between phosphosites. The key idea of RoKAI is to propagate the phosphosite quantifications on this heterogeneous network to capture the coordinated changes in the signaling, which are used to infer the kinase activities in a more robust manner.

RoKAI is implemented in MATLAB and the source code is available from [link]. To obtain the up-to-date version, please visit the Github page [link]. If you are interested in performing RoKAI through a user-friendly online interface, please visit RoKAI Web Application [link].

Using RoKAI Web Application

1) Input Data Format

RoKAI requires an input file following a specific format: The data must be a comma separated file (.csv) containing the following three columns with the specified headers:

A	B	C
Protein	Position	Quantification
ENSP00000016946	788	-0.62
ENSP00000019317	116	-0.32
ENSP00000019317	463	0.295
ENSP00000025301	448	0.1

Protein: Ensembl Protein (ENSP) Identifier
Position: Position of the site on the protein
Quantification: Phosphorylation of the specified site, provided as log2 fold change

You can use a sample input file by clicking on Sample Data button on RoKAI web application. You can also download this file by clicking on the Download button. Alternatively, you can download it from the Github repository [link].

2) Upload Data

After preparing the data, you can upload it simply by clicking on the Upload button on the left panel and selecting the file on your computer.

3) Select Parameters

1. Select "Method" for kinase activity inference:
   This is the inference method that will be used to predict the activity of a kinase based on the refined phosphorylation profiles created by RoKAI. There are three options:
   - "Mean" (default): The mean phosphorylation (log2-fold changes) of the known kinase substrates are taken as the activity of a kinase.
   - "Z-Score": It is the normalized version of the mean substrate phosphorylation ("Mean") according to statistical significance. Kinase-substrate enrichment analysis (KSEA) of Casado et al. (2013) [link] uses this approach.
   - "LinearModel": The linear model, considered by Hernandez-Armenta et al. (2017) [link], aims to take into account of the dependencies between kinases that phosphorylate the same site. A similar (but more complex) approach is also utilized by IKAP (Mischnik et al., 2015 [link]). In this model, the phosphorylation of a site is modeled as summation of the activities of kinases that phosphorylate the site.

2. Select *"Network"* for RoKAI*:*

   This option determines the functional network used by RoKAI to propagate the phosphosite quantifications to obtain robust phosphorylation profiles. There are four options:

   - "KinaseSubstrate" network (default): The network consists only of known human kinase-substrate interactions obtained from PhosphositePlus [link]. RoKAI uses kinases as bridges to propagate the quantifications on this network in order to take into account of shared-kinase interactions between sites that are phosphorylated by the same kinase.
   - "KS+PPI" network: In addition to "KinaseSubstrate", this network contains weighted edges between kinases based on the human protein-protein interaction network obtained from STRING [link].
   - "KS+PPI+SD" network: In addition to "KS+PPI", this network contains edges between phosphosites with structure distance evidence obtained from PTMcode [link].
   - "KS+PPI+SD+CoEv" network: In addition to "KS+PPI+SD", this network contains edges between phosphosites with co-evolution evidence obtained from PTMcode [link].

3. Select "Missingness" option for RoKAI:

   This option essentially determines whether phosphosites without quantifications should be kept in the network. There are two choices:

   - "Type I": Only phosphosites with quantifications (provided in the specified input file) are kept in the network.
   - "Type II" (default): All phosphosites are kept in the network. The phosphosites without quantifications are used as bridge nodes in RoKAI. This option is typically slower than "Type I".

4) Click Run!

This will direct you to Visualization tab once the analysis is done. Note that, depending on the size of the input and the specified options, this may take some time.

5) Save the results

You can download the inferred kinase activities as well as the refined phosphorylation profiles generated by RoKAI from the "Kinase Activities" tab.

Chapter C

Installing RoKAI desktop application

In order to run RoKAI locally, a MATLAB environment is required. If you do not have one already, you can install MATLAB on your computer by following the installation steps.

Once the MATLAB is installed, you can use the RoKAI application as follows:

1. Download the source code from [Link1] or [Link2].
2. Extract the compressed '.tar.gz' file (e.g., by using 7-zip file manager).
3. Start the MATLAB runtime environment.
4. Run the provided '*demo_rokai.m'* script and make sure it runs without errors.
5. Store your phosphorylation data as a csv file into the "data" folder. Note that, as a protein identifier of a phosphosite, UniprotKb identifiers must be used. To map proteins another reference database, you can use Uniprot ID mapping tool.

   See "/src/data_preprocessing/load_sample_phospho_data.m' for an example on mapping from Ensembl protein (ENSP) identifiers to UniprotKB.
6. Modify the input path in '*demo_rokai.m'* script to load your csv file.

How to Cite

Please cite the following papers if using this app:

- Yılmaz S., Ayati M., Schlatzer D., Çiçek E., Chance M. R., Koyutürk M. (2020) Robust Inference of Kinase Activity Using Functional Networks. *Submitted for review.*
- Hornbeck, P. V. et al. (2015). Phosphositeplus, 2014: mutations, ptms and recalibrations. *Nucleic acids research*, 43(D1), D512–D520 [Link]
- Minguez, P. et al. (2012). Ptmcode: a database of known and predicted functional associations between post-translational modifications in proteins. *Nucleic acids research*, 41(D1), D306–D311 [Link]
- Szklarczyk, D. et al. (2014). String v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(D1), D447–D452 [Link]

Limitations

The RoKAI web application stores the input data temporarily. The data is removed once the website is reloaded or disconnected (i.e., once a session is terminated). Thus, please make sure to save your results before leaving the website. To conserve resources (i.e., server run time), the app terminates a session automatically after 5 minutes of inactivity. Note that, RoKAI web application only includes network data for human phospho-proteome. For other organisms, you can use the RoKAI desktop application by following the data pre-processing steps. Similarly, for sensitive and/or confidential data, we strongly encourage you to run RoKAI locally on your computer using the source code. This tool is intended for educational or academic purposes and comes with no warranty. See license for more information.

Contact

RoKAI is designed by Serhan Yilmaz and Mehmet Koyuturk, at Case Western Reserve University.

If you have any questions, please send an email to serhan.yilmaz@case.edu.

To report an issue, please use the Github issue reporting page.