

RoKAI App

Robust Inference of Kinase Activity

using network propagation on functional networks

User Manual

RoKAI v0.11

Table of Contents

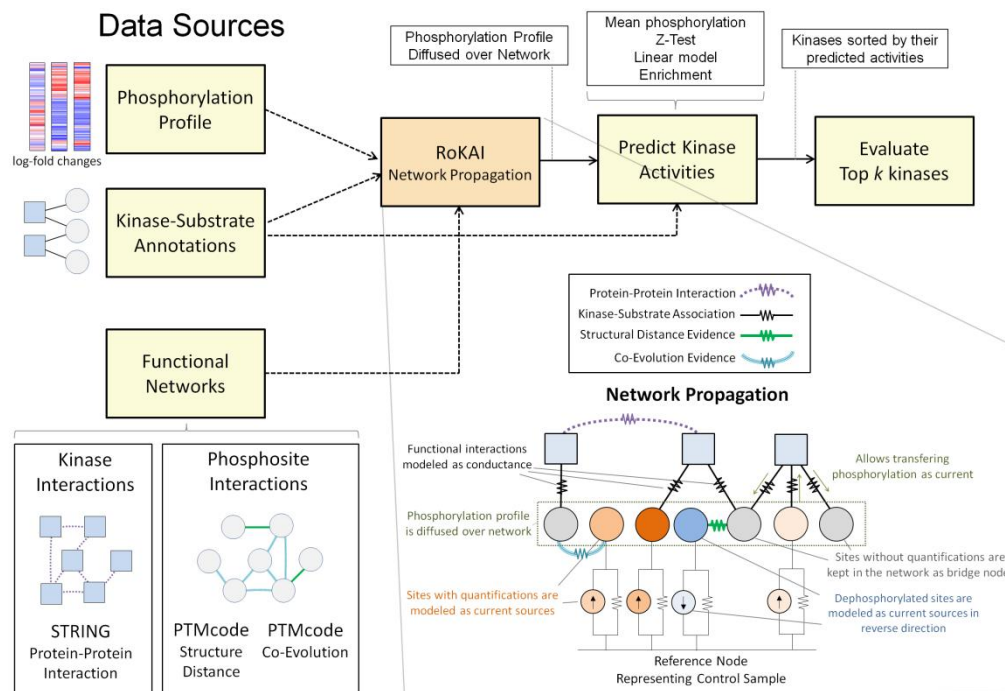
Introduction.....	1
Using RoKAI web application	2
Using RoKAI source code	4
Input data format	4
Output format	5
Code Schematic	6
Limitations	7
Contact.....	7
Acknowledgments	7
References.....	7

© 2020 Serhan Yilmaz

Case Western Reserve University

Email: serhan.yilmaz@case.edu

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 button on RoKAI web application. You can also download this file by clicking on the 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 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 "Results" tab.

Using RoKAI source code

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. To test on the sample data, run the provided '*demo_rokai.m*' script. It should run without errors if everything is correct up to this point.
5. To use RoKAI on your data, simply store your phosphorylation data as a csv file in the "data" folder (see below for input data format) and modify the input path in '*demo_rokai.m*' script to load your csv file.

Input data format

To run RoKAI on your data, all you need is a single csv file having phosphosites identifiers and phosphorylation value (as log2 fold change). Overall, RoKAI desktop applications accepts three types of phosphosite identifiers:

1. Ensembl protein identifiers and position

The sample in '*demo_rokai.m*' script uses this type of identifiers. In this case, you need an input csv file that has three columns:

- **Protein:** The Ensembl protein (ENSP) identifier
- **Position:** The position of the site on the protein
- **Quantification:** The phosphorylation of the site as log2 fold change

2. UniprotKB protein identifiers and position

Alternatively, you can use UniprotKB protein identifiers and position information to specify the phosphosites. See '*demo_rokai_uniprotkb.m*' script for an example on how to run RoKAI with this type of input data.

3. Flanking sequence

You can also use the +-7 flanking sequence to specify phosphosites. See the script '*demo_rokai_flankseq.m*' for an example that uses this type of data.

Output format

RoKAI provides two types of outputs which are (1) the inferred kinase activities and (2) the refined phosphorylation profiles after network propagation.

1. Inferred kinase activities

The first output of RoKAI is a table '*KinaseTable*' containing the inferred activity of kinases. Each row is a kinase and there are total of four columns:

- **KinaseID:** The UniProt accession id of the kinase
- **KinaseName:** The name/description of the kinase
- **Gene:** The gene symbol of the kinase
- **Activity:** The inferred activity of the kinase based on the phosphorylation of sites in their functional neighborhood (including their known targets).

Note that, the table is sorted by kinase activities in descending order.

2. Refined phosphorylation profile

The second output of RoKAI is a table '*SiteTable*' containing the refined phosphorylation profile after network propagation. These refined profiles are essentially aggregates of the phosphorylation of the sites themselves as well as the sites in their functional neighborhood. Thus, these are expected to capture coordinated changes in signaling which are potentially more robust and less noisy.

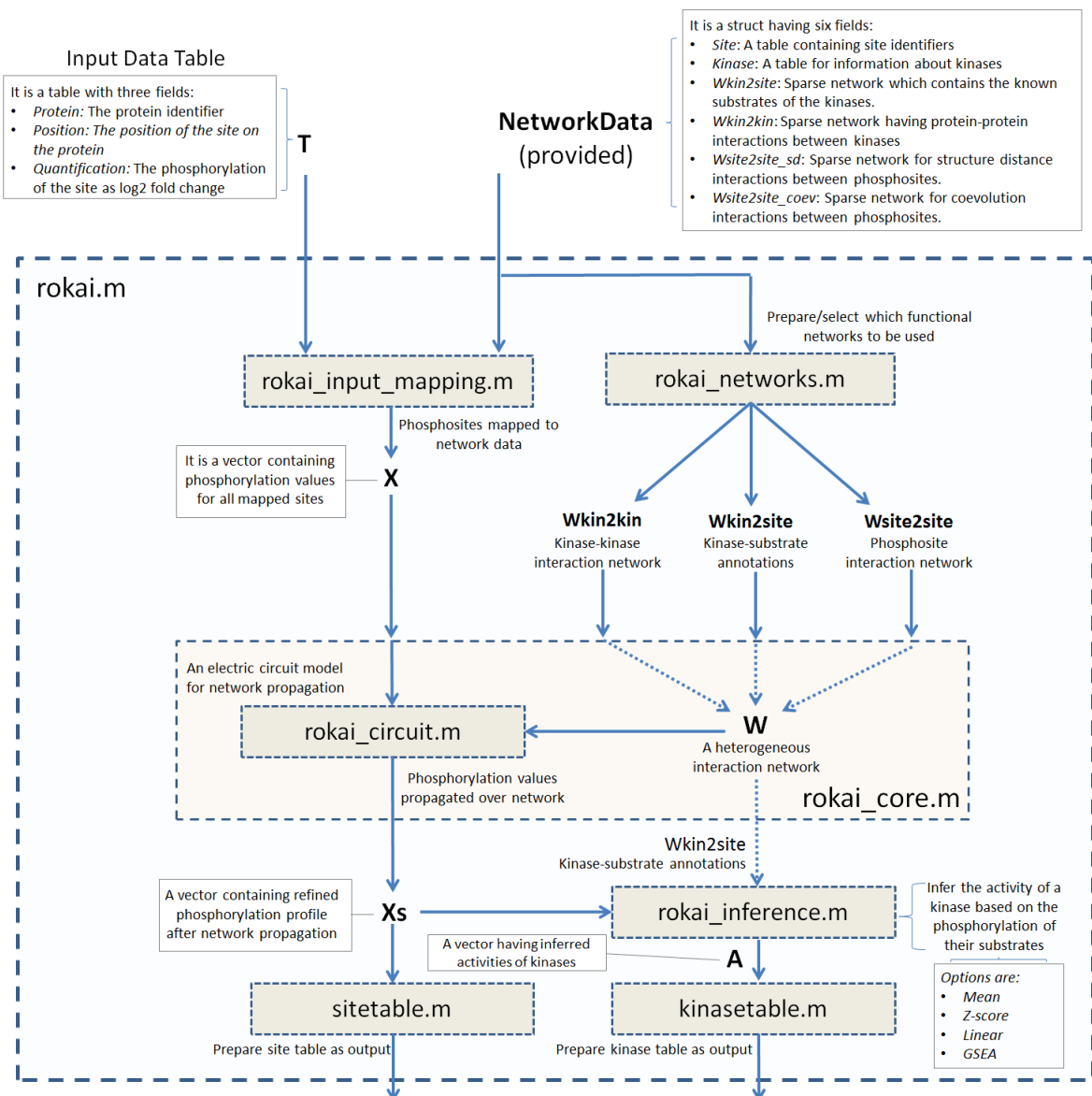
This table contains the following two columns (in addition to site identifiers):

- **Raw_Q:** The raw phosphorylation value (log2 fold change) of the site. These are equal to the input quantifications without any modification.
- **RoKAI_Q:** The refined phosphorylation value of the site after network propagation.

Note that, the table is sorted by the absolute values of the refined profile obtained by RoKAI. Thus, the sites on the top exhibit higher differences in phosphorylation compared to the control sample.

Code Schematic

The RoKAI web application and 'rokai.m' function in the source code provides an easy to use interface to run RoKAI with a predefined set of functional networks. For more advanced use cases that require an adaptation, you can refer to the schematic below for the code structure used:



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 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 or to suggest a new feature, please use the [Github page](#).

Acknowledgments

This work was supported by the National Library Of Medicine of the National Institutes of Health under Award Number [R01-LM012980](#). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Please cite the following paper(s) if you use RoKAI in your research:

- Yilmaz S., Ayati M., Schlatzer D., Çiçek E., Chance M. R., Koyutürk M. (2020) Robust Inference of Kinase Activity Using Functional Networks. *Preprint in bioRxiv* [[Link](#)].

Resources used for functional networks:

- Hornbeck, P. V. et al. (2015). Phosphositeplus, 2014: mutations, ptms and recalibrations. *Nucleic acids research*, 43(D1), D512–D520 [[Link](#)]
- Minguéz, 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](#)]