PhosphoPath: a Cytoscape plugin for network analysis on phosphoproteomics data

L.M. Raaijmakers, Piero Giansanti, Patricia A. Possik, Judith Mueller, Daniel S. Peeper, Albert J.R. Heck, A.F. Maarten Altelaar

June 9, 2015

Contents

1 Introduction

PhosphoPath is a Cytoscape app developed for the use with Cytoscape version 3. This manual shows how to use PhosphoPath to create a network in Cytoscape using phospho-proteomics data. It describes how to import the experimental data in Cytoscape and import information from PhosphoSitePlus, WikiPathways and Biogrid. As example files, we included a case study with quantitative information of proteins and phosphorylation sites. To follow this manual with the example files, you need to download the following two files: **Download from github**. Internet connection is required to retrieve the available data.

1. **dataset.phos** contains the uniprot ids and phosphorylation sites that will be used to create the network `https://github.com/linseyr/PhosphoPath/blob/master/Data/dataset.phos`

2. **regulation_2_tp.txt** contains the quantitative information for each of the quantified phosphorylation sites for two time points `https://`

github.com/linseyr/PhosphoPath/blob/master/Data/regulation_2_tp.txt

2 Usage

2.1 Installing the PhosphoPath app locally

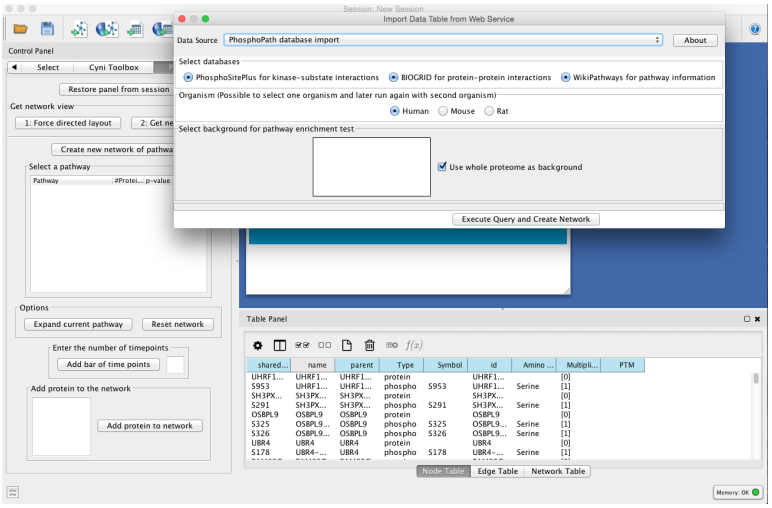
1. Download the PhosphoPath App Jar file from https://github.com/linseyr/PhosphoPath/blob/master/phosphoPath-0.0.1.jar
2. Start Cytoscape and install the PhosphoPath app via Apps → App manager. Click on 'Install from File' and browse for the jar file you just downloaded
3. Click 'Install' and the app will be installed in Cytoscape

2.2 Loading experimental data

Import the example data file 'dataset.phos' via File → Import → Network → File... The network will show up as a network called 'PhosphoPath-Network' with 721 nodes. Uniprot ids are automatically converted to gene names. This example file contains an extra column which represents the multiplicity (the number of phosphorylation sites measured on the quantified peptide). In this way it is possible to have multiple quantification data for one particular site. This column is optional.

2.3 Database import

Import database information from PhosphoSitePlus, Biogrid and Wikipathways via File → Import → Table → Public databases...
As Data Source select 'PhosphoPath database import' This step could take a while to perform. We are working on getting this step faster.



1. Select the 'PhosphoPath database import' option at Data Source
2. Select the data sources you want to retrieve information from and the organism (For example data use the default settings).
3. Select the background you want to use for the enrichment test. By default it is set to the whole proteome background of the selected organism. It is also possible to specify the background yourself.
4. Click the 'Execute Query and Create Network' button
5. Close the import panel
6. All identified pathways from WikiPathways are put together in the Pathway table in the 'PhosphoPath' pane in the Control Panel on the left.

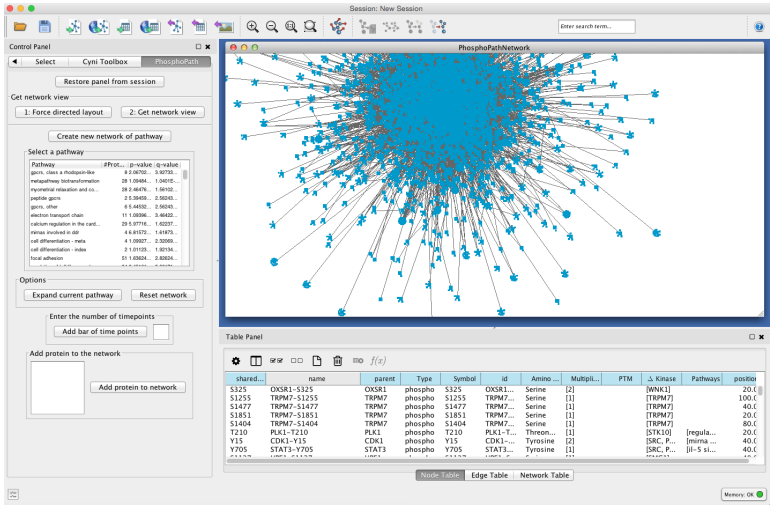
Information extracted from BioGrid about the type of interaction can be retrieved in the 'Edge table'

2.4 Create network view

Once the data is loaded, the network view can be created. Go to the 'PhosphoPath' pane in the 'Control panel' on the left. In this pane, all PhosphoPath tasks can be performed.

1. Click '1: Force directed layout', this will place the nodes according to the force directed layout algorithm of Cytoscape.
2. Click '2: Get network view', this will place the phosphosites on top of the protein node.

3. The protein-protein interactions from Biogrid are shown by a line, the kinase-substrate interactions from PhosphoSitePlus are shown by an arrow.

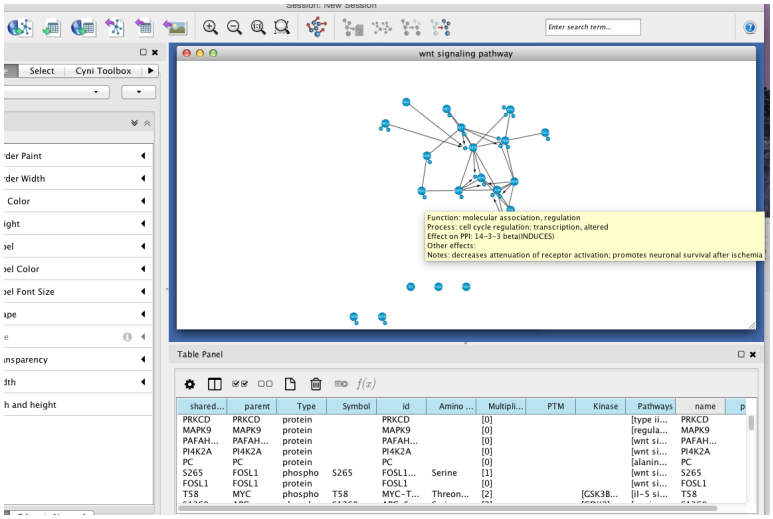


2.5 Create subnetwork for pathway

Also in the 'PhosphoPath' pane, all pathways for the proteins in the dataset are placed in the pathway table. These pathways are ordered according to their enrichment, with the most enriched pathways on top. The table can also be ordered in alphabetical order, by clicking on the 'Pathway' column header. It is possible to visualize the network for one specific pathway, or multiple of these pathways.

1. Select the pathways of interest from the table
2. Click the 'Create new network of pathway' button. The name of the newly created network will be the selected pathway's name.
3. Create the PhosphoPath network view by clicking the '2: Get network view' button.
4. Available information about a phosphorylation site can be shown by moving the mouse over the specific site. This information is extracted from the PhosphoSitePlus database.

The shape of the nodes can be modified in the 'Style' pane. Here we changed the shape of the nodes to 'ellipse'



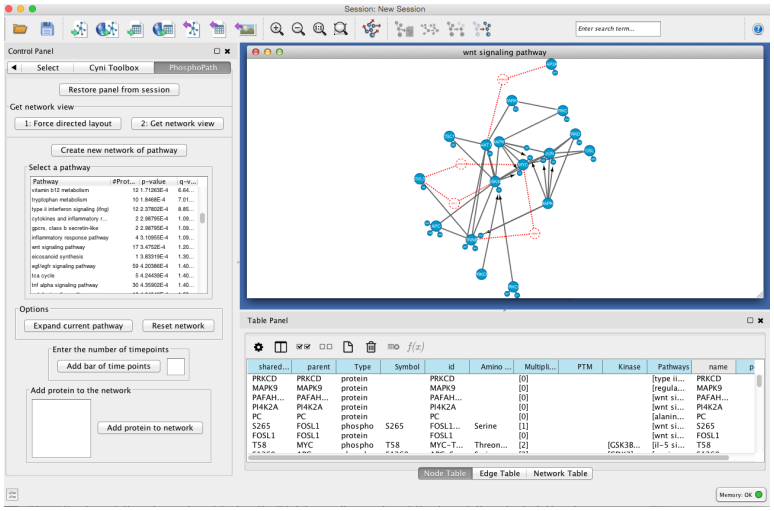
2.6 Expand network with database proteins

The current network can be expanded with proteins from the PhosphoSitePlus and Biogrid database. These proteins will be added if they follow these requirements;

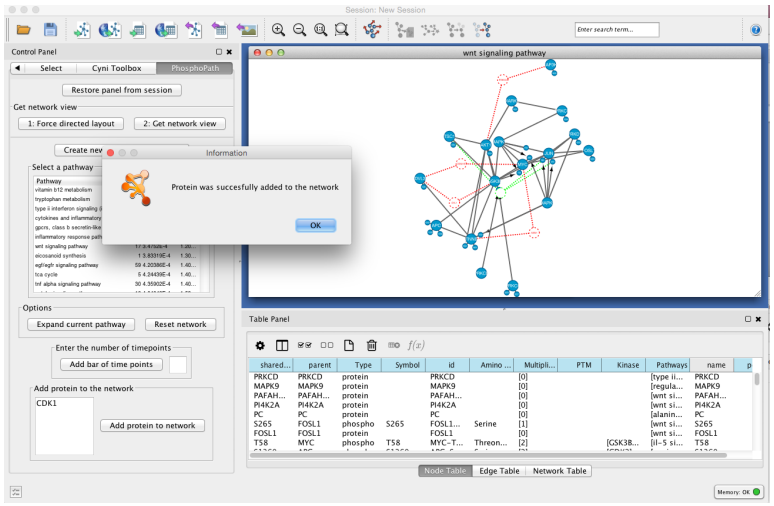
1. Connect unconnected proteins together
2. Are kinases of phosphorylation sites in the dataset
3. Are members of the selected pathway (if multiple pathways are selected, members of all pathways can be added)

If a protein is added to connect unconnected proteins together, and this protein is also a kinase of a phosphorylation site, it will also add an arrow describing the kinase-substrate interaction. The newly added proteins will be shown in red. Also the edge between the newly added proteins will be red.

1. Click the 'Expand current network' button
2. Click the '1: Force directed layout' button
3. Click the '2: Get network view' button



These three steps can be performed multiple times, each round adding more interconnecting proteins. The added proteins will be added to the node table, with as 'shared name' the round number in which it was added. The is also an option to place a user defined protein to the network (as gene name). This protein should be placed in the text box 'Add protein to the network'.



The added proteins can be removed again to get the original network back by clicking the 'Reset Network' button.

2.7 Adding quantitative information to the network

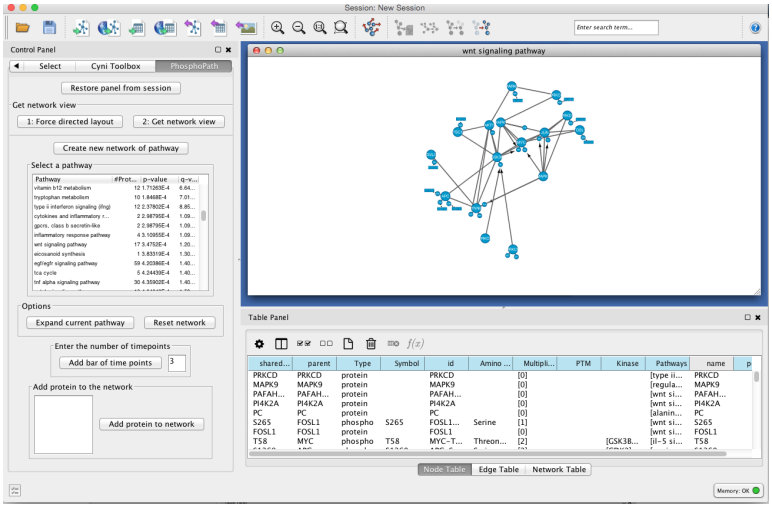
Nodes can be colored according to quantitative information. It is also possible to do this per time point. The attribute file downloaded from github

shows an example of how the attribute file should look like with multiplicities and time points.

When there is quantification data of 3 time points for phosphorylationsite Q13619-S10, the attribute file should contain entry; Q13619-S10-1, Q13619-S10-2, Q13619-S10-3. If the multiplicity is specified and the multiplicity is 1, the entries should be; Q13619-S10-1-1, Q13619-S10-1-2, Q13619-S10-1-3. For multiplicity 2 this should be; Q13619-S10-2-1, Q13619-S10-2-2, Q13619-S10-2-3. The easiest case is without any time point and multiplicity information. In that case it would just be Q13619-S10.

1. Select the nodes for which you want to add time point information

2. In the 'PhosphoPath' pane, enter the number of time points you want to add. This can be any number larger than 1 (If there is only one condition it is easier to just color the protein or phosphorylation node). In the example data there is data from 2 time points.



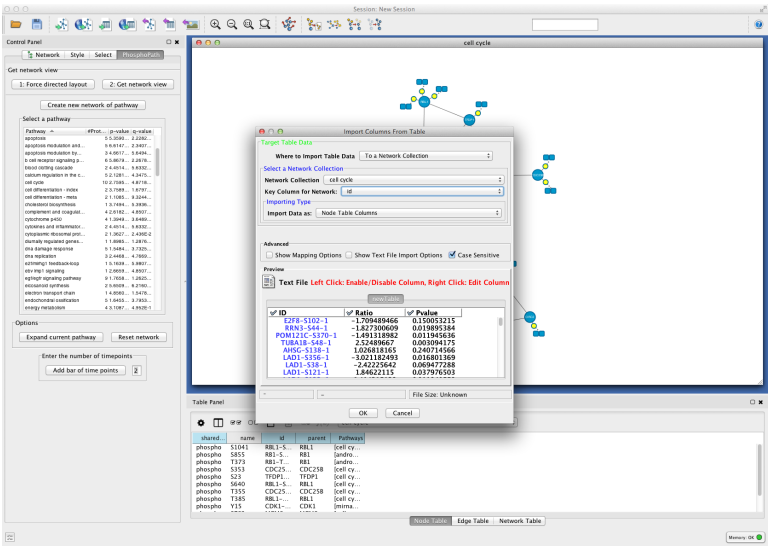
3. The attribute file can be loaded via File → Import → Table → File

4. Select the attribute file containing the quantitative information. The first column should contain the node id, followed by a dash and the time point nr ('node id-1,-2,-3.....'). The example attribute file 'regulation_2_tp.txt' shows the correct format

5. Select as 'Network Collection' the network you want to map the information to

6. As 'Key Column for Network' select 'id'. This is the column in the network which should match to the first column in the attribute file

7. Click 'OK' and the attribute data will be loaded

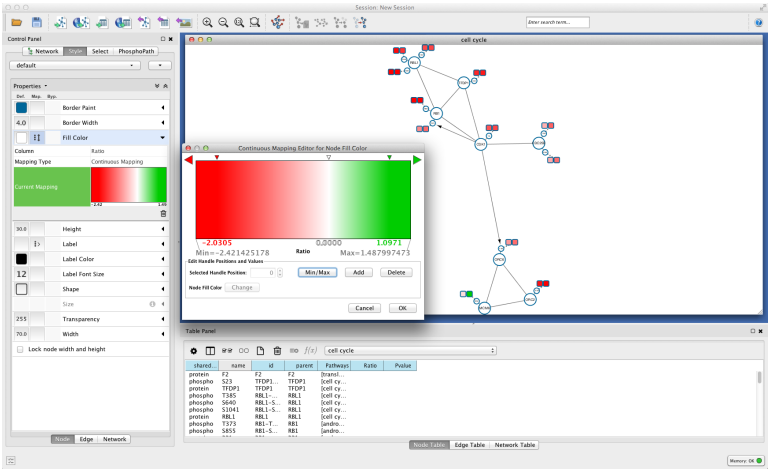


8. Once the attribute file is loaded, go to the ‘Style’ pane in the Control Panel

9. Go to ‘Fill Color’ and select the column you want to take the information from. In the example attribute file this is ‘Ratio’

10. Select ‘Continuous Mapping’ as mapping type and select the color scheme

11. The color will automatically be mapped to the matching nodes in the network



2.8 Changing the visual style of the network

The network visualization can be modified according to the user’s preferences. All options are available in the ’Style’ pane. Node border, shape or

color can be changed for phosphoryalation type (T, S, Y..) and the visual style for the edge can be modified to differentiate between interaction types (direct interaction, physical interaction..) or experiment (imaging, affinity chromatography..). But also the default settings of the network can be modified.