Network Biology

Abstract:

Introduction- Protein kinases are a group of enzymes that modify other proteins by adding a phosphate group to them in a process called phosphorylation. The aim of this practical was to visualise a kinase signalling network in Cytoscape using data downloaded from SIGNOR database, calculate graph theory metrics and analyse how biological disruptions affect the network.

Method- The human phosphorylation data was downloaded from SIGNOR data base as a 'tsv' format. It was then read into python as a panda's data frame. It was filtered to only include protein level interactions and phosphorylation mechanism. The data was filtered further to only include data on kinase, action, target protein and residue. The data frame was downloaded as a 'sif' file and opened in Cytoscape for visualisation. The network was analysed and a histogram showing degree distribution produced.

Another data frame was created, only including target proteins that were also kinases to reduce a cluttered look of the network. The new data frame was opened in Cytoscape to visualise the network and analyse it. To find out how the network would be affected in response to MAPK1 being unable to phosphorylate, a dataframe without MAPK1 in the kinase column was created. It was uploaded to Cytoscape and analysed again.

Results and discussion- The original dataframe looked very cluttered compared to the dataframe with only kinase as target proteins. It was necessary to reduce the cluttering as the original network was difficult to visualise due to many nodes overlapping.

When MAPK1 was unable to phosphorylate, it affected the network as 70 target proteins couldn't be phosphorylated, 134 outdegree connections were lost from the network. In addition, when a mutation of the 642nd residue of WEE1 changed from serine to glutamine, WEE1 was unable to be phosphorylated, leading to an

increased expression of WEE1 because phosphorylation downregulated this protein. Similarly, an AKT1 mutation from Tyr474 to Phe474 results in AKT1 not being able to be phosphorylated leads to a 55% inhibition of the AKT pathway.

The reason only some amino acid residues, usually serine, threonine and tyrosine can be phosphorylated is because they have a hydroxyl group present on their side, allowing a phosphate group to bind.

The overall degree distribution is very right skewed, showing an overall low degree network with a few high degree, critical nodes.

Conclusion- It was difficult to view human phosphorylation data in Cytoscape without filtering due to it being too cluttered looking. MAPK1 was one of the kinases with the largest out degree connections in the network, so when inhibited it the network lost many connections. Specific residue mutations in WEE1 and AKT1 caused inhibition of phosphorylation. The overall degree distribution was low with a few highly connected nodes.

Introduction:

Protein kinases are a group of enzymes that modify other proteins by covalently adding phosphates to them in a process called phosphorylation. The target proteins are phosphorylated at specific sites called residues, usually amino acids serine, tyrosine or threonine are phosphorylated. Protein kinases also phosphorylate each other, forming a cell signalling network.

The aim of this practical was to build a kinase signalling network using prior knowledge, visualise this network in Cytoscape, calculate graph theory metrics and see how the network can be disrupted by biological disruptions such as mutations. The human phosphorylation data was extracted from SIGNOR database, an easily accessible database containing lists of biological interactions available for download.

Method:

To begin with, the human phosphorylation data was downloaded from SIGNOR database under the ‘Downloads’ tab. Under ‘Available Downloads’ the phosphorylation data was downloaded in a ‘tsv’ format. This was then read into python as a panda’s data frame. The data was then filtered to only include interactions specified at protein level and then filtered further to only include phosphorylation mechanism. Some columns were removed, remaining were columns ‘ENTITYA’ which were the protein kinases, ‘MECHANISM’ which was the phosphorylation action, ‘ENTITYB’ which were the target proteins and ‘RESIDUE’, the site at which phosphorylation occurred.

For clarity, columns were renamed as the following:
ENTITYA: Kinase
MECHANISM: Action
ENTITYB: Target
RESIDUE: Residue

This data frame was then named ‘orignaldata.tsv’ and was downloaded as a ‘sif’ file and uploaded to Cytoscape (version 3.9.1) for visualisation. In Cytoscape, the network was analysed by going to Tools > Analyse Network, then ‘Analyse as Directed Graph?’ was checked. This provided graph theory metrics such as degree, indegree, outdegree, number of nodes and number of edges. To conjure a histogram showing degree distribution, ‘Node Degree Distribution’ was selected.

To reduce cluttering in the network, the data was updated to only include target proteins that were also kinases. This new data frame, ‘kinasematched.tsv’ was downloaded as a ‘sif’ file and uploaded to Cytoscape for visualisation. The network was analysed in the same way as above and a histogram showing degree distribution was produced.

To find out how the network may change in response to MAPK1 being unable to phosphorylate, MAPK1 was removed from the kinase column of dataframe ‘kinasedmatched.tsv’. This new data frame named ‘nomapk’ was then converted to a ‘sif’ file and viewed in Cytoscape to visualise differences between the network with

MAPK1 and the network without MAPK1. Once again, the network was analysed, and a degree distribution histogram was produced.

Results and discussion:

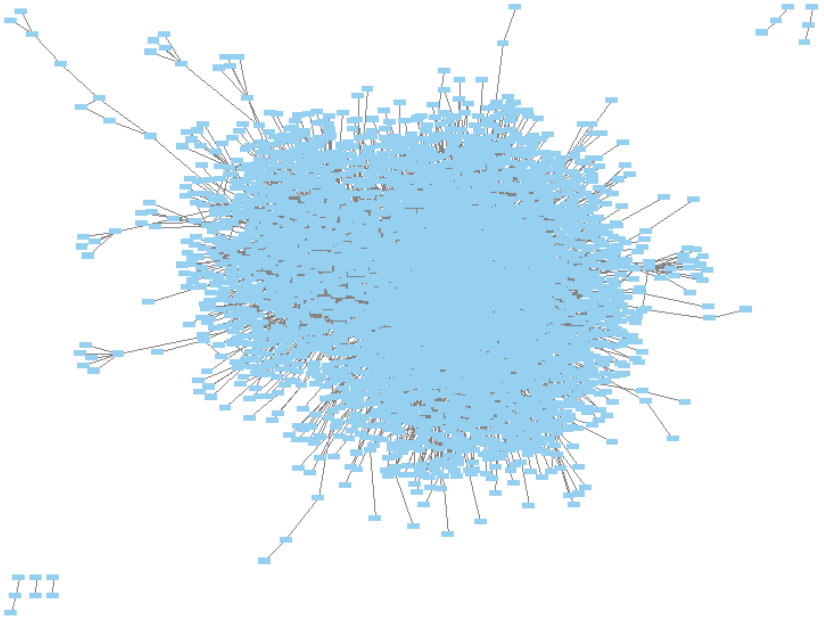

Figure 1- Signalling network visualised in Cytoscape from data frame ‘orignaldata.tsv’

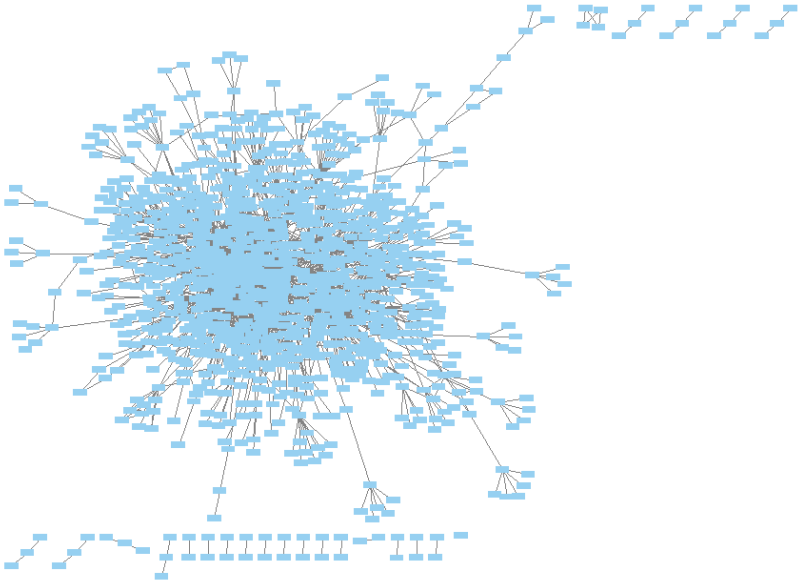

Figure 2- Signalling network visualised in Cytoscape from data frame ‘kinasedreduced.tsv’

Figure 1 is the original data containing all kinases, target proteins and residues from the human phosphorylation data downloaded from SIGNOR. It is very cluttered looking and there is lot of node overlap making it difficult to visualise the network and its connections. In figure 1 there are 4346 nodes and 17,096 edges. This is greatly reduced in figure 2 with only 1167 nodes and 3401 edges, by excluding target proteins that are not kinases.

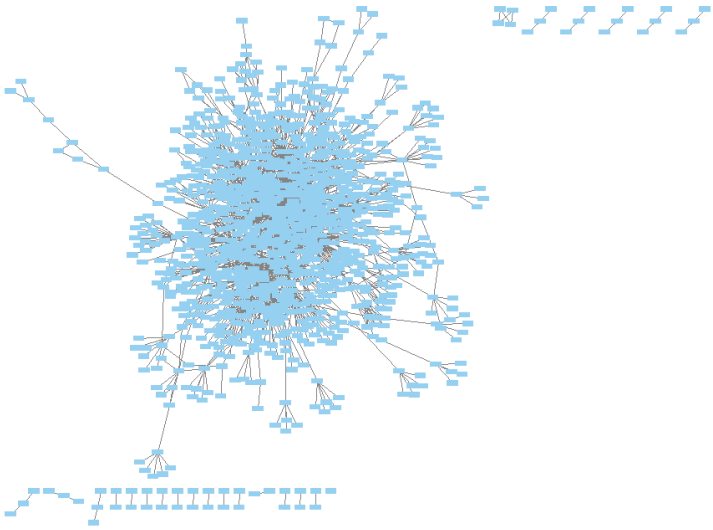

Figure 3- Signalling network visualised in Cytoscape from data frame ‘nomapk’.

MAPK1 is a part of the MAPK pathway and plays an important role in the MAPK cascade. This cascade is involved in many functions depending on cellular context, such as cell growth, adhesion, survival and differentiation through regulating transcription, translation and cytoskeletal translation. For this kinase to be activated, it needs to be phosphorylated at both residues Thr185 and Tyr187 (Uniprot, n.d.). A mutation causing inhibition of this kinase will lead to the inhibition of the MAPK cascade and therefore suppress cell proliferation, survival and apoptosis.

Figure 3 shows what the signalling network in figure 2 would look if a mutation were to occur and MAPK1 was no longer able to phosphorylate. The signalling network in figure 3 without MAPK1 kinases contain 134 less outdegree connections compared to figure 2, with MAPK1 kinases. In the figure 2 network, MAPK1 had the second highest number of outdegree connections of 134 and in figure 1 MAPK1 had the highest number of outdegree connections of 873 out of all kinases. Outdegree

connections are edges coming out of the node. In this instance there are 70 target proteins being phosphorylated by MAPK1 in the figure 2 network which are no longer being phosphorylated in figure 3.

Evidently MAPK1 is a very important kinase in the phosphorylation network, as it is one of the kinases with the greatest number of degree connections. Therefore, when it is inhibited and unable to phosphorylate, it greatly affects the network, and a lot of connections are lost and many target proteins remain unphosphorylated.

WEE1 plays a crucial role in mitotic regulation and DNA damage identification and repair. Its main role as a tumour suppressor is to prevent replication of cells with damaged DNA by preventing entry into mitosis to allow DNA repair. WEE1 activity increases in the S and G2 phase of mitosis and then it is hyperphosphorylated in the mitosis stage causing downregulations of WEE1 (Ghelli Luserna di Rorà et al., 2020).

When there is a mutation causing the 642nd residue of WEE1 to change from serine to glutamine, it affects the network as the glutamine cannot be phosphorylated. Normally, when BRSK1 phosphorylates WEE1 at Ser642, it leads to a down-regulation of WEE1 activity (Antibodies Online, n.d.). Therefore, there is an increased expression of WEE1 when it fails to be phosphorylated. Increased WEE1 expression has been implied in many types of cancers.

The reason why a residue change from Ser642 to Gln642 would result in WEE1 not being phosphorylated is because in eukaryotes, protein phosphorylation occurs usually on Serine, Threonine and Tyrosine. These amino acids have a hydroxyl group on their sidechain and so the protein kinase can bind a phosphate group to these amino acid side chains (Raju, 2019). Some other amino acids that can be phosphorylated post-translationally are arginine, lysine, aspartic acid, glutamic acid and cysteine.

Similarly, an AKT1 mutation at residue Tyr474 to phenylalanine abolishes phosphorylation. AKT1 is one of three kinases of the AKT kinase pathway. They are involved in processes such as metabolism, cell survival, cell growth, proliferation and

the formation of new blood vessels. Three specific residues, Thr308, Ser473 and Tyr474 need to be phosphorylated for full activation of this kinase (Uniprot, n.d.). Substitution at residue 474 from tyrosine to phenylalanine showed to prevent phosphorylation at this site therefore resulting in 55% inhibition of AKT pathway activation (Conus et al., 2002).

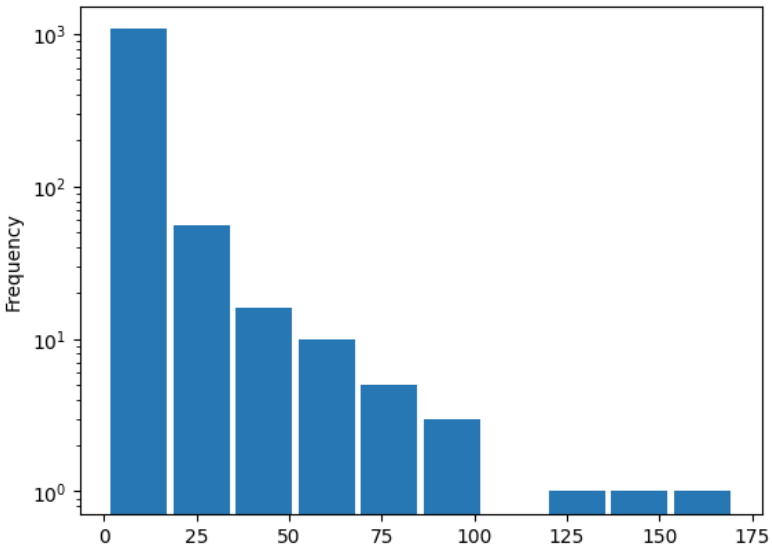

Figure 4 shows the degree distribution across the network using data frame ‘kinasereduced.tsv’. The x axis is showing degree and the y axis is showing frequency.

According to figure 4, there are many low degree nodes, with only a few nodes that are extremely connected, with a degree higher than 100. The few nodes with a degree past 100 are SRC, MAPK1 and PDPK1, these are critically important nodes as they have a significantly larger number of connections compared to the rest of the nodes. All 1164 other nodes, have a degree under 100. The data is right skewed with a peak left of the centre and a gradual tapering to the right of the graph as degree number increases. Overall, the degree of the entire network relatively low. This is like other biological networks as it is typical to find a lower degree network in which each node is connected to a subset of nodes rather than all nodes being highly connected to one another. Another example of a low degree network in biology is a food web detailing species within a community and their relationship to one another, rather than all species of the food web being connected to all other

species, these organisms are more connected to a few rather than being connected to many organisms in the food web.

Conclusion:

To conclude it is difficult to visualise the human phosphorylation data in its full form in Cytoscape due to it being so large and many of the nodes overlapping. It needed to be reduced to only include kinase target proteins. MAPK1 was one of the kinases with the most outdegree connections in the network and when a mutation occurred inhibiting this protein kinase from phosphorylating, it affected the network greatly as many connections were lost. Other mutations such as mutations in WEE1, residue Ser642 to Tyr642 and AKT1 residue Tyr474 to Phe474 also caused inhibition of phosphorylation, affecting the network and its connections. Generally, inhibition of phosphorylation depends on the type of amino acid mutation, as a handful of amino acids can be phosphorylated, whereas others cannot. The degree distribution of the network is right skewed with mostly low degree connections and a few critical highly connected nodes.

References:

Ghelli Luserna di Rorà, A. *et al.* (2020) "A Wee1 family business: Regulation of mitosis, cancer progression, and therapeutic target," *Journal of Hematology & Oncology*, 13(1). Available at: https://doi.org/10.1186/s13045-020-00959-2.

BRK1 Proteins (no date) *Antibodies Online*. Available at: https://www.antibodies-online.com/br/brsk1-58996/brsk1-proteins-32049/ (Accessed: December 25, 2022).

Raju, T.S. (2019) "Phosphorylation of proteins," *Co- and Post-Translational Modifications of Therapeutic Antibodies and Proteins*, pp. 163–175. Available at: https://doi.org/10.1002/9781119053354.ch13.

P31749 · AKT1_HUMAN (no date) *UniProt*. Available at: https://www.uniprot.org/uniprotkb/P31749/entry (Accessed: December 25, 2022).

Conus, N.M. *et al.* (2002) “Direct identification of tyrosine 474 as a regulatory phosphorylation site for the AKT protein kinase,” *Journal of Biological Chemistry*, 277(41), pp. 38021–38028. Available at: https://doi.org/10.1074/jbc.m203387200.

P28482 · MK01_HUMAN (no date) *UniProt*. Available at: https://www.uniprot.org/uniprotkb/P28482/entry (Accessed: December 25, 2022).