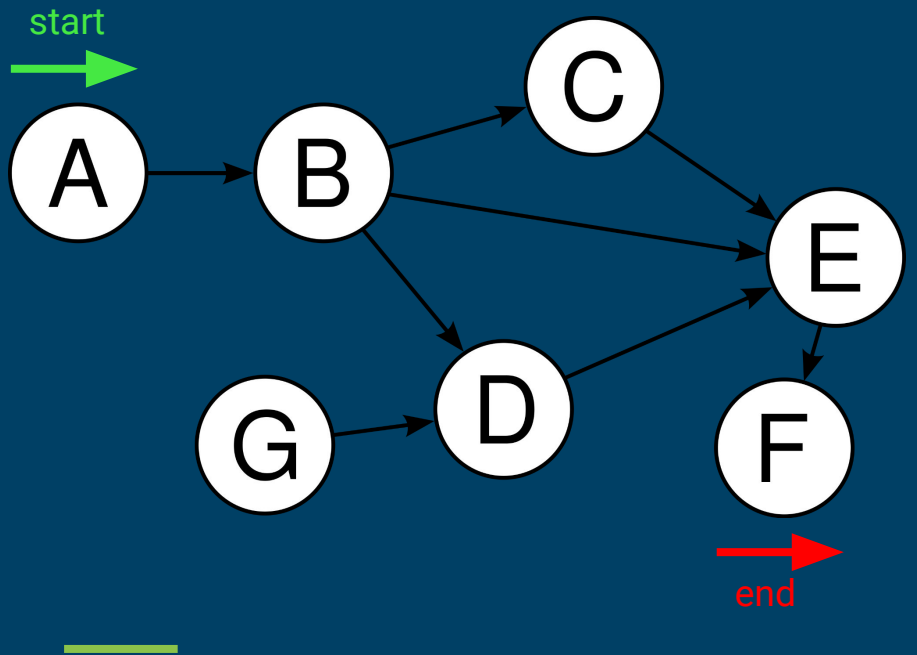
TBRU Computational Update

Shortest Path Network Analysis

What is the ‘Shortest Path’ ?

The shortest path is the list of nodes in a network that connect nodes A and F with the fewest number of edges *(or with the highest/lowest weighted score*).*

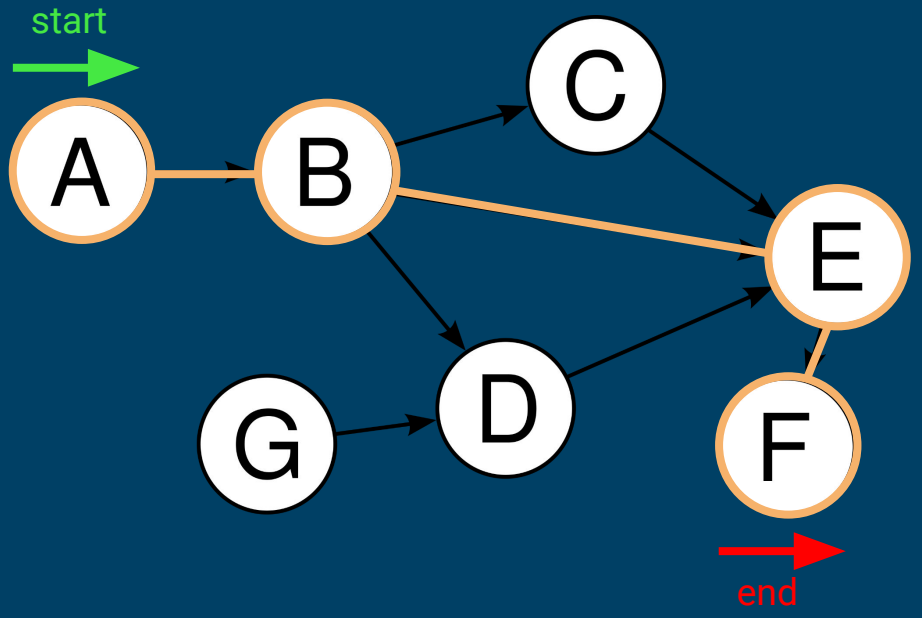
**weighted edges might be summed or multiplied*



What is the ‘Shortest Path’ ?

The shortest path is the list of nodes in a network that connect nodes A and F with the fewest number of edges *(or with the least cost if edges are weighted)*.

A, B, E, F

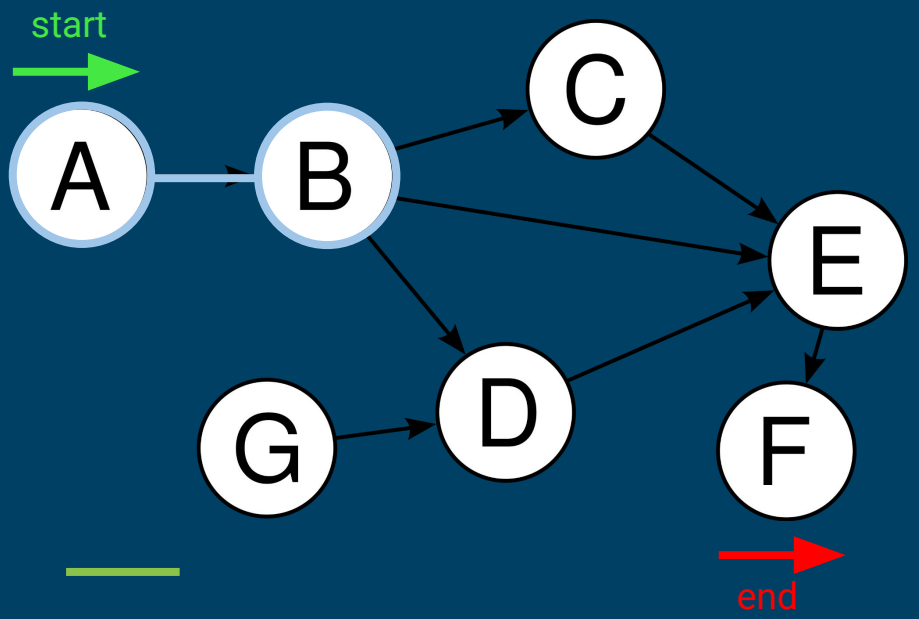


How to compute the shortest path?

In an unweighted graph, we'll use the **Breadth First Search** algorithm.

Beginning with the start node, iterate through every neighbor and track its distance from the start node.

Node	A	B	C	D	E	F	G
Dist	0	1	Inf	Inf	Inf	Inf	Inf
Prev	-	A	-	-	-	-	-

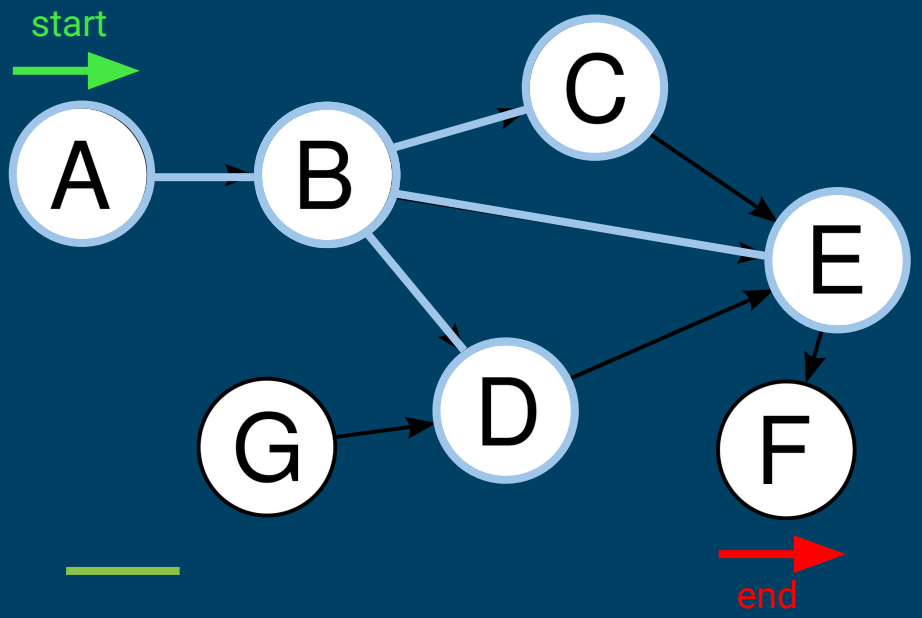


How to compute the shortest path?

In an unweighted graph, we'll use the **Breadth First Search** algorithm.

Record the distance of each node by adding 1 to the distance of the previous node.

Node	A	B	C	D	E	F	G
Dist	0	1	2	2	2	Inf	Inf
Prev	-	A	B	B	B	-	-

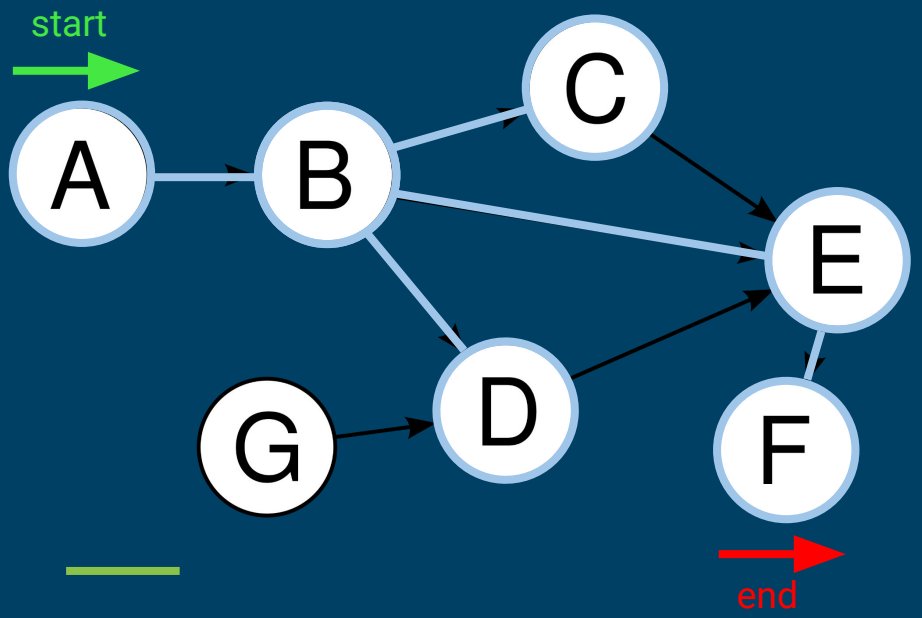


How to compute the shortest path?

In an unweighted graph, we'll use the **Breadth First Search** algorithm.

Record the distance of each node by adding 1 to the distance of the previous node.

Node	A	B	C	D	E	F	G
Dist	0	1	2	2	2	3	Inf
Prev	-	A	B	B	B	E	-

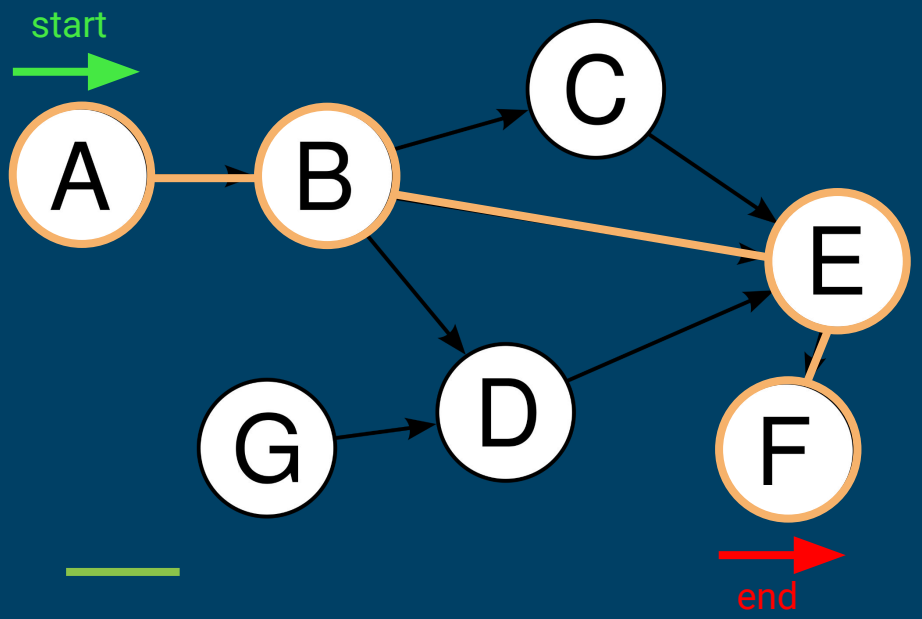


How to compute the shortest path?

In an unweighted graph, this is accomplished using the “breadth first search” algorithm.

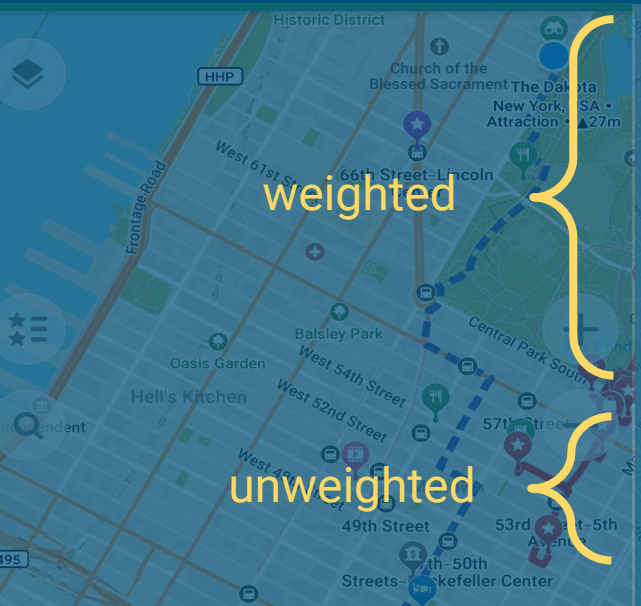
Working backwards from the end node, retrace the nodes traversed.

Node	A	B	C	D	E	F	G
Dist	0	1	2	2	2	3	Inf
Prev	-	A	B	B	B	E	-



Shortest Path Examples & Use Cases

Imagine an app that gives directions–in its network model, every intersection is a node and every stretch of road is an edge. The shortest path might be:

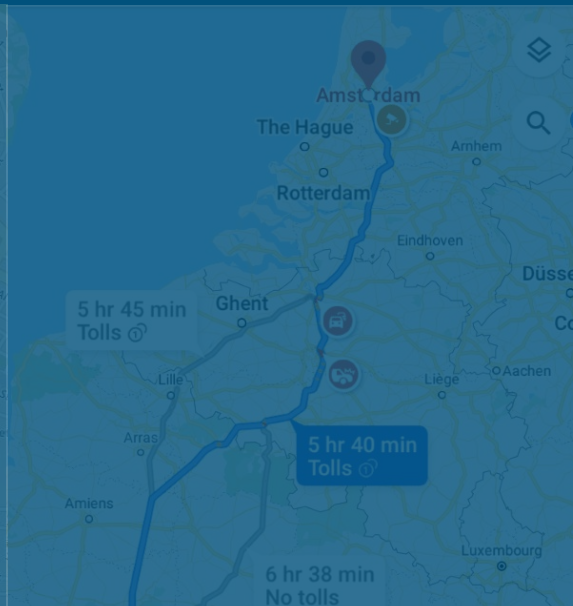


The shortest length of road travelled,

The roads that can be travelled in the least time,

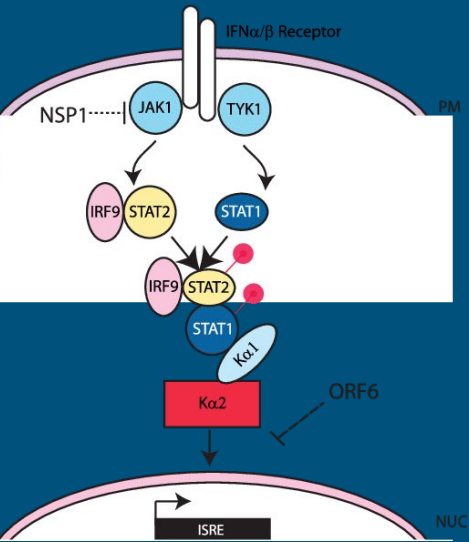
The roads that can be travelled by burning the least amount of fuel,

The roads that can be travelled to make the fewest number of turns, etc...



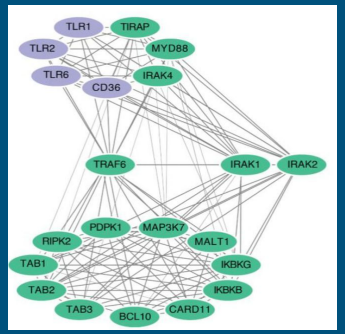
Shortest Path Examples & Use Cases

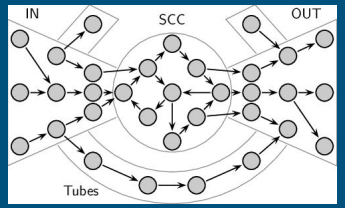
In Biology, shortest paths are a useful tool for modelling **signalling pathways** and other cascading cellular interactions… In a biological network, every gene is a node and every protein-protein interaction might be an edge.



A signalling pathway starts with receptors and ends with effector proteins; i.e. a transcription factor or ion channel.

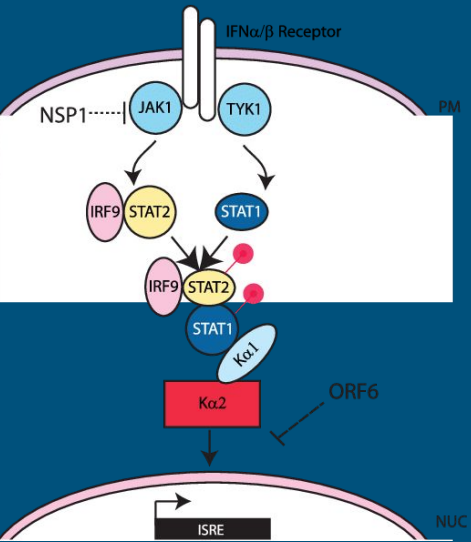
Critical genes in the signalling pathway might form “bowties” where many upstream and downstream nodes converge on a small number of nodes in the pathway.





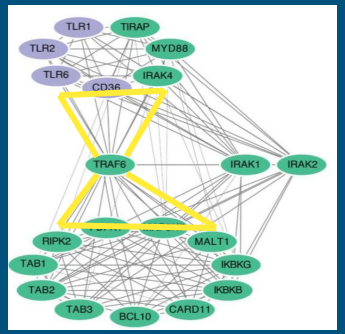
Shortest Path Examples & Use Cases

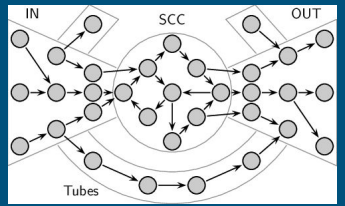
In Biology, shortest paths are a useful tool for modelling **signalling pathways** and other cascading cellular interactions… In a biological network, every gene is a node and every protein-protein interaction might be an edge.



A signalling pathway starts with receptors and ends with effector proteins; i.e. a transcription factor or ion channel.

Critical genes in the signalling pathway might form “bowties” where many upstream and downstream nodes converge on a small number of nodes in the pathway.





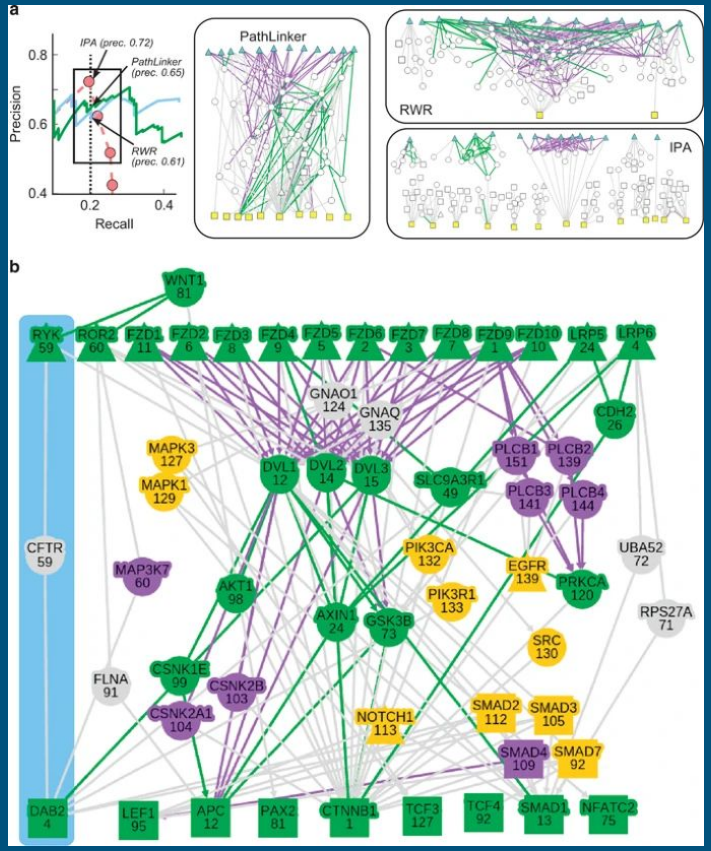
Shortest Path Examples & Use Cases

In Biology, shortest paths are a useful tool for modelling **signalling pathways** and other cascading cellular interactions… In a biological network, every gene is a node and every protein-protein interaction might be an edge.

The underlying network is critically important—what edges are used & how are they weighted?

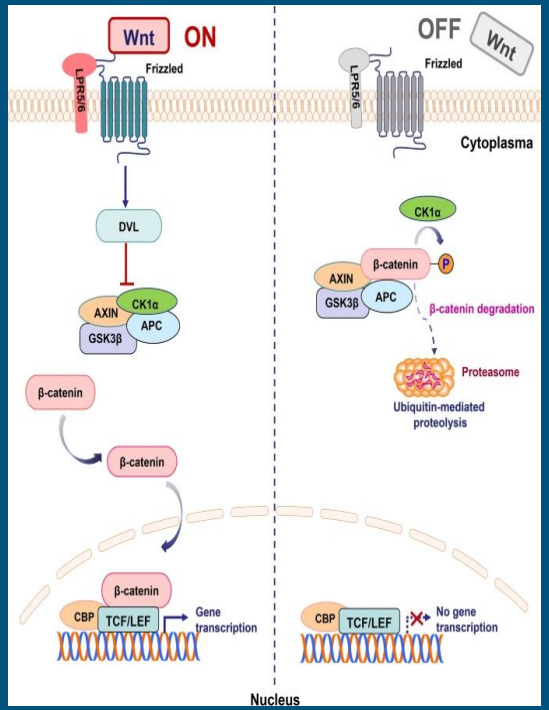
Biological use-cases often have multiple starts and ends, and therefore multiple paths, some algorithms calculate and optimize k-shortest paths.

Are we interested in the intersection or the union of the paths connecting them?

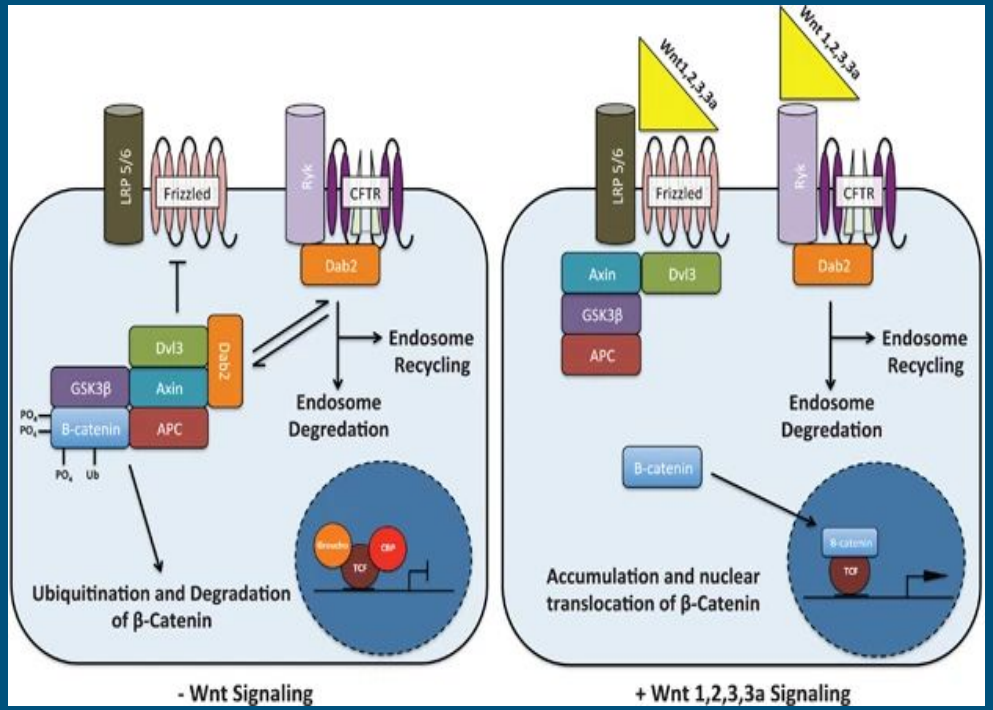


Pathways on Demand: Automated Reconstruction of Human Signalling Networks
Ritz et al. 2016

Old



New



Pathways on Demand: Automated Reconstruction of Human Signalling Networks
Ritz et al. 2016

First Approach:
Simple Union of Unweighted Shortest Paths.

Concept

TB secreted factor – host protein interaction data (PPI) and CRISPR KO/KD immune screen data (Crispr) both have particular mechanistic significance.

PPI hits represent the beginning of TB disease pathology in the host cell and Crispr hits are likely drivers of macrophage intra- and intercellular immune phenotypes (maybe ultimately responsible for the resistor phenotype).

Pathways in the PPI network connecting these 2 groups of genes could reveal the mechanisms of macrophage innate immune response to TB infection.

Creating the subnetwork.

Started with STRING Human interactome data.

- Limited edges to those with cumulative score > 600. [20,000 nodes, 800,000 edges]
- Limited nodes to Mehdi's network propagation results. [5000 nodes, 180,000 edges]
- Reconstruct network from union of shortest paths. [600 nodes, 800 edges]

Implementation.

1.) Prepare significant PPI and Crispr hits and retrieve Mehdi's TB subnetwork.

2.) Calculate unweighted network distance between every node.

3.) For each Crispr hit, lookup PPI hit(s) with the shortest distance; then find the shortest path(s) from Crispr hit to PPI hit(s).

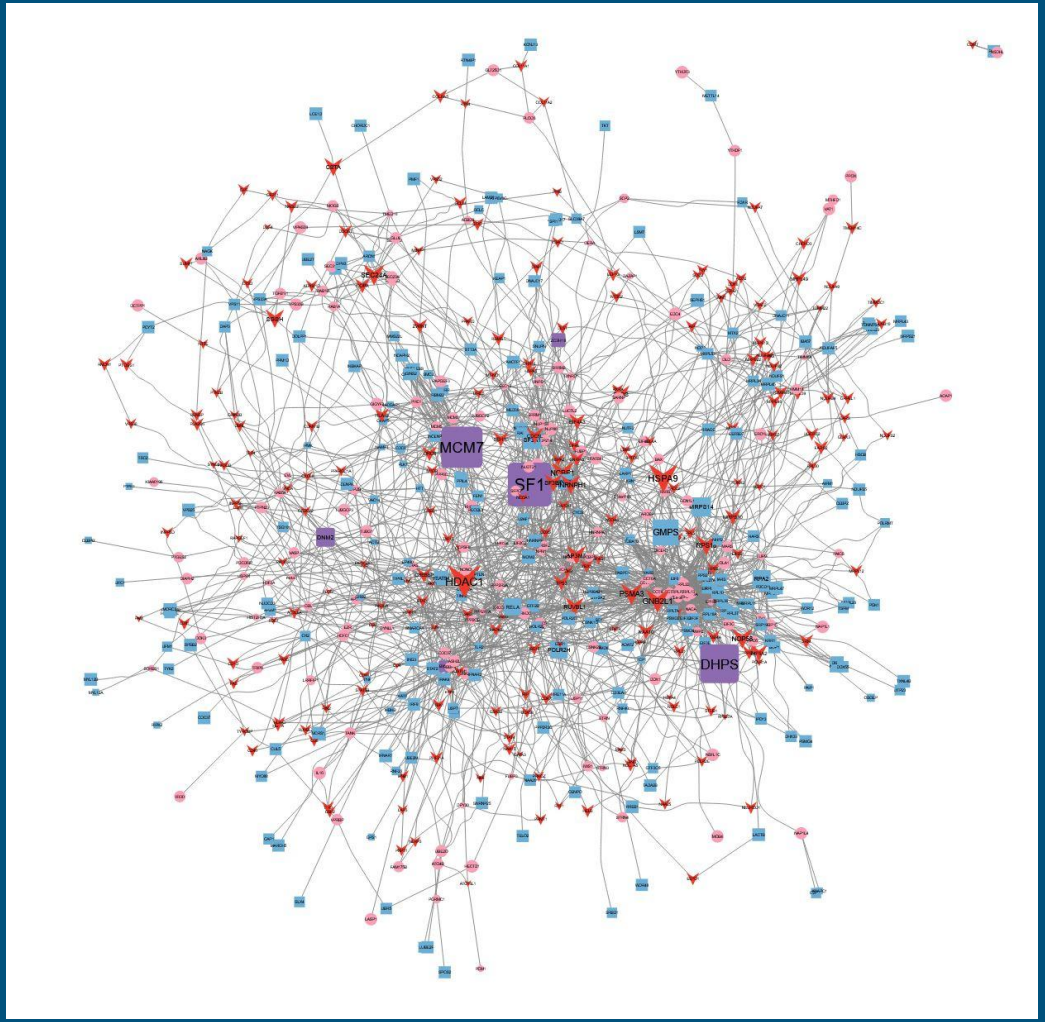
4.) Simplify the union of all edges in those shortest paths and then calculate centrality measures for each node from the new network.

5.) Normalize nodes' centrality measures by generating 1000 shortest paths subnetworks using the above protocol with random sources and targets.

vs. All KD Network

Network of shortest paths from each Crispr KD/KO to the closest PPI hit.

- Node size = normalized betweenness centrality
- Node shape & color based on source/target/pathway
- Force-directed layout shows network density



What Genes Have the Highest Centrality?

NEW FROM SHORTEST PATH ANALYSIS:

HDAC1 - histone deacetylase

HSPA9 - heatshock response protein involved in coordinating stress response all over the cell, esp. in mitochondria

RACK1 - ribosomal subunit/intracellular receptor for C kinase 1 responsible for rejecting bad mrna and controlling rate of protein translation

NCBP1 - mrna cap binding protein involved in rna splicing, rna-gene-silencing, and rna stabilization/decay

SEC24a - vesicle coat protein involved in directing vesicular transport and vesicle fusion. Related coat proteins also came out as interactors of TB secreted proteins--might be important part of autophagocytosis interaction

What Genes Have the Highest Centrality?

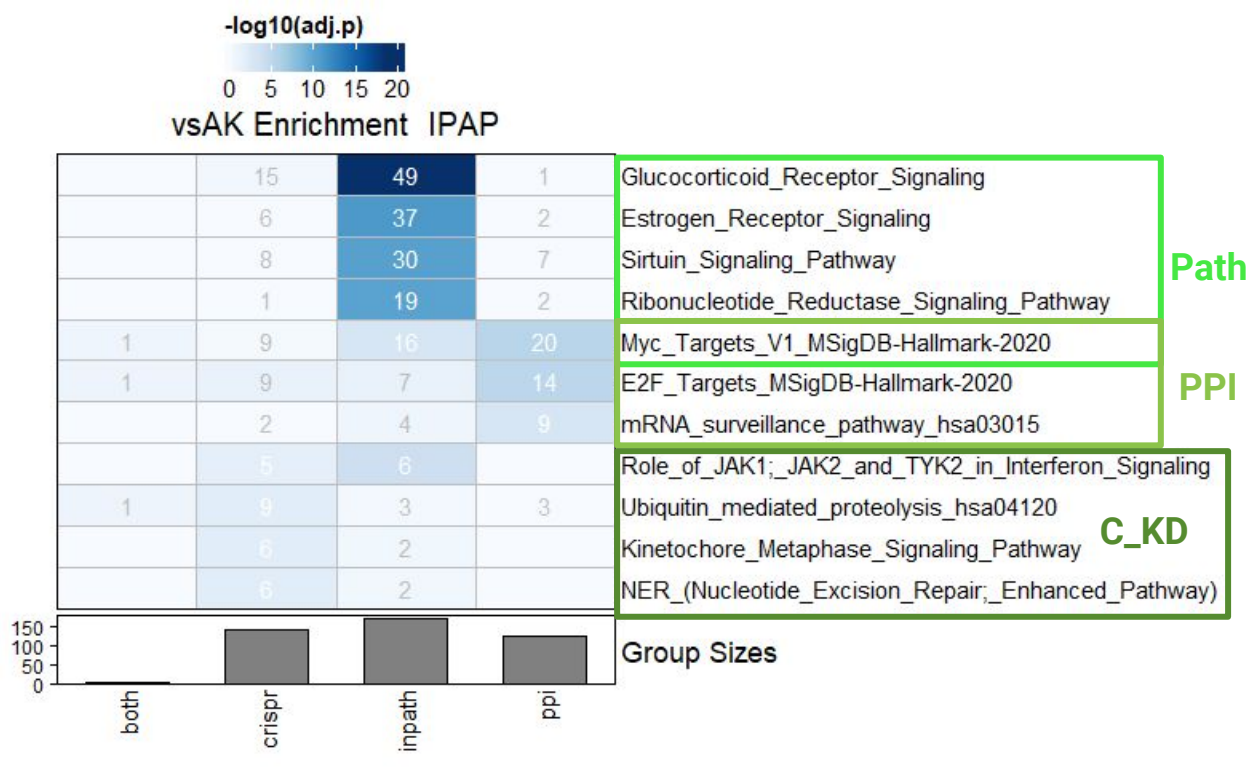
IN BOTH PPI AND CRISPR:

SF1 - Zn finger protein that binds mRNA and is involved in splicing and stabilization

MCM7 - part of cell cycle initiating complex that prepares chromosome for dna replication

DHPS - activates EIF-5a, a translation initiation factor

IPA Enrichment Results



Future Approaches:

- k-shortest paths (look to PathLinker)
- Modularize using dsd
- Weight network (using clinical datasets?)
- Add TB-baits to network (group crispr hits by closest bait, find shortest paths from meningitis-assc genes to validate clinical data)
- Build a directed, logical network