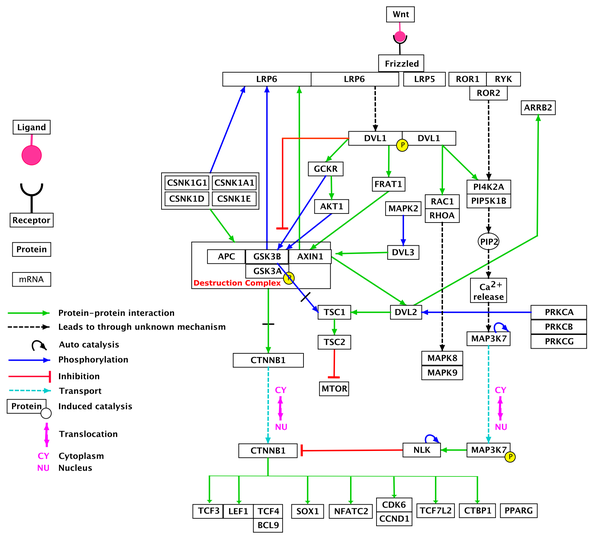
Integrating Protein Localization Information in Signaling Pathway Reconstructions

Ibrahim Youssef[1,2] and Anna Ritz[1]

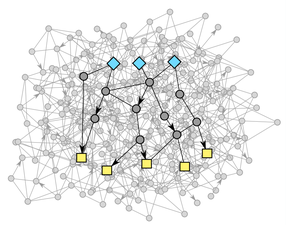
[1]Biology Department, Reed College, OR, USA; [2]Biomedical Engineering Department, Cairo University, Egypt

1. Signaling Pathway Reconstructions

Databases such as NetPath [1] provide annotated signaling pathways; however the manual curation of these databases is tedious and pathways are incomplete.

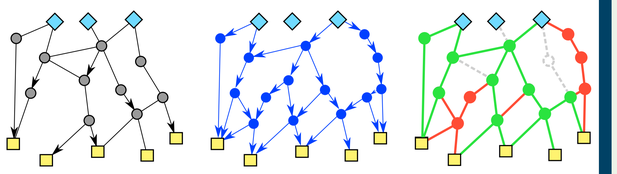


Suppose we consider the receptors and transcriptional regulators (TRs) of a pathway of interest that is embedded within a much larger interactome (12K nodes and 152K directed edges). A *pathway reconstruction* is a subnetwork that connects the receptors (blue diamonds) to the TRs (yellow squares).



Evaluating Pathway Reconstructions

We calculate true and false positives by comparing a pathway reconstruction to the curated NetPath pathway.



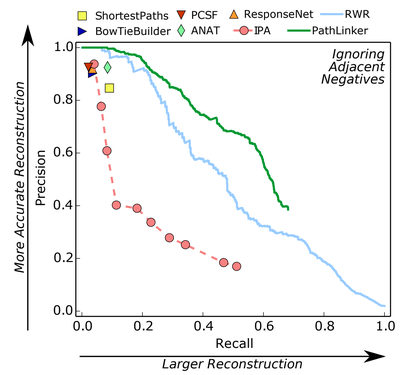
Curated Pathway | Reconstruction | True & False Positives

Here, we incorporate protein localization information to improve signaling pathway reconstructions.

2. PathLinker [2]

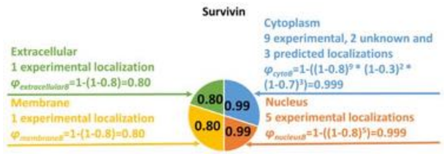
PathLinker computes the k shortest paths from any receptor to any TR in the interactome. Pathway reconstructions can "grow" by the number of shortest paths k, and the paths ensure that the interactions in the reconstruction remain "close" to the receptors and TRs.

Reconstructions Aggregated Over 15 Pathways



3. Protein Localization with ComPPI [3]

ComPPI integrates protein interactions and compartment information to compute confidence scores of protein localization and interactions.



Example compartment probabilities (adapted from Fig. 2a in [3]).

PathLinker performance improves at low levels of recall when the interactome is constrained to proteins that have compartment information (blue curves) compared to the original (orange curves).

We use the probabilities that proteins appear in each compartment. An edge (u, v) in a path is *consistent* with respect to localization if the following is true for some threshold t:

$\varphi_{membrane}(u) \times \varphi_{cytosol}(v) \geq t$ if (u, v) is the first edge in the path

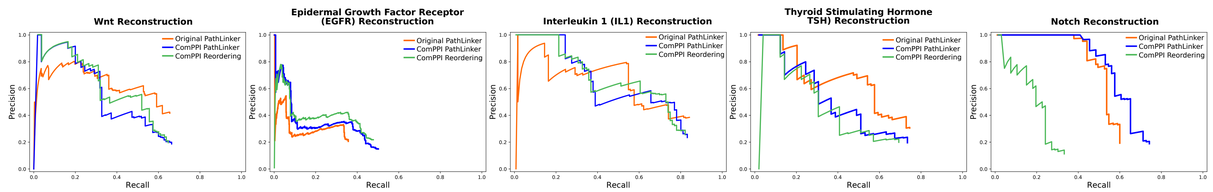
$\varphi_{cytosol}(u) \times \varphi_{cytosol}(v) \geq t$ if (u, v) is an internal edge in the path

$\varphi_{cytosol}(u) \times \varphi_{nucleus}(v) \geq t$ if (u, v) is the last edge in the path.

After running PathLinker, we identify all the consistent interactions within the paths for $t = 0.8$. We reorder paths according to the proportion of consistent interactions.

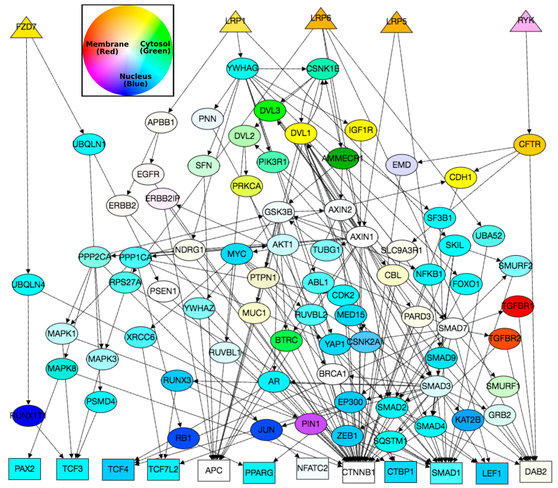
In some cases, we show improvement at intermediate levels of recall (green curves) if we prioritize paths where at least half of the edges are consistent.

Performance suffers for pathways with nuclear receptors (e.g. hormone signaling) or receptors with an intracellular domain (e.g., the Notch signaling pathway).



4. Wnt Reconstruction by Compartment

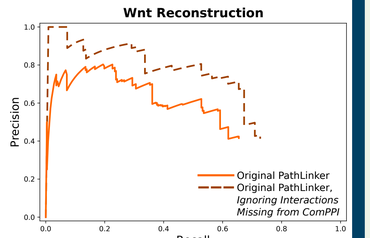
Top 200 paths computed by PathLinker [2] on the interactome limited to proteins with compartment information from ComPPI [3].



5. Conclusions and Future Work

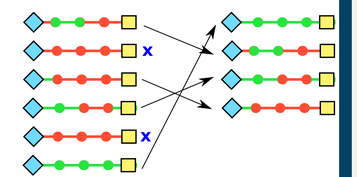
Protein localization information improves pathway reconstruction performance by reducing the false positives from the original interactome.

If we consider the *original* PathLinker rankings, in many cases we improve the performance if we ignore any interactions with proteins that do not have ComPPI information from the original PathLinker rankings.



Ongoing work incorporates localization probabilities to identify paths indicative of intracellular signaling.

Our goal is to reprioritize paths by developing a Hidden Markov Model (HMM) or other probabilistic model to prioritize likely paths. We are also incorporating gene expression data from healthy and diseased tissues.



Acknowledgements

Thanks to T. M. Murali at Virginia Tech and to the other GraphSpace [4] co-authors. The Wnt reconstruction is available at http://graphspace.org/graphs/23672?user_layout=2051.

M.J. Murdock CHARITABLE TRUST

Contact Information

Web: http://www.reed.edu/biology/ritz
Email: aritz@reed.edu & youssefi@reed.edu

REED COLLEGE

[1] K. Kandasamy *et al.*, "NetPath: a public resource of curated signal transduction pathways," *Genome Biology*, vol. 11, p. R3, Jan 2010.

[2] A. Ritz *et al.*, "Pathways on demand: automated reconstruction of human signaling networks," *NPJ Systems Biology and Applications*, vol. 2, p. 16002, 2016.

[3] D. V. Veres *et al.*, "Comppi: a cellular compartment-specific database for protein–protein interaction network analysis," *Nucleic Acids Research*, vol. 43, no. D1, pp. D485–D493, 2014.

[4] A. Bharadwaj *et al.*, "Graphspace: stimulating interdisciplinary collaborations in network biology," *Bioinformatics*, 2017.