Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data

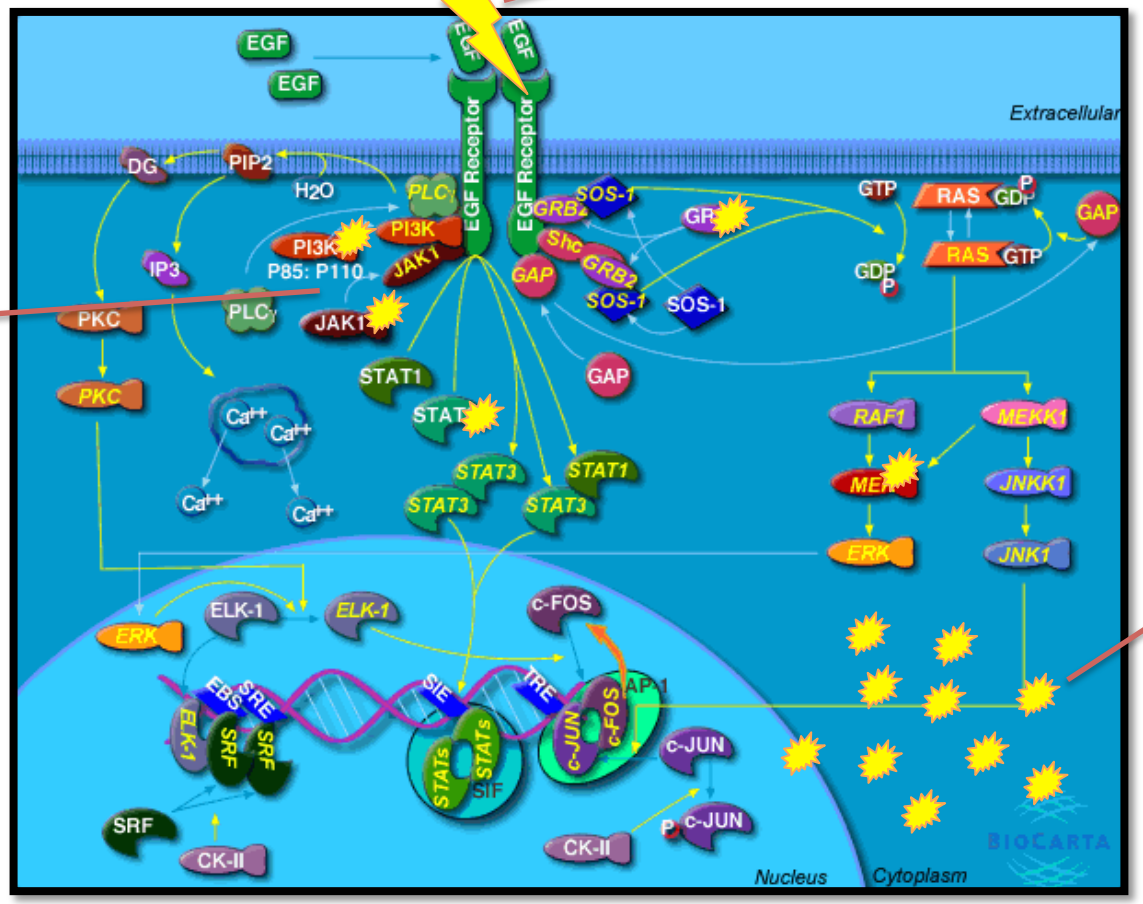
Ali Sinan Köksal[1], Anthony Gitter[2], Kirsten Beck[3], Aaron McKenna[3], Saurabh Srivastava[4], Nir Piterman[5], Rastislav Bodík[1], Alejandro Wolf-Yadlin[3], Ernest Fraenkel[6], Jasmin Fisher[7]

[1]University of California, Berkeley, [2]University of Wisconsin-Madison, [3]University of Washington, [4]20n, [5]University of Leicester, [6]Massachusetts Institute of Technology,[7]Microsoft Research Cambridge

Beyond Pathway Maps

Stimulate receptor

31% of pathway is activated

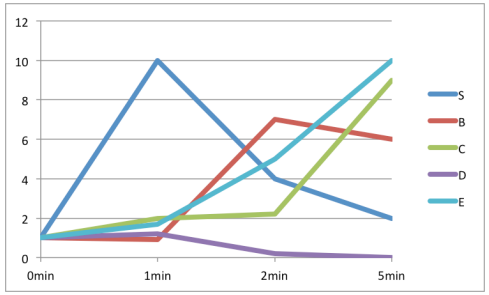


98% of activity is not covered

BioCarta EGF Signaling Pathway

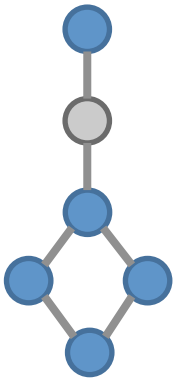
http://www.biocarta.com/pathfiles/h_egfpathway.asp

Data for Inferring Specific Pathways



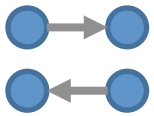
Temporal phosphorylation

- Global response to receptor stimulus
- Not all activity is phosphorylation
- Irrelevant/spurious phosphorylation



Undirected network topology

- Sparse, high-confidence connections
- Obtained by methods such as PCSF
- No temporal precedence knowledge

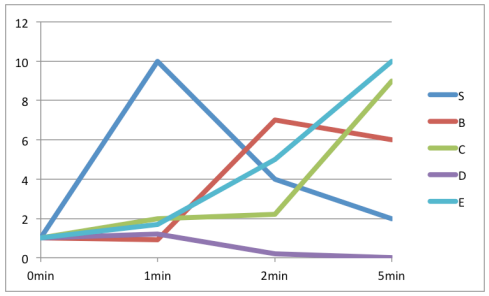


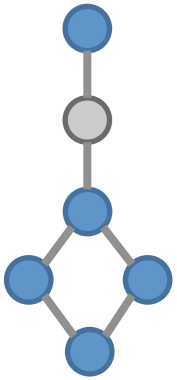
Prior knowledge

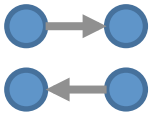
- Directed kinase-substrate interactions

Inferring Network Models

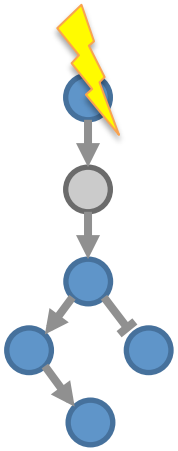
input







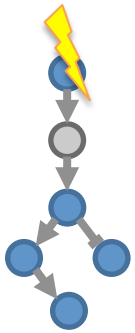
output



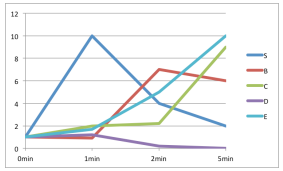
- Condition-specific pathway maps
- Signed directed graphs
- Signaling event timing annotations

From Data to Constraints

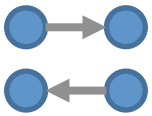
Explore **all** signed directed graphs that satisfy:



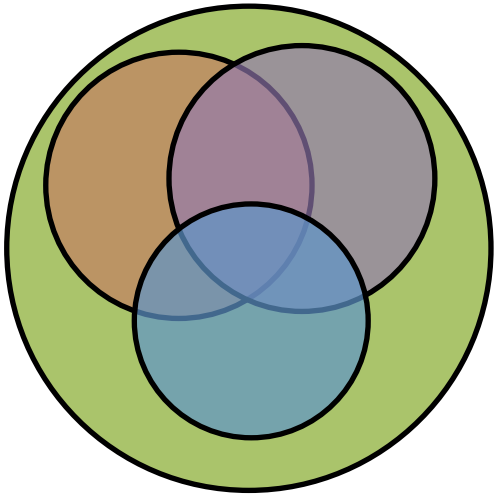
Topological constraints
All interactions must
originate from the source



Temporal constraints
Sequences of interactions must
agree with temporal precedence



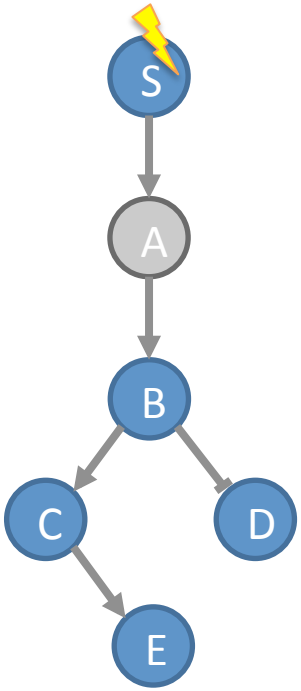
Prior knowledge constraints
Inferred networks may not
violate known directions



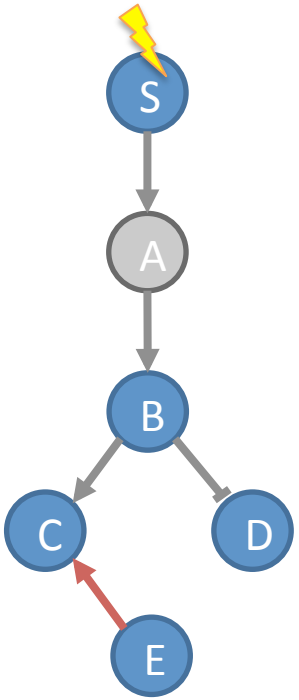
All network models

Topological Constraints

All interactions must originate from the source.



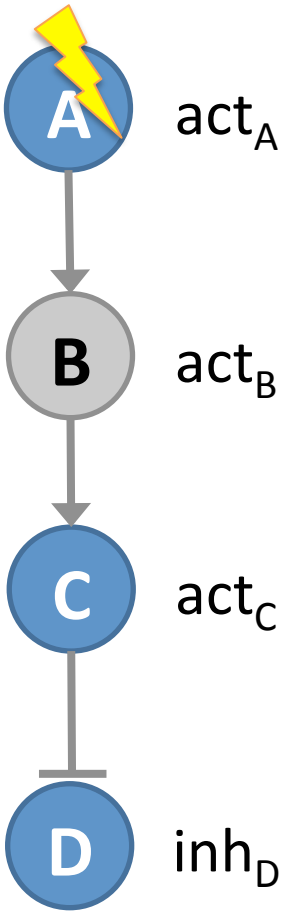
Valid model
All interactions are on a path from the source.



Invalid model
E activates C but is not reachable from the source.

Temporal Constraints

Interaction paths **consistent** with temporal data:

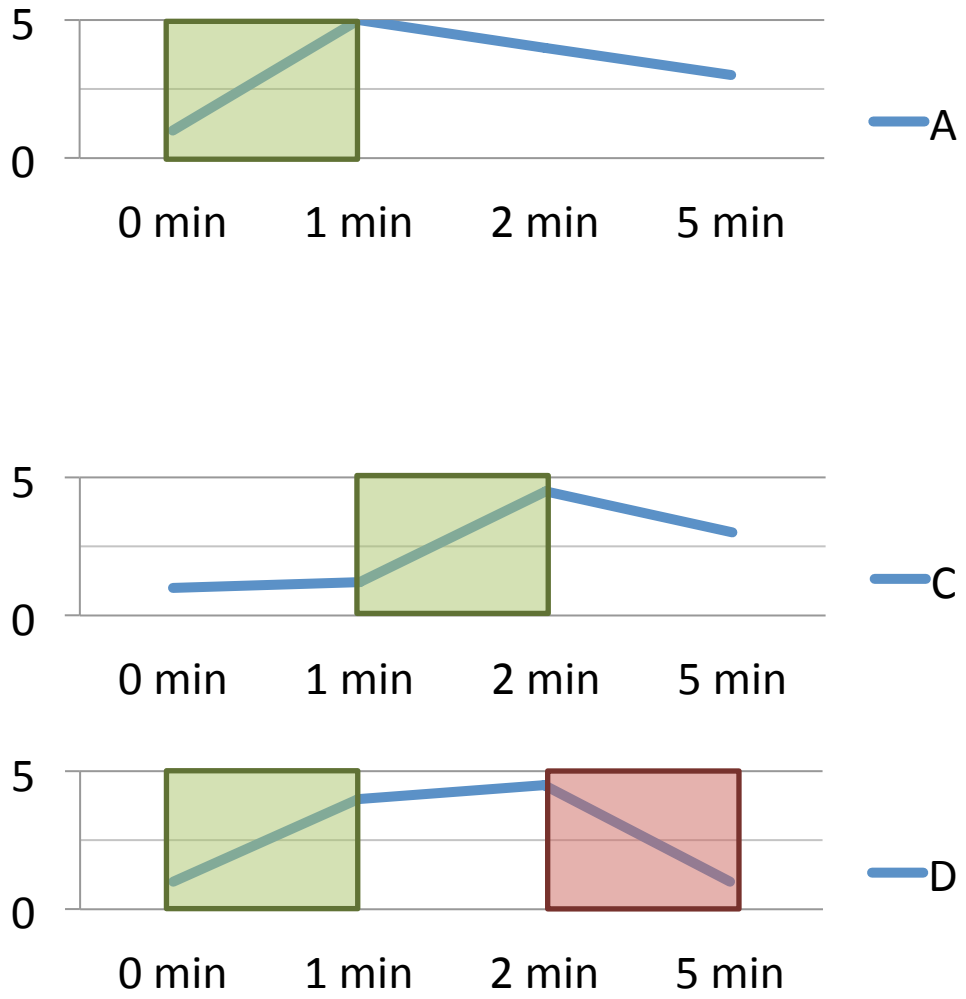


act_A

act_B

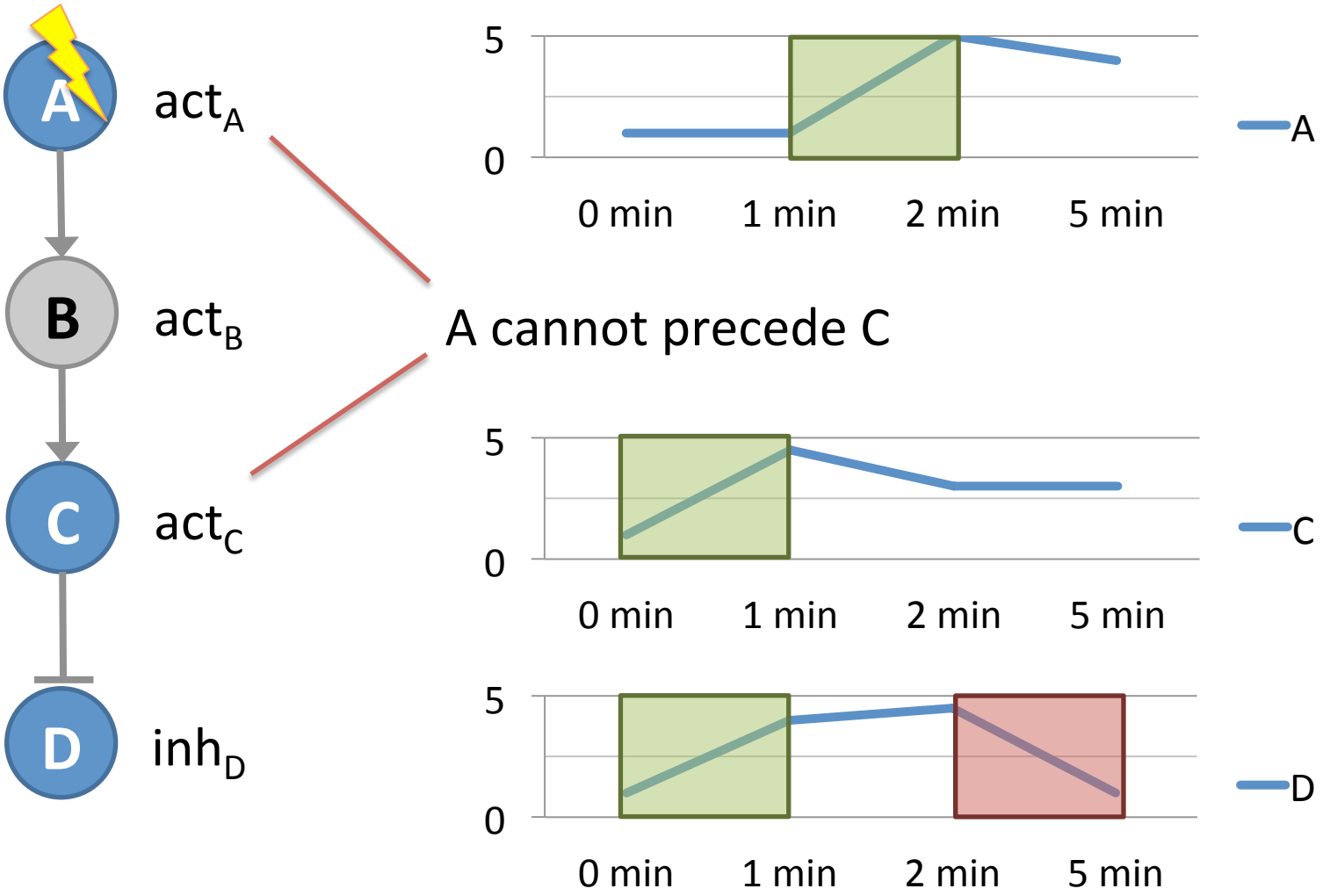
act_C

inh_D

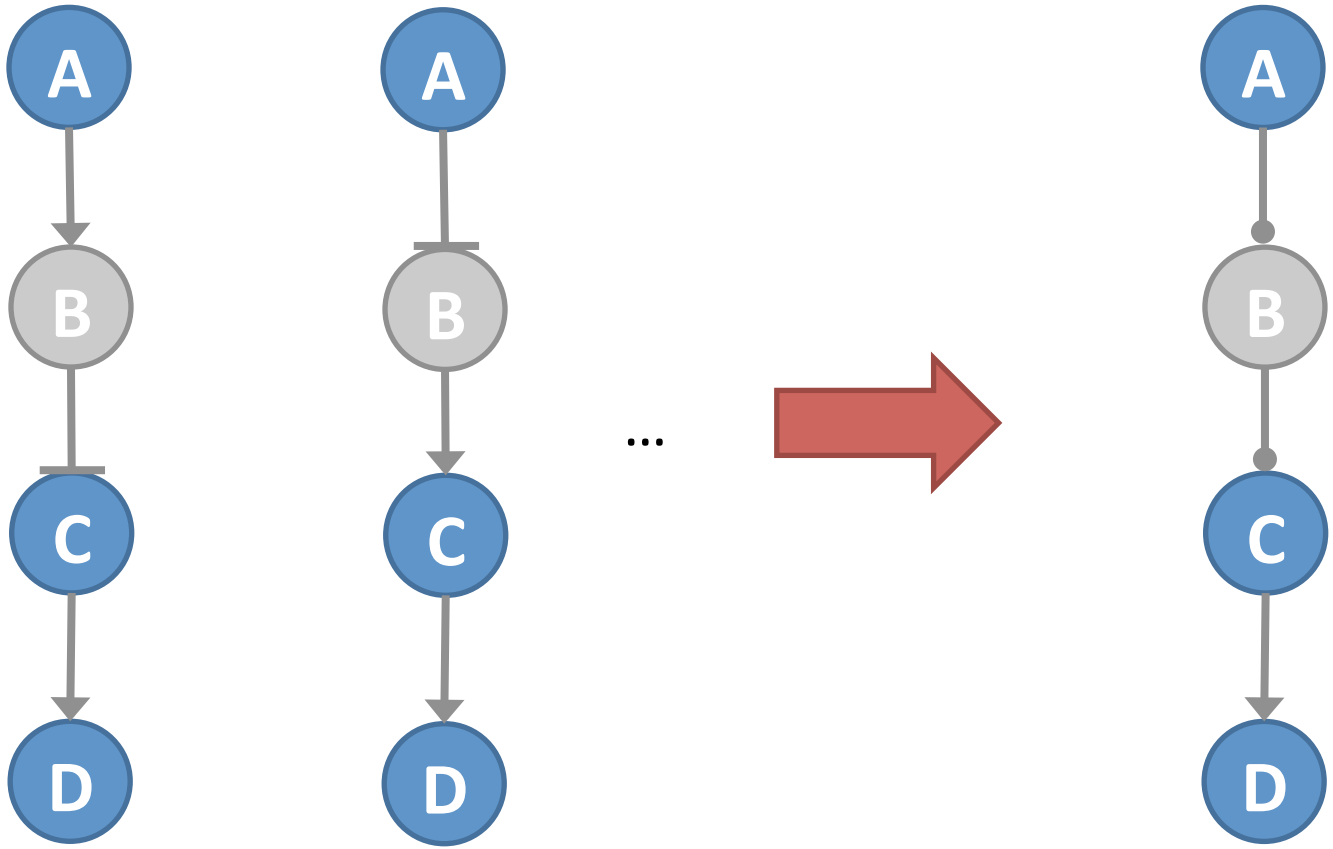


Temporal Constraints

Interaction path **inconsistent** with temporal data:



Summarizing All Valid Models



B is unobserved, we can't determine edge sign

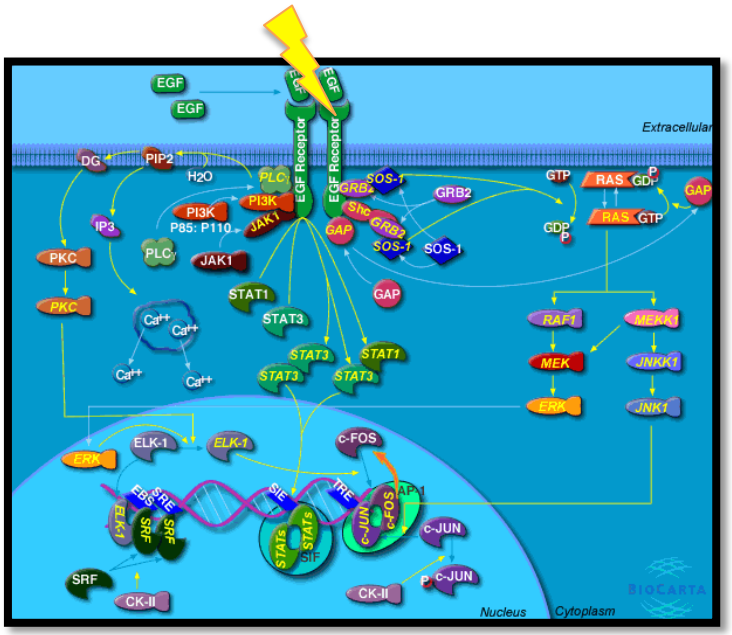
Summarize all valid solutions as a union of networks

EGFR Case Study: Materials

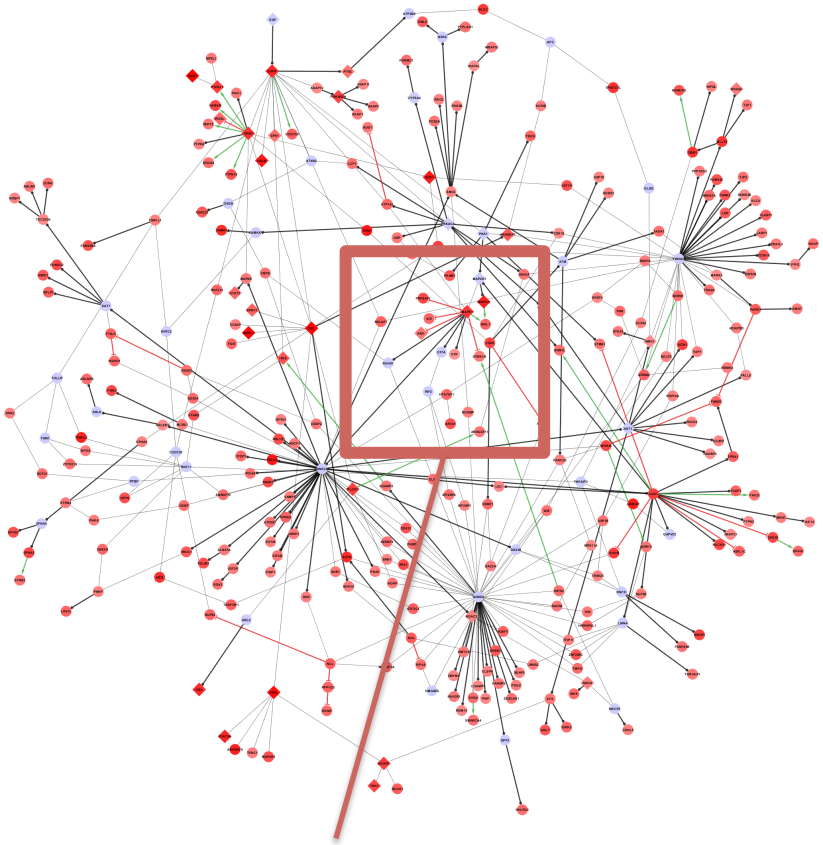
Stimulate EGFR Flp-In HEK-293 cells with EGF ligand.

Mass spectrometry at 0, 2, 4, 8, 16, 32, 64, 128 mins.

Observe 203 significantly phosphorylated proteins.



Inferred EGFR Pathway Map



MAPK subnetwork

We inferred a summary network of **413** edges

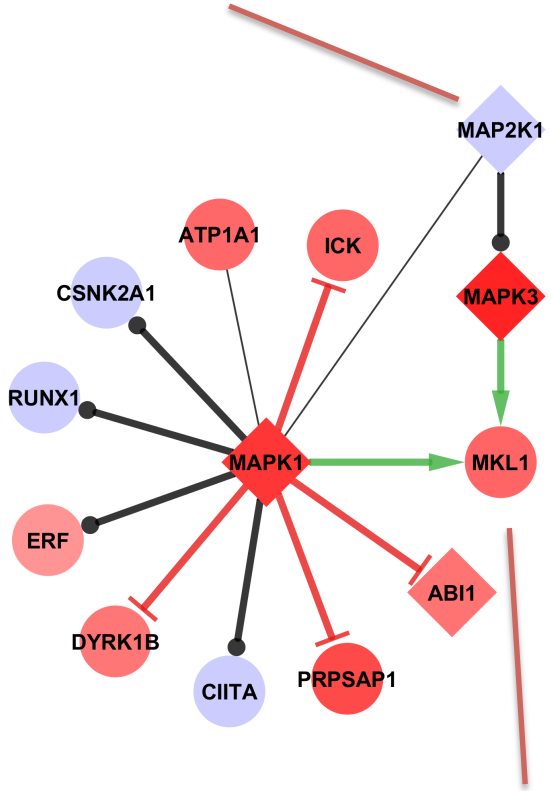
202 edges have the same direction in all models

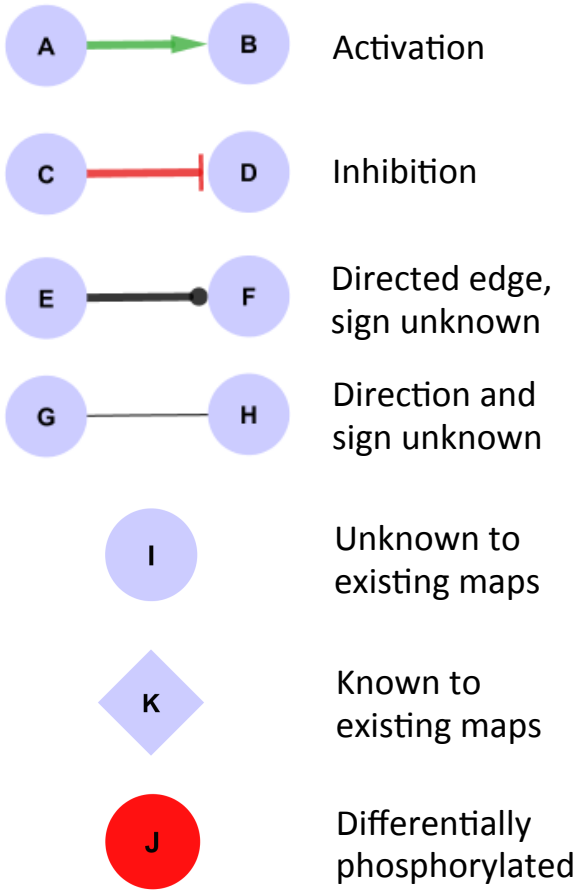
38 edges have the same sign and direction in all models

83% of phosphorylated proteins are included

Inferred MAPK Subnetwork

MAP2K1 unobserved but recovered by topology

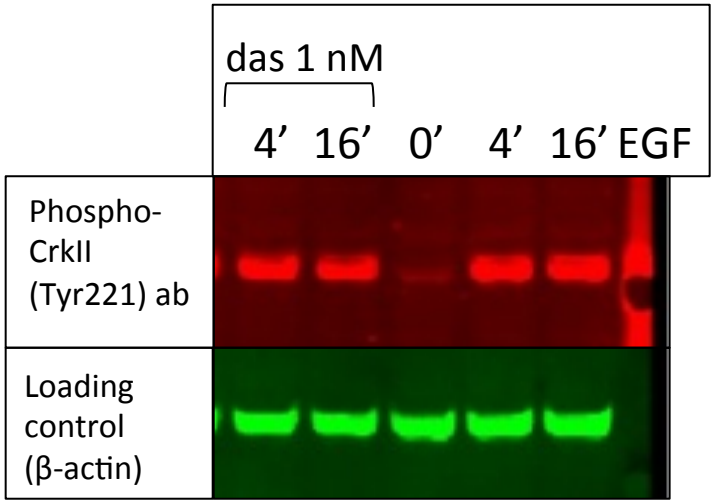




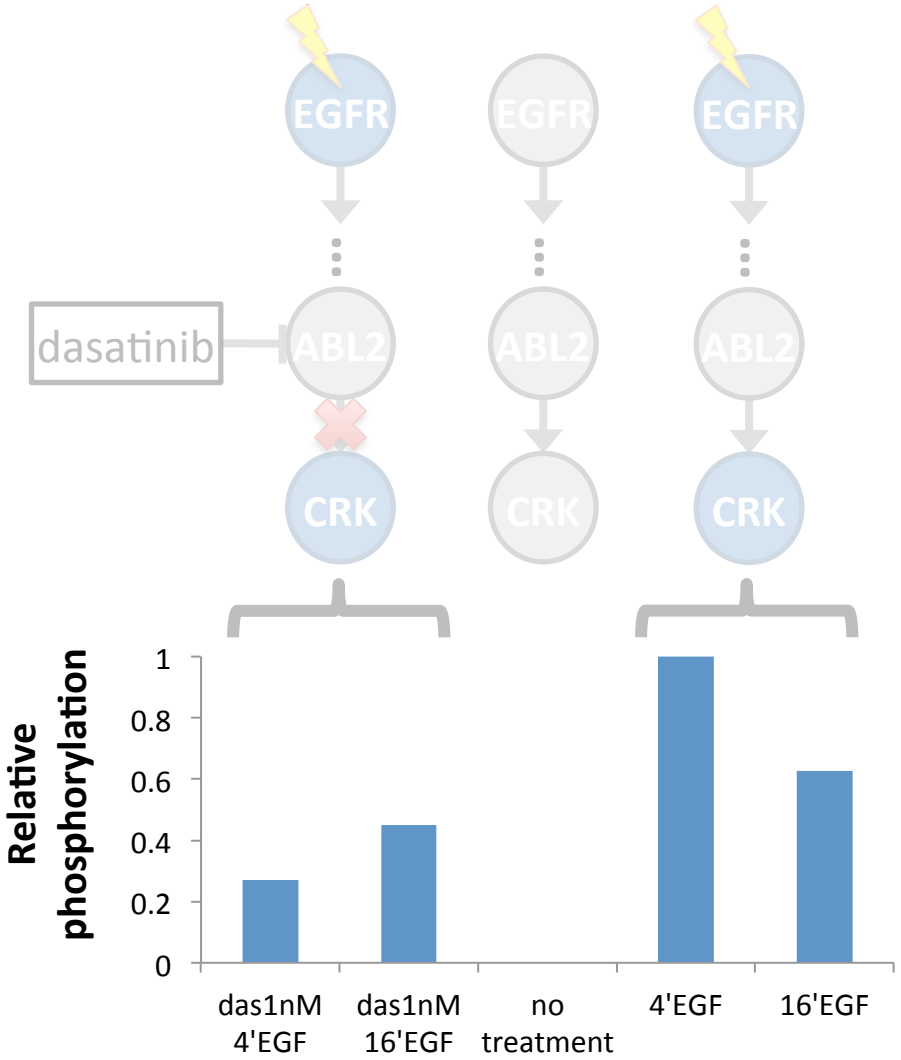
Activation of MKL1 inferred via temporal data and prior knowledge

Experimental Validation

Preliminary results validating the prediction Abl2 → Crk



a-pCrk blot, inhibition of Abl2 with 1 nM dasatinib



Conclusion

Pathway models that agree with actual dynamic signaling events

Joint inference with multiple types of constraints (topological, temporal, prior knowledge, …)

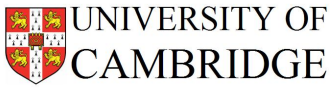
Detect non-ambiguous predictions across all valid models for experimental validation

Acknowledgements

Anthony Gitter

Ernest Fraenkel

Rastislav Bodík
Saurabh Srivastava



Jasmin Fisher

Nir Piterman

Alejandro Wolf-Yadlin
Kirsten Beck
Aaron McKenna