Synthesizing Signaling Pathways from Temporal Phosphoproteomic Data

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

\textsuperscript{1}University of California, Berkeley, \textsuperscript{2}University of Wisconsin-Madison, \textsuperscript{3}University of Washington, \textsuperscript{4}20n, \textsuperscript{5}University of Leicester, \textsuperscript{6}Massachusetts Institute of Technology, \textsuperscript{7}Microsoft Research Cambridge
Beyond Pathway Maps

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

- 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\textsubscript{A} for A
- Act\textsubscript{B} for B
- Act\textsubscript{C} for C
- Inh\textsubscript{D} for D
Temporal Constraints

Interaction path **inconsistent** with temporal data:

- **act\(_A\)**
- **act\(_B\)**
- **act\(_C\)**
- **inh\(_D\)**

A cannot precede C
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.
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.

MAPK subnetwork
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 $\text{Abl2} \rightarrow \text{Crk}$

- **Loading control** (β-actin)
- **Phospho-CrkII** (Tyr221) ab

| das 1 nM | 4’ | 16’ | 0’ | 4’ | 16’ EGF |

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

**Relative phosphorylation**

- das1nM 4’EGF
- das1nM 16’EGF
- no treatment
- 4’EGF
- 16’EGF
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