An Egf Signaling Map in Pathway Logic

M. Knapp, L. Briesemeister, S. Eker, P. Lincoln, I. Mason, C. Talcott, and K. Laderoute

SRI International
Menlo Park, CA 94025

Abstract

Pathway Logic is an approach to modeling cellular processes based on rewriting logic, a simple logic designed for modeling and analysis of distributed systems. It allows one to model aspects of the structure and state of interacting components as elements of an abstract data type; to represent individual process steps (reactions) as rewrite rules; and to study possible ways a system might evolve using techniques based on logical inference. Given a network of reactions and a specification of cellular components one can query the network about possible reaction pathways and outcomes. Knockouts that prevent a given outcome can be computed, competing reactions can be found, and pathways can be compared to look for potential cross-talk.

Epidermal growth factor receptor (EgfR) signaling regulates growth, survival, proliferation, and differentiation in mammalian cells. We present a Pathway Logic model of early response to Epidermal growth factor (Egf) stimulation in adherent cells expressing Egf-receptors. The model is entirely based on experimental results and data curated from the published scientific literature.

We explain the curation process that extracts information about state changes from experimental data to determine the components of a reaction rule. A reaction network assembled from data supporting events that might be downstream of EgfR signaling has over 370 reactions involving more than 460 species (signaling molecules in different states and locations). The network was then constrained to outcomes that have been demonstrated experimentally to happen in response to a short stimulus with Egf. So far more than 80 such outcomes have been collected and used to test the adequacy of the model---does it predict the observed outcomes? We will also point out some surprises that appear in the generated pathways due to the richness of the considered biological context.

Pathway Logic models are qualitative and thus answer different kinds of questions than quantitative differential equation-based or stochastic models. They can also serve as useful road maps for developing quantitative models of subsystems and understanding the potential cross talk between signaling submodules.

Introduction

Various representations or models of Egf-to-Erk signal transduction have been described over the last decade. In general, this process is modeled as follows:

Egf ➔ EgfR ➔ Grb2 ➔ Sos1 ➔ a Ras family member
➔ Raf1 ➔ MEK1/2 ➔ ERK1/2

In this canonical pathway, Egf binds to the EGF receptor (EgfR) and stimulates its protein tyrosine kinase activity to cause autophosphorylation. Next, a complex containing the adaptor protein Grb2 and the guanine nucleotide exchange factor Sos1 docks (binds) to the autophosphorylated (activated) EgfR. The Sos1-containing EgfR complex activates a Ras family GTPase, and the activated Ras protein activates Raf1, a member of the RAF serine/threonine protein kinase family. Raf1 then activates the dual-specificity protein kinases Mek1 and/or Mek2 (MEK1/2), which activate Erk1 and/or Erk2 (ERK1/2).

Here we present a Pathway Logic model of early signal transduction by EgfR that is entirely based on experimental results curated from the published scientific literature. We demonstrate that the series of events between activation of EgfR by Egf and activation of Erk2 may not be as simple as those described in the canonical pathway.

A note about protein nomenclature: the symbols used for proteins in this presentation are those used in Pathway Logic. Other identifiers and synonyms are provided in the Glossary in the Appendix.

Data Collection

The first step in the construction of the model was to collect the data: 174 papers were searched for appropriate experiments and the results were listed as 1373 evidence items. Below is an example of an item used as evidence for rule E19 which requires the presence of tyrosine phosphorylated Gab1 for the activation of Erk2 in response to Egf.

An Evidence Item

Source	PMID: 11323411 Type: data Figure: 2
Subject	Pathway Logic name: Erk2 Expressed: yes Identification method: expression tag antibody
State Change	Type: kinase activity Direction: increase Assay: IP Kinase assay, MBP as substrate
Cause	Stimulus: Egf Time: 5 minutes Concentration: 0.25 ng/ml
Requirement	Pathway Logic name: Gab1 Method: Y627F dominant-negative coexpression Method: Y659F dominant-negative coexpression
Environment	Cells: COS-7 State: serum starved for 20 hr

Evidence items do not use the conclusions of the authors, only the experimental results.

Writing the Rules

Initially a set of Pathway Logic rules were derived from published pathway diagrams and review articles. The rules were annotated with evidence items that support or contradict them. If no evidence could be found to support a rule, it was either removed or modified to agree with the data. The remaining evidence items were used to construct rules not found in conventional pathway diagrams.

The rules in this model are divided into two sets:

(1) **Common Rules** are based on state changes caused by ectopically expressed proteins.

(2) **Egf Rules** are based on state changes caused by short-term Egf treatment. In most cases, the treatment was 10 minutes or less, which is the time window corresponding to the peak response of Erk phosphorylation.

Choosing an Initial State

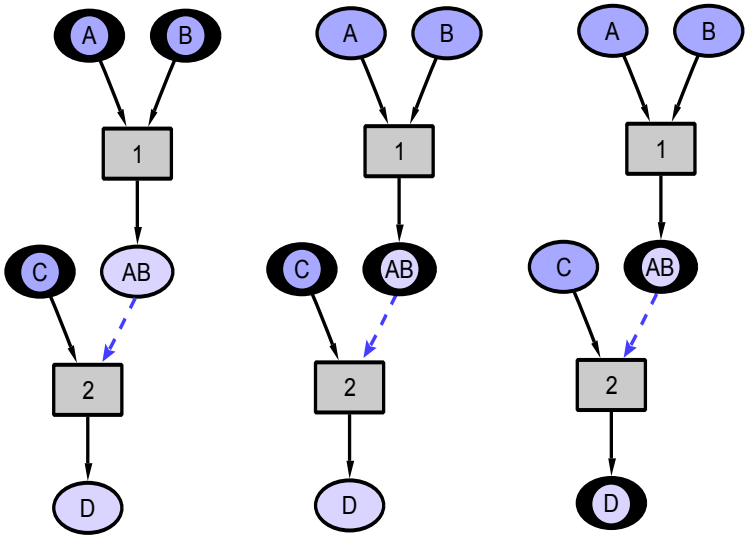
The initial state for a Pathway Logic model consists of a list of components (proteins, chemicals, or nucleic acids), their modifications, and locations. In this case, the initial state represents a serum starved, adherent cell expressing EgfR. The list was curated from published experimental data (**Table I**, **Appendix**).

The initial state also contains one ligand - Egf.

About Petri Nets

In Pathway Logic, reaction networks and pathways are represented as Petri nets. Petri nets were invented to model execution of concurrent processes (such as signals propagating through a cell). A Petri net can be thought of as graph with two kinds of nodes: **transitions/rules** (reactions -- shown as squares) and **places/ occurrences** (reactants, products, modifiers -- shown as ovals). Each occurrence represents a chemical, protein or complex in a specific state and location.

The **reactants** of a rule are the occurrences connected to the rule by arrows from the occurrence to the rule. The **products** of a rule are the occurrences connected to the rule by arrows from the rule to the occurrence. The **modifiers** of a rule are the occurrences connected to the rule by a dashed arrow. For example rule 2 has reactant C, product D and modifier AB.



To execute a Petri net model one puts tokens on the ovals corresponding to occurrences present in the initial state (represented by darker colors). A rule can fire if all of its reactants and modifiers have tokens. When a rule fires the tokens are moved from reactants to products, indicating that the products are now present. Modifiers are unchanged. An execution is a sequence of firings of enabled rules. (Note there may several possible executions.) A pathway is a network with initial marking that corresponds to rules fired in one possible execution.

Building a Petri Net

The rules and initial state files were loaded into the Pathway Logic Assistant (PLA). PLA converted all the rules into one Petri net and then removed any transitions not allowed by the initial state. The result was a Petri net that included all curated events that might occur in response to a short-term stimulus with EGF. At this point the Petri net looked like this:

The Unconstrained Petri Net:

Events that could occur in response to Egf

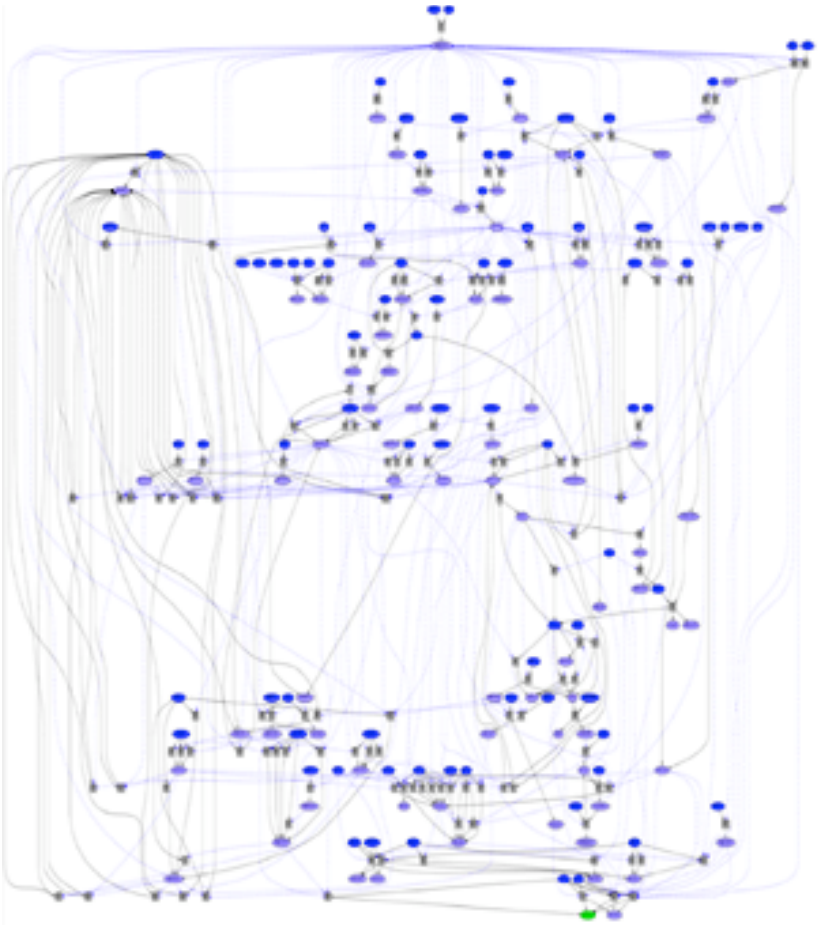
Although PLA can be used to browse this network, it is clearly too complex to understand as a whole.

Making a Subnet

We were interested in the mechanism of the initial activation of Erk2 in response to a stimulus by Egf. Once a Petri net had been built, the PLA Viewer was used to make a subnet in which anything not relevant to the Erk2 activation was stripped away. Unfortunately, there are still so many potential paths to Erk2 activation that the Petri net is still too complicated to comprehend without visualization tools.

A Subnet of the Unconstrained Petri Net

All possible routes to from Egf to Erk2 activation



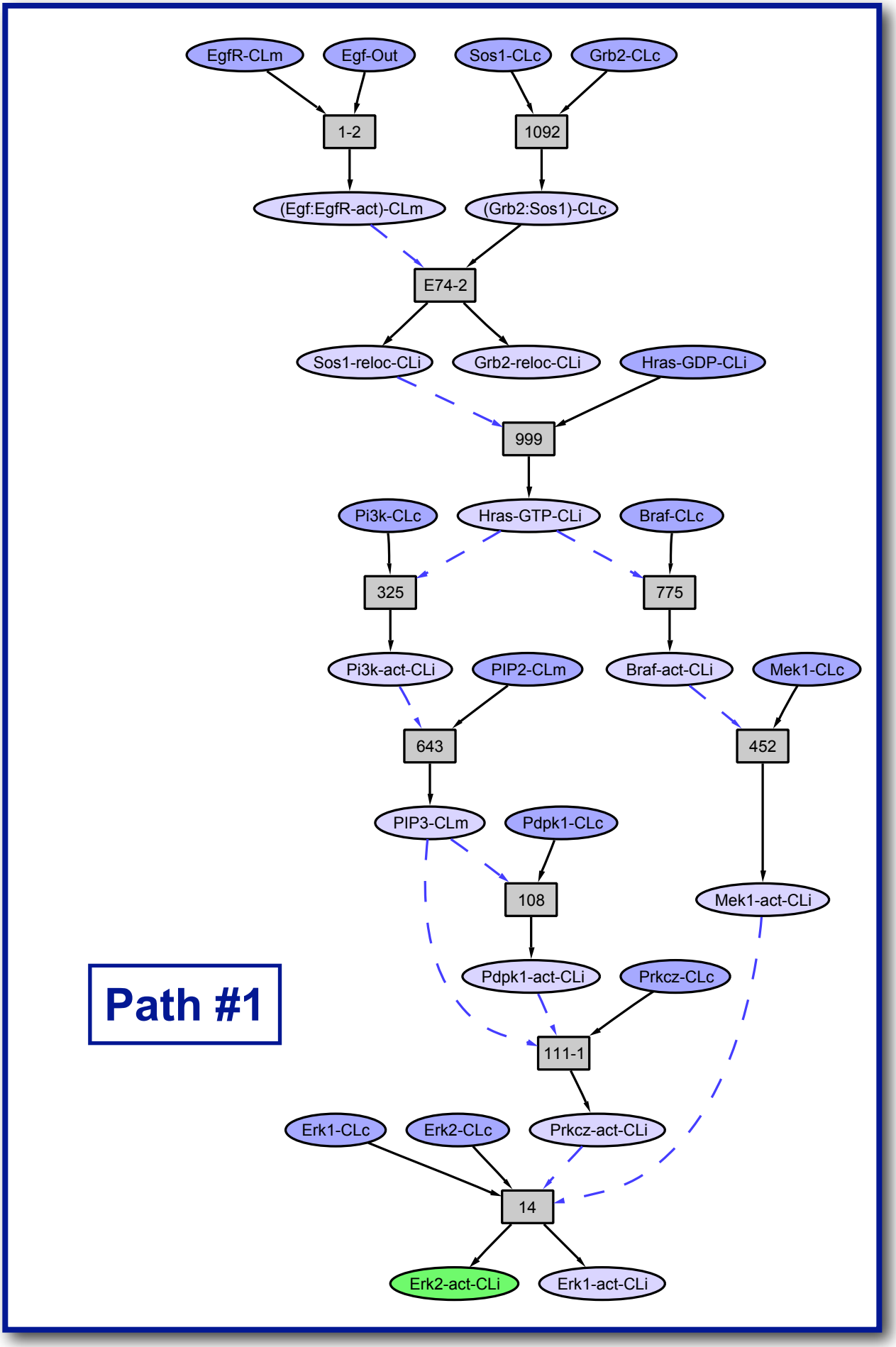
Finding Paths

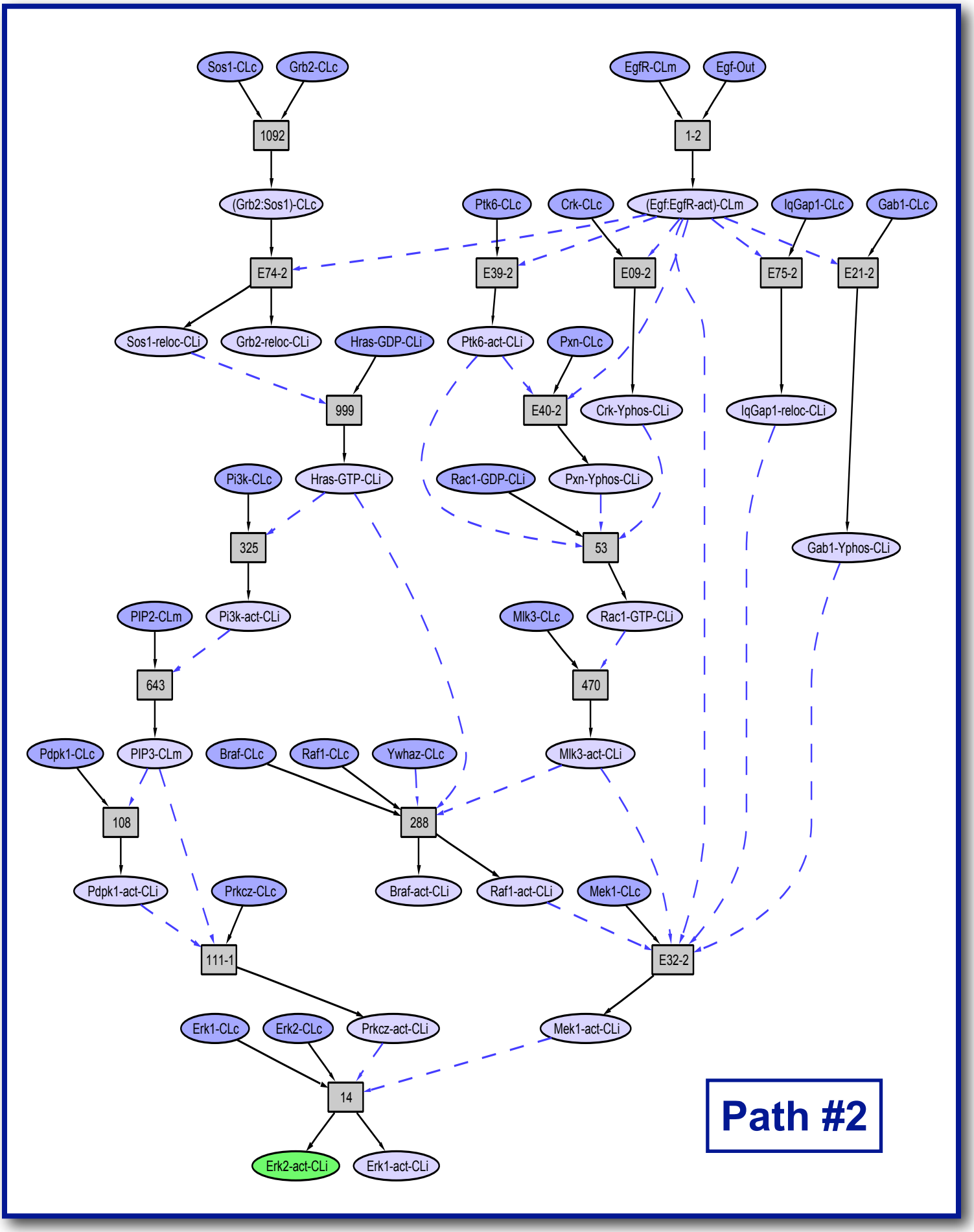
Given an initial state and a goal state, the PLA viewer can isolate possible paths between them. It uses a model-checking tool named LoLa (www.informatik.hu-berlin.de/~kschmidt/lola.html) that returns the first path it finds -- usually the path with the least number of steps. We asked LoLa to find a Path from Egf to Erk2 activation.

The first path that LoLa found (**Path #1**) is similar to the canonical pathway but uses Braf instead of Raf1 to activate Mek1 and requires activated Prkcz as well as Mek1 to activate Erk1 and Erk2.

Is this the "correct" path? Unfortunately, we have no tools to answer that question automatically. But we have provided the evidence collected to date to support each rule. By inspecting the evidence for each of the rules, it becomes apparent that the evidence supporting rule 452 is weak; there is only one datum and it is an in vitro kinase assay showing that recombinant Mek1 can be activated by recombinant Braf.

PLA allows the user to "hide" rule 452 and ask for another path. The new path (**Path #2**) uses rule 288 in which both Braf and Raf1 are activated by Hras, Mlk3, and Ywhaz. This result correlates well with the next rule E32 in which Mek1 activation in response to Egf requires Raf1, Mlk3, Gab1, and IqGap. The user can access the evidence for a rule via PLA or in the curated model. It is left up to the user to decide whether there is sufficient data to justify the rules.


EgfR-CLm
Egf-Out
Sos1-CLc
Grb2-CLc
1-2
1092
(Egf:EgfR-act)-CLm
(Grb2:Sos1)-CLc
E74-2
Sos1-reloc-CLi
Grb2-reloc-CLi
Hras-GDP-CLi
999
Pi3k-CLc
Hras-GTP-CLi
Braf-CLc
325
775
Pi3k-act-CLi
PIP2-CLm
Braf-act-CLi
Mek1-CLc
643
452
PIP3-CLm
Pdpk1-CLc
108
Mek1-act-CLi
Pdpk1-act-CLi
Prkcz-CLc
111-1
Erk1-CLc
Erk2-CLc
Prkcz-act-CLi
14
Erk2-act-CLi
Erk1-act-CLi
Path #1


Sos1-CLc
Grb2-CLc
1092
(Grb2:Sos1)-CLc
E74-2
Sos1-reloc-CLi
Grb2-reloc-CLi
Hras-GDP-CLi
999
Hras-GTP-CLi
Pi3k-CLc
325
Pi3k-act-CLi
PIP2-CLm
643
PIP3-CLm
Pdpk1-CLc
108
Pdpk1-act-CLi
Prkcz-CLc
111-1
Prkcz-act-CLi
Erk1-CLc
Erk2-CLc
14
Erk2-act-CLi
Erk1-act-CLi
Braf-CLc
Raf1-CLc
Ywhaz-CLc
288
Braf-act-CLi
Raf1-act-CLi
EgfR-CLm
Egf-Out
1-2
(Egf:EgfR-act)-CLm
Ptk6-CLc
Crk-CLc
IqGap1-CLc
Gab1-CLc
E39-2
E09-2
E75-2
E21-2
Ptk6-act-CLi
Pxn-CLc
E40-2
Crk-Yphos-CLi
IqGap1-reloc-CLi
Rac1-GDP-CLi
Pxn-Yphos-CLi
53
Gab1-Yphos-CLi
Mlk3-CLc
Rac1-GTP-CLi
470
Mlk3-act-CLi
Mek1-CLc
E32-2
Mek1-act-CLi
Path #2

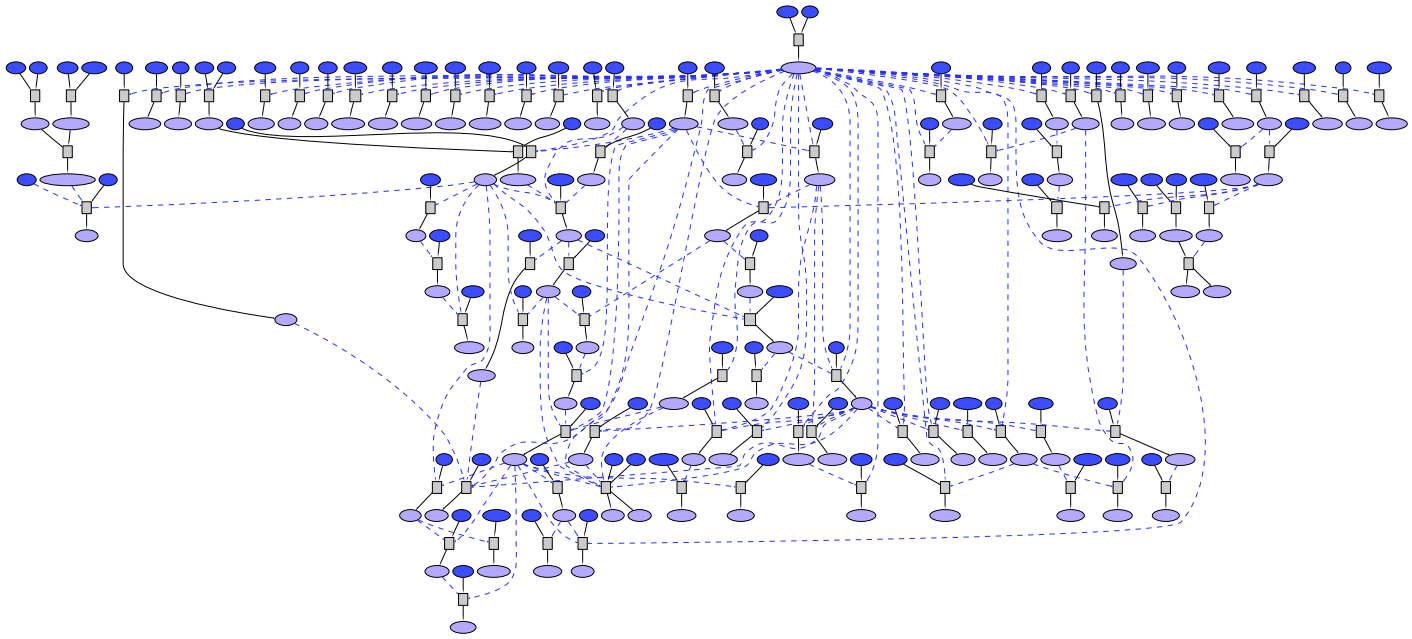
Constraining the Egf Map

There are many possible paths in addition to Paths 1 and 2. One could go on evaluating each one manually or use the PLA tools to limit (or constrain) the number of possible interactions by giving precedence to events found in response to Egf. We applied two constraints to the network of Panel 7.

The first constraint took advantage of a list of 85 state changes demonstrated experimentally to occur in response to a short stimulus with Egf (**Table II, Appendix**). These occurrences (protein states) were set as goals and a set of concurrent paths were produced by PLA. This selection ensures that the paths used to reach the chosen goals are mutually compatible (i.e., no rules would be used that would allow one goal to be reached at the expense of any other goal).

The second constraint used PLA's ability to hide rules, thus giving precedence to Egf Rules over Common Rules. The Egf Rules contain requirements specific to Egf signaling that must be satisfied before they can fire. By using Egf Rules instead of Common Rules wherever possible, LoLa is forced to use events must happen before Erk2 is activated.

The Constrained Egf Map



The Egf Map above was constructed from the network of **Panel 7** using both constraints. It is much simpler, and more focused, but still sufficiently complex that tools are needed to explore and understand it.

The Path to Erk2 in the Constrained Egf Map

Panel 15 shows the path from the constrained Egf map **(Panel 13)** that leads to activated Erk2. This differs from the path found by searching in the unconstrained network **(Panel 7)** as we have forced the model-checking tool to work in the context of realizing all the other observed goals.

This path is a great deal more complicated than the canonical pathway. This result is not surprising if you look at all the reported requirements for Erk2 to be activated. **(Table III, Appendix)** contains a list of the proteins that have been tested for their influence on Erk activation and/or phosphorylation in response to Egf. The shaded rows in the table are events that the curator chose to use in the Egf Rules.

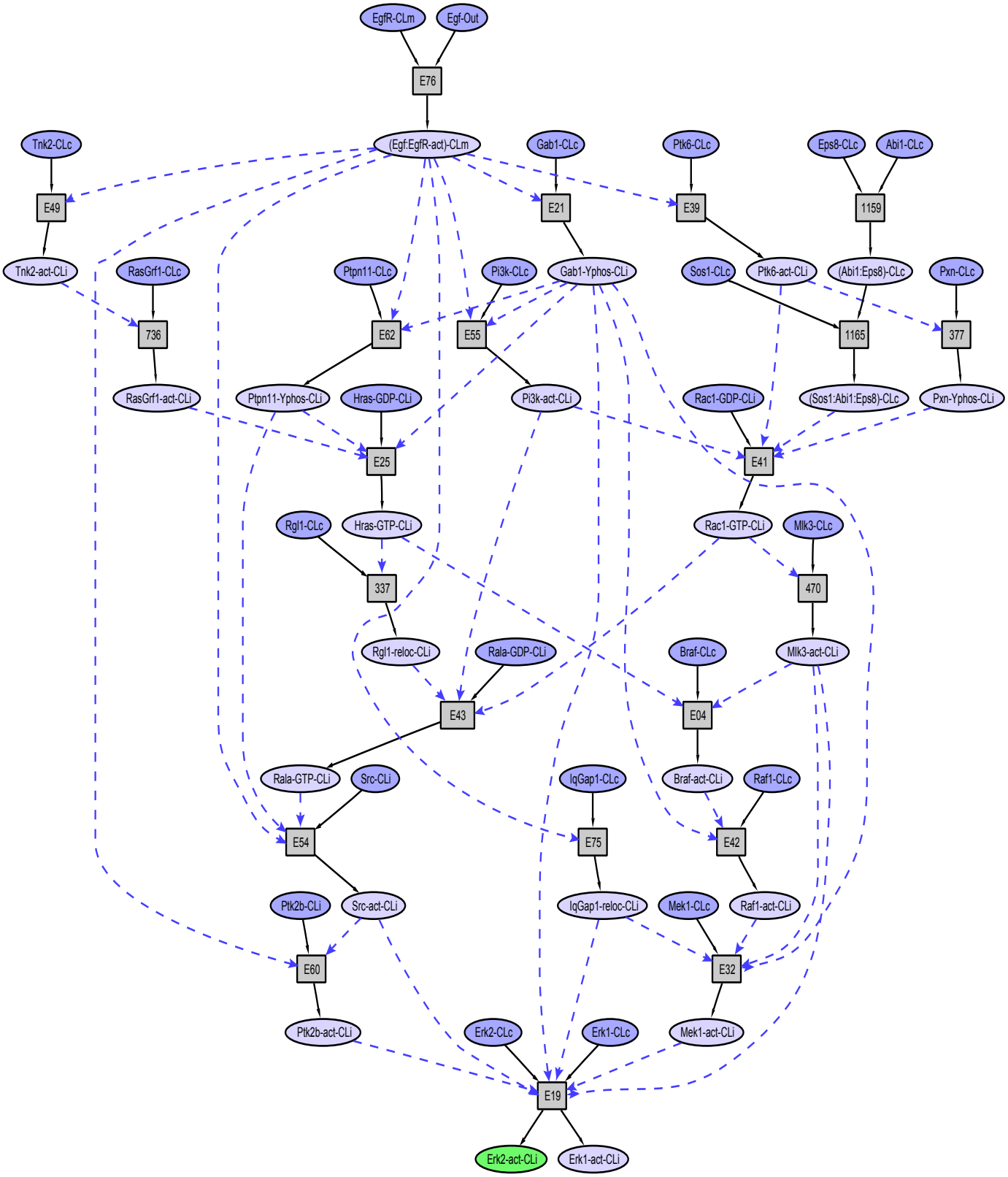
Clearly, there are many unfamiliar events within the new Egf to Erk2 path. We point out three such events in the following panels.

(1) Rala is required for Src activation in response to Egf **(Panel 16)**.

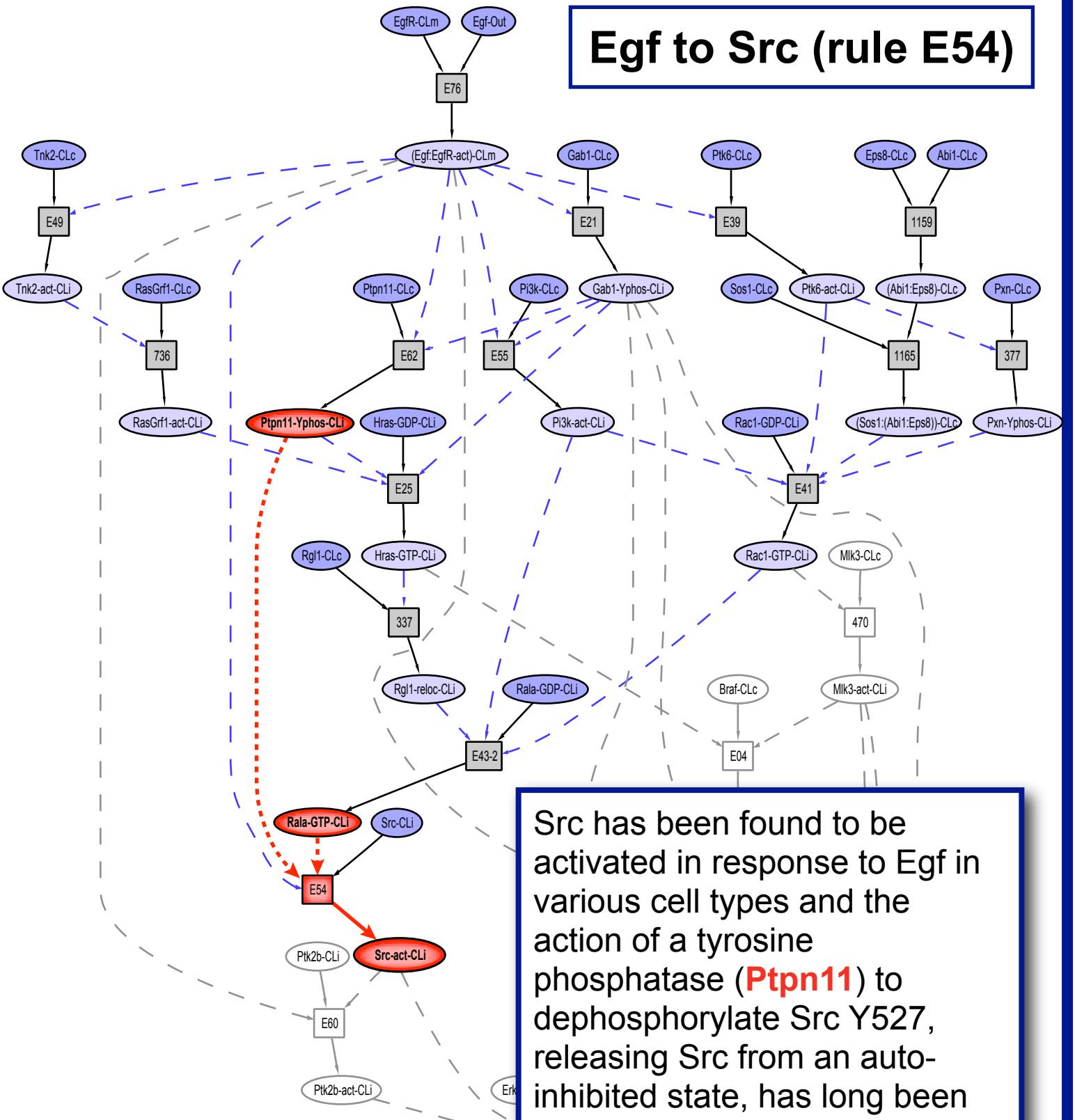
(2) Sos1 is not required for Hras activation in response to Egf **(Panel 17)**.

(3) Mlk3 is required for the activation of Braf, Mek1/2, and Erk1/2 **(Panel 18)**.

A Path from Egf to Erk2 activation derived from Egf Specific Rules

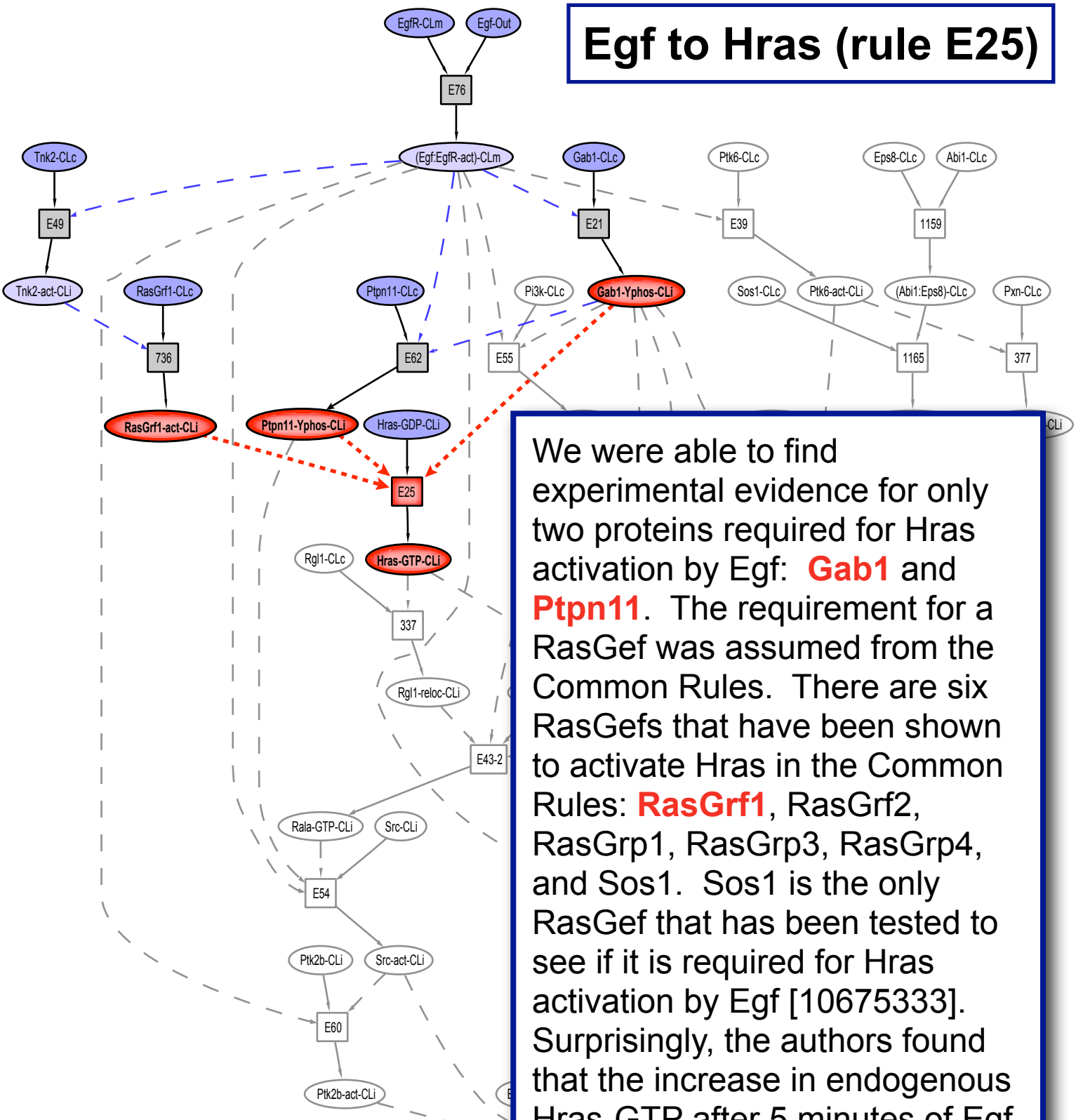


Egf to Src (rule E54)



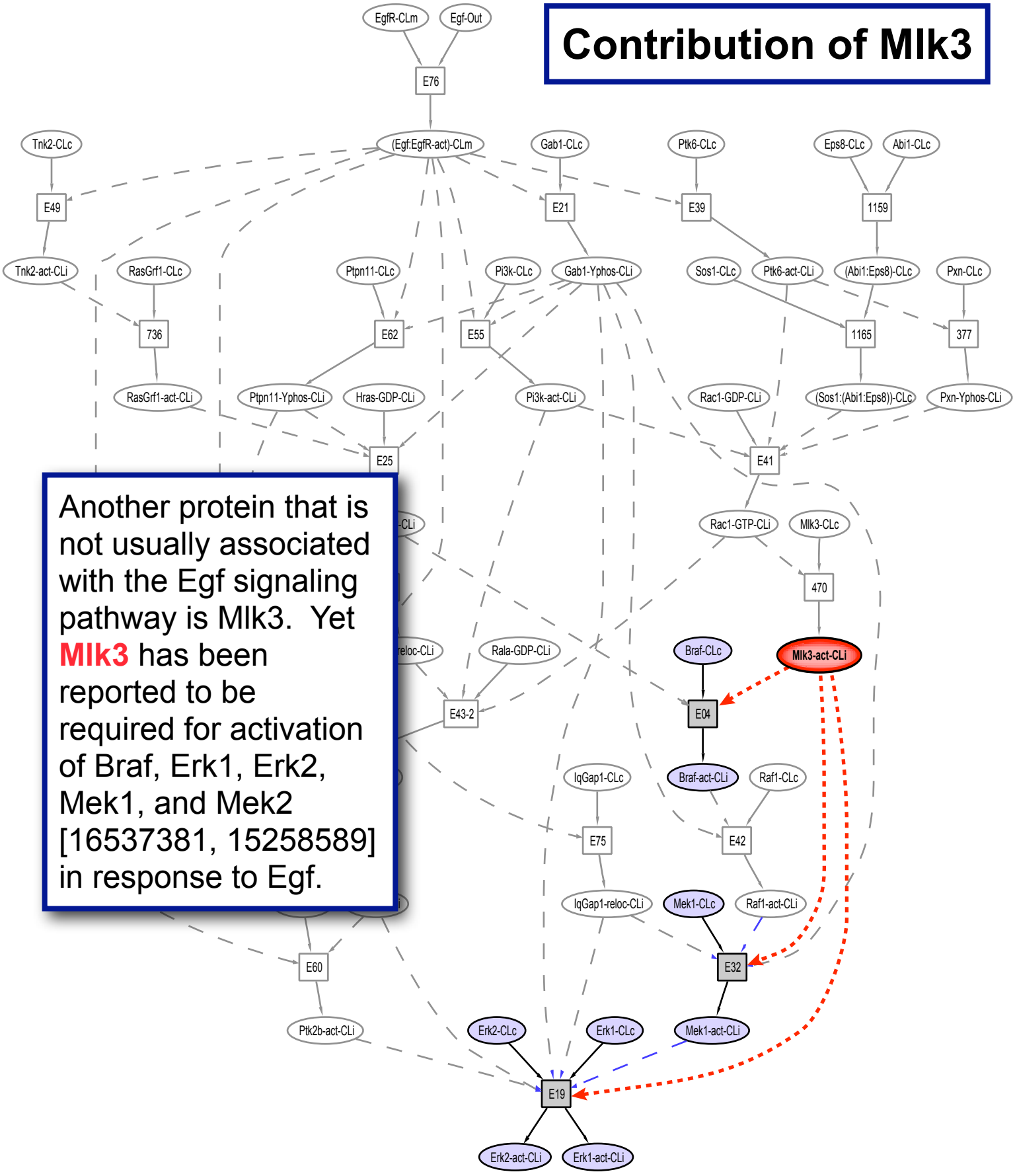
Src has been found to be activated in response to Egf in various cell types and the action of a tyrosine phosphatase (**Ptpn11**) to dephosphorylate Src Y527, releasing Src from an auto-inhibited state, has long been recognized. Yet a possible requirement for **Rala** in the activation of Src [10675331] has been ignored.

Egf to Hras (rule E25)



We were able to find experimental evidence for only two proteins required for Hras activation by Egf: **Gab1** and **Ptpn11**. The requirement for a RasGef was assumed from the Common Rules. There are six RasGefs that have been shown to activate Hras in the Common Rules: **RasGrf1**, RasGrf2, RasGrp1, RasGrp3, RasGrp4, and Sos1. Sos1 is the only RasGef that has been tested to see if it is required for Hras activation by Egf [10675333]. Surprisingly, the authors found that the increase in endogenous Hras-GTP after 5 minutes of Egf treatment was not affected by knockout of Sos1 in mouse embryo fibroblasts.

Contribution of Mlk3



Another protein that is not usually associated with the Egf signaling pathway is Mlk3. Yet **Mlk3** has been reported to be required for activation of Braf, Erk1, Erk2, Mek1, and Mek2 [16537381, 15258589] in response to Egf.

Validation of the Egf Map

The most obvious way to validate our model is by experimentation. The data used to prepare this map was obtained from published experiments from different research groups using different cell lines, times of Egf stimulation, levels of EgfR saturation, lysis methods, antibodies, plasmids, and knock-out technologies. The analysis used to generate our model assumes that variations in the assays and materials did not significantly change the conclusions used to generate the rules. This major assumption could be tested by reproducing the Egf to Erk experiments with a standardized cell line and protocol.

Conclusion

The Pathway Logic model of Egf stimulation was developed by curation of a knowledge base of signaling reactions from experimental data in the published literature. Both general reactions and reactions specific to cells stimulated by Egf have been curated. Subnets and paths were assembled by postulating an initial state (proteins expressed in a cell and external ligands) and using logical tools to:

(1) find all reactions that might be reached given the intial state

(2) find a pathway realizing specified observations (occurrences) within the full network, or within subnetworks.

This often leads to pathways that are more complex than the usual predefined pathways. These inferred pathways can serve as a guide for developing dynamic models as well as to suggest further experimental studies.

References

[10675331] Goi T, Shipitsin M, Lu Z, Foster DA, Klinz SG, Feig LA. An EGF receptor/Ral-GTPase signaling cascade regulates c-Src activity and substrate specificity. Embo J 2000;19:623-30.

[10675333] Qian X, Esteban L, Vass WC, Upadhyaya C, Papageorge AG, Yienger K, Ward JM, Lowy DR, Santos E. The Sos1 and Sos2 Ras-specific exchange factors: differences in placental expression and signaling properties. Embo J 2000;19:642-54.

[15258589] Chadee DN, Kyriakis JM. MLK3 is required for mitogen activation of B-Raf, ERK and cell proliferation. Nat Cell Biol 2004;6:770-6.

[16537381] Chadee DN, Xu D, Hung G, Andalibi A, Lim DJ, Luo Z, Gutmann DH, Kyriakis JM. Mixed-lineage kinase 3 regulates B-Raf through maintenance of the B-Raf/Raf-1 complex and inhibition by the NF2 tumor suppressor protein. Proc Natl Acad Sci U S A 2006;103:4463-8.

Find out more about Pathway Logic at:

http://pl.csl.sri.com

This work was supported by grants GM068146 and CA112970-01 from the National Institutes of Health.