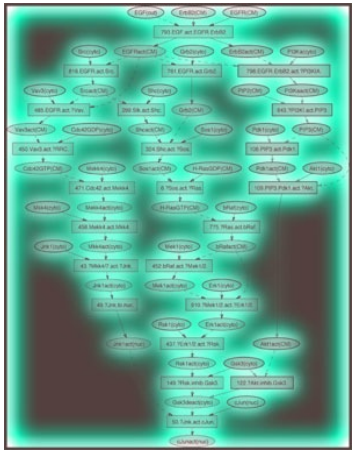


The Pathway Logic Formal Modeling System

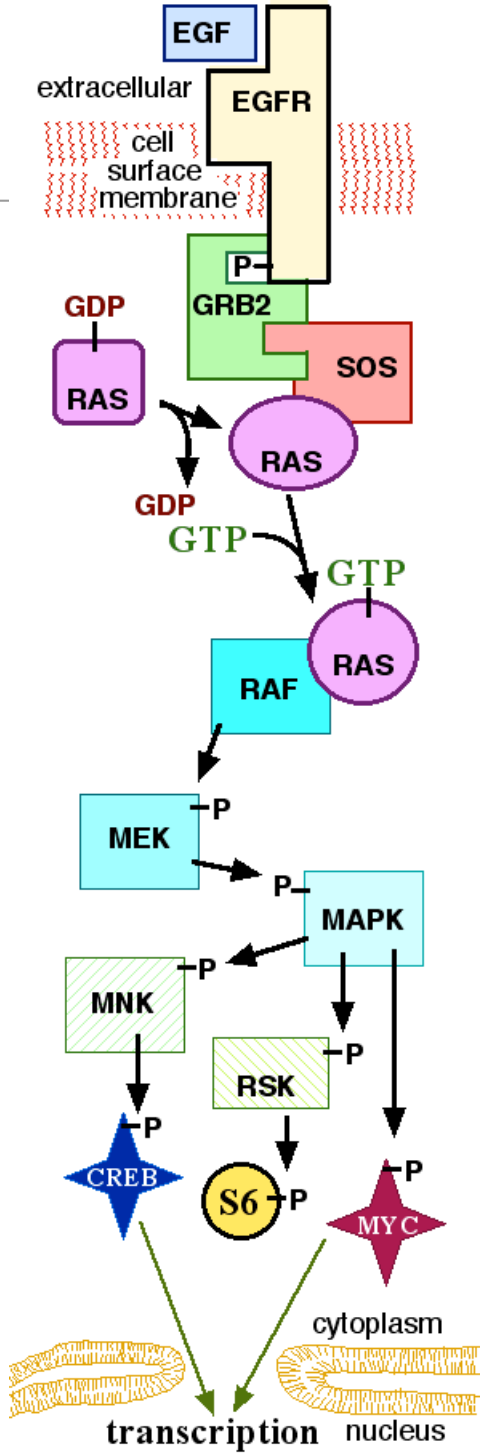
Diverse views of a formal representation of signal transduction

http://pl.csl.sri.com

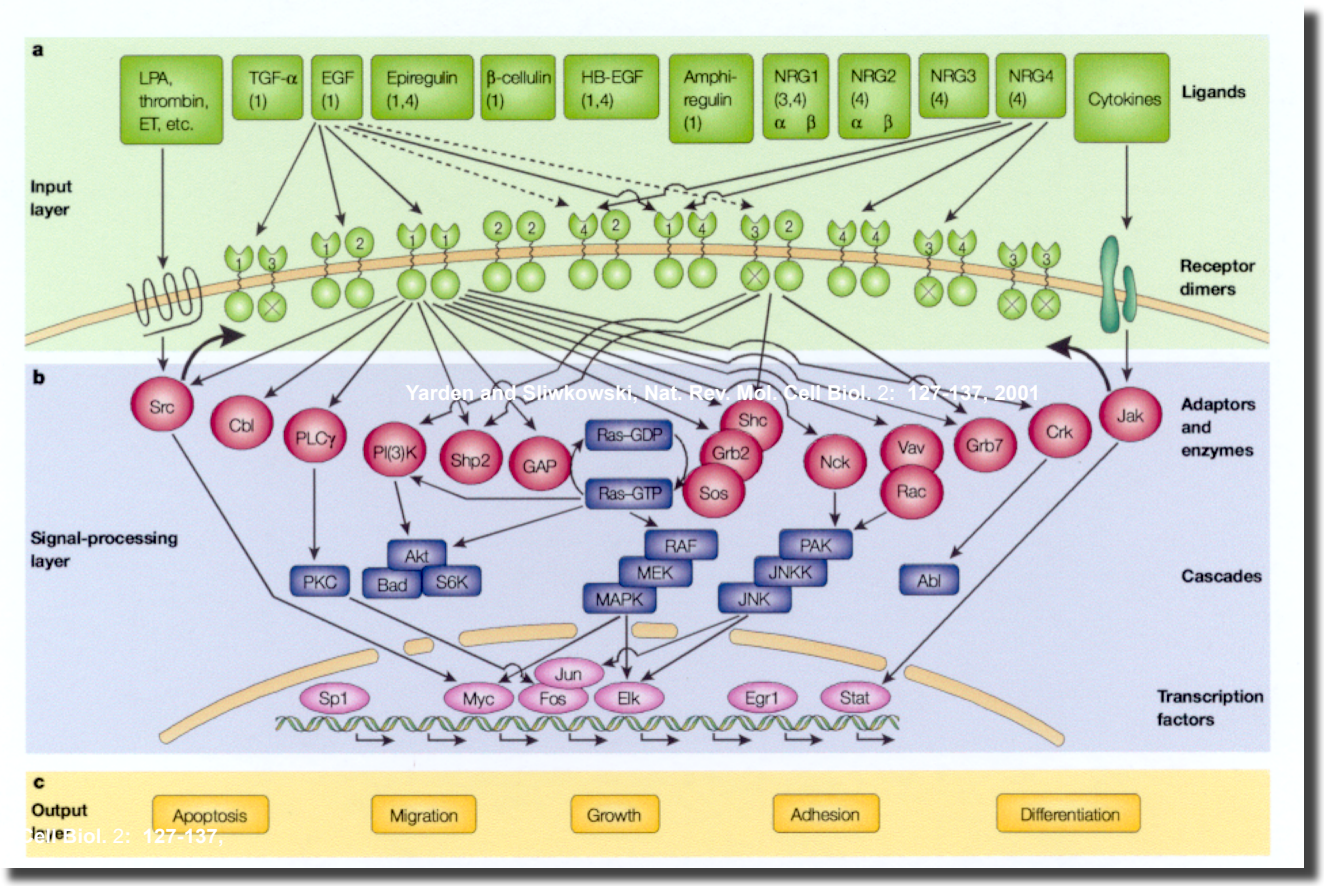
Carolyn Talcott
Formal Methods
for Biological and Biomedical Systems
December 2016

Pathway Logic (PL) Goals

- Understanding how cells work
- Formal models of biomolecular processes that
  - capture biologist intuitions
  - can be executed
- Tools to
  - organize and analyze experimental findings
  - carry out gedanken experiments
  - discover/assemble execution pathways
- New insights into the inner workings of a cell.
- A new kind of review

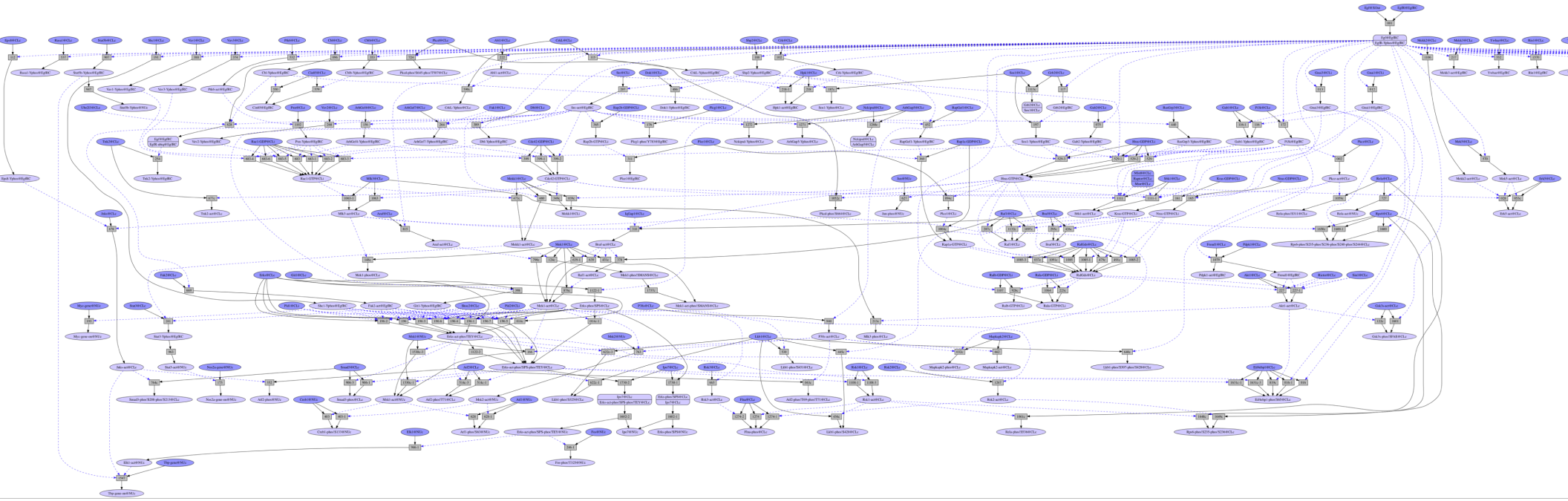


ErbB network cartoon—biologists review model



Yarden and Sliwkowski, Nat. Rev. Mol. Cell Biol. 2: 127-137, 2001

Pathway Logic Egf dish
An interactive, executable review model



Plan

- Modeling 101
- What are we modeling
- Rewriting Logic / Maude
- PL representations
- PL at work
- Conclusion

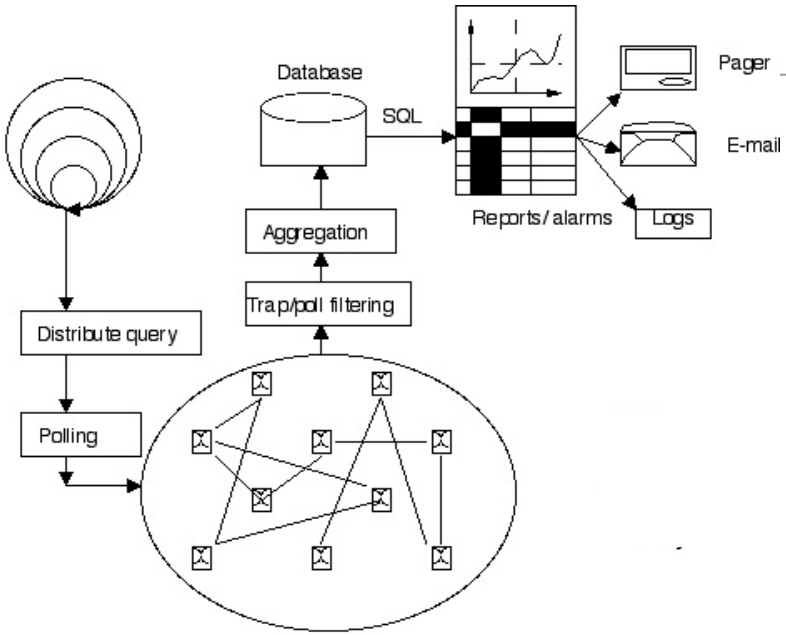
Modeling 101

- What questions do you want the model answer?
- What can you observe/measure?
- What does that mean?
- Explain it to a computer!
    - Need a formal representation system with tools for reasoning

Formal Modeling Methodology

Curator/model builder

data



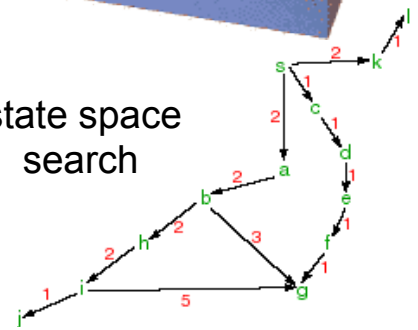
model

asking questions

$S \models \Phi$

model checking

rapid prototyping



state space search

What is a formal system?

- Language
  - to describe system states and properties
- Semantics/System behavior
  - what does a description describe?
- Reasoning principles
  - proving/disproving properties
- Reflection
  - to model and reason about models and reasoning
- Formal model
  - a formal theory that can answer questions using formal reasoning

Executable formal models

- Something to play with — ala model train, architectural model, Sim City…
- Computer Representations of
  - System state: collections of entities (name,location,knowledge,resources..)
    - Initial state: an experimental setup/cell state -- what is expressed, where; what modifications to the cell; what treatment has been applied
  - State transition rules
    - metabolic reaction, signal transduction step, secretion, cell mobility
- Watch it run, poke it, do in silico experiments
- Execution: set of rule applications -- possible behavior
  - How does a signal propagate, watch things light up as modified ...
  - Find collections of cellular components that function together
  - One notion of “Pathway”

Symbolic analysis -- answering questions

- Forward collection
  - upper bound on possible states and transitions
- Backward collection
  - initial states and transitions leading to states of interest
- Search -- for (symbolic) state of interest
- Model checking
  - do all executions satisfy ϕ
  - find counter examples
- Constraint solving -- steady state analysis

Rewriting Logic and Maude

The Pathway LogicFormal Representation System

What is Rewriting Logic?

- Rewriting Logic is a logical formalism that is based on two simple ideas
  - states of a system are represented as elements of an equationally specified algebraic data type
  - the behavior of a system is given by local transitions between states described by rewrite rules
- It is a logic for executable specification and analysis of software systems, that may be concurrent, distributed, or even mobile.
- It is also a (meta) logic for specifying and reasoning about formal systems, including itself (reflection!)

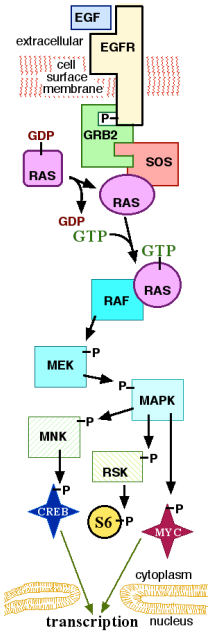
Rewriting Logic Formally

- Rewrite theory: (Signature, Labels, Rules)
- Signature: (Sorts, Ops, Eqns) -- an equational theory
  - Specifies data types that structure of system state, and functions
  - Sorts are partially ordered
- Rules have the form label : t => t’ if cond
- Rewriting operates modulo equations
  - rules apply locally, by matching
  - rule application generates computations (pathways)

Maude

- Maude is a language and tool based on rewriting logic
- See: http://maude.cs.uiuc.edu
- Features:
  - Executability -- position /rule/object fair rewriting
  - High performance engine --- {ACI} matching
  - Modularity and parameterization
  - Builtins -- booleans, number hierarchy, strings
  - Reasoning: search and model-checking
  - Reflection -- using descent and ascent functions

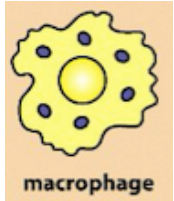
What are we modeling?



About signal transduction

What are cells? A CS view.

- Cells are tiny distributed systems
  - the actors/agents are proteins, chemicals, genes acting concurrently, and occasionally interacting to exchange/transmit information.
- Cells are elements of organ and organism level distributed systems.
  - they make decisions based on local information
  - they communicate by transmitting signals or by binding
  - some are couriers — carrying information from one place to another
  - some are garbage collectors

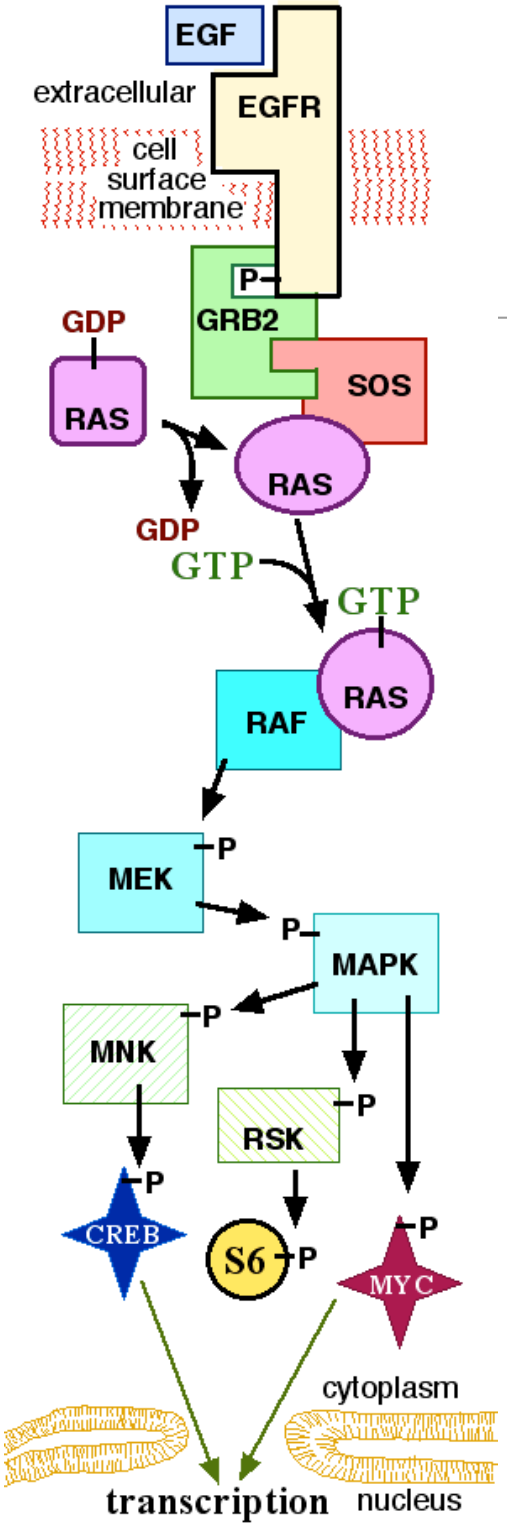


Signaling Pathways

- Signaling pathways involve the modification and/or assembly of proteins and other molecules within cellular compartments into complexes that coordinate and regulate the flow of information.

- Signaling pathways are organized in networks having stimulatory (positive) and inhibitory (negative) feedback loops, and cross talk to ensure that signals are propagated and interpreted appropriately in a particular cell or tissue.

- Signaling networks are robust and adaptive, in part because of combinatorial complex formation (several building blocks for forming the same type of complex), redundant pathways, and feedback loops.



Egf stimulation of the Mitogen Activated Protein Kinase (MAPK) pathway.

Egf → EgfR → Grb2 → Sos1 → Ras → Raf1 → Mek → Erk

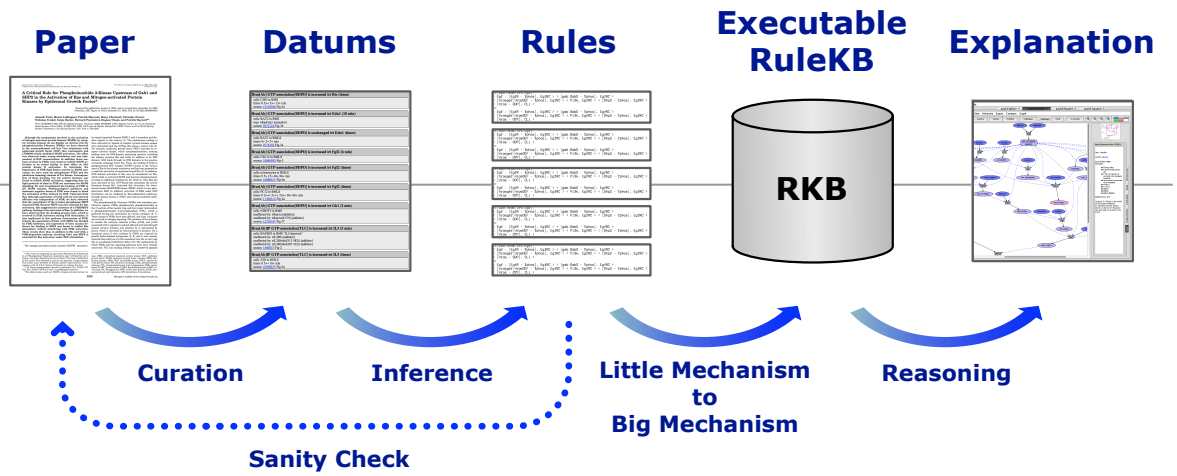
- Egf (EGF) binds to the Egf receptor (EgfR, EGFR) and stimulates its protein tyrosine kinase activity to cause auto-phosphorylation, thus activating EgfR.
- The adaptor protein Grb2 (GRB2) and the guanine nucleotide exchange factor Sos1 (SOS) are recruited to the membrane, binding to EgfR.
- The EgfR complex activates a Ras family GTPase by exchanging GDP for GTP.
- Activated Ras activates Raf1, a member of the RAF serine/threonine protein kinase family.
- Raf1 activates the protein kinase Mek (MEK), which then activates Erk (MAPK) ...

Challenges

- Naming
  - different biologists use different names for the same protein Egf vs EGF, Erk vs MAPK, EgfR vs ErbB1 vs HerbB1
  - solution: link name to `standard' source: UniProt, KEGG, HUGO ...
- Activity / state -- a protein may need to be is a specific state (active) to carry out its function. Activity is a function, what modification, context leads to activity?
- Location -- what compartment, where in the compartment?
  - media -- outside a cell
  - associated to a cell: membrane, cytoplasm, nucleus, endoplasm …

Pathway Logic (PL) as a Rewrite Theory

PL from 1k feet



Key components

- Representation system
  - controlled vocabulary
  - datums (formalized experimental results)
  - rules
- Curated datum knowledge base (KB) and search tool
- Evidence based rule networks
  - STM, Protease, Mycolate, GlycoSTM
- Executable models
  - generated by specifying initial conditions and constraints
  - query using formal reasoning techniques
- PLA to visualize and browse rule networks

Roles of Maude in PL

- Controlled Vocabulary--Equational theory with *metadata*
- Rule Knowledge base (RKB) -- symbolic rules
- Inferring models -- Initial state + RKB -> Executable rule network
- Export
  - Interactive Graph
  - May I borrow your logic: export to LoLA for path finding
  - Export to JSON for dbquery/tool exchange
- Pathway Logic Assistant (PLA) -- Maude Actor

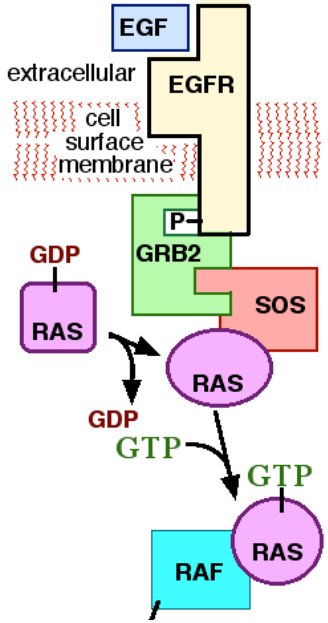
Pathway Logic Model Organization

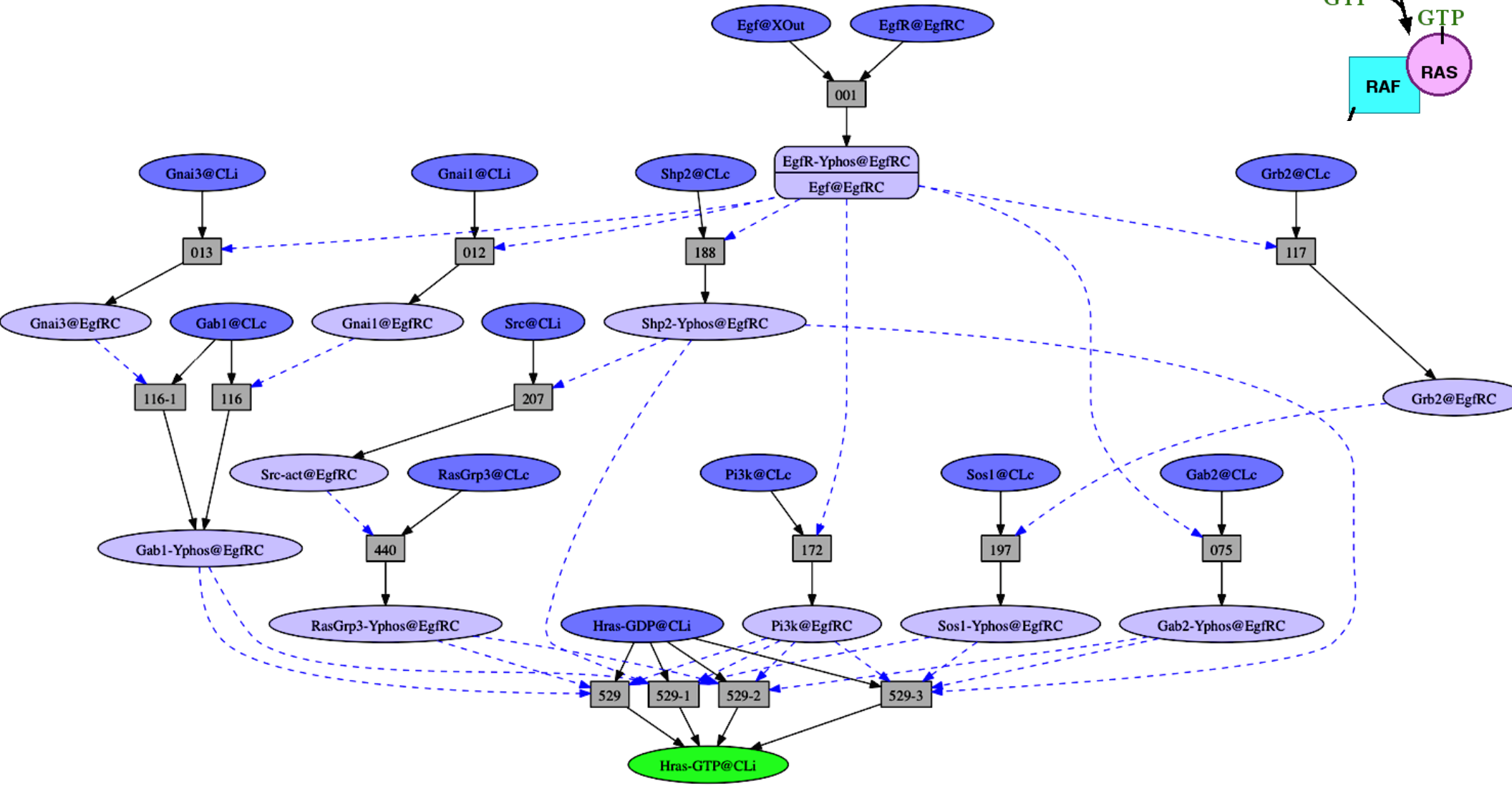
- Ops (Equational part)
- Rules --- signal transduction reactions
- Dishes --- candidate initial states
- Key Sorts
  - BProtein (gene product), Peptide (AA+) < Protein (can have mods)
  - Chemical, Gene, ….
  - Modification / ModSet
  - LocName — cellular location, named transient complexes
  - Occ / OccSet — entity instances (modifications,location)

Running example.

The subnet of the Egf model for activating **Hras**.

(Exchanging GDP for GTP.)





Sorts and ops

```
sort HrasSort .
subsort HrasSort < RasS < BProtein .

op Hras : -> HrasSort [ctor metadata (  (spnumber P01112)  (hugosym HRAS)
  (synonyms "GTPase HRas"
            "Transforming protein p21"
            "v-Ha-ras Harvey rat sarcoma viral oncogene homolog"
            "Harvey murine sarcoma virus oncogene"
            "H-Ras-1"
            "c-H-ras"
            "HRAS1"
            "RASH1"
            "RASH_HUMAN"))] .

op Rass : -> RasS [ctor metadata (  (category Family)  (members Hras Kras Nras))] .

op Pi3k : -> Composite [ctor metadata "(
  (subunits Pik3cs Pik3rs)
  (comment  "PI3 Kinase is a heterodimer of:"
            "a p110 catalytic subunit: Pik3ca, Pik3cb, Pik3cd or Pik3cg"
            "a p85 regulatory subunit: Pik3r1, Pik3r2, or Pik3r3"))] .
```

Modifications and LocNames

```
sorts Site Modification ModSet .
subsort Modification < ModSet .
op __ : AminoAcid Nat -> Site .
op [_-_] : Protein   ModSet -> Protein   [right id: none ] .

op phos :       -> Modification .      *** phosphorylated
op phos : Site  -> Modification .      *** phosphorylated on a
op Yphos : -> Modification .           *** phosphorylated on Tyrosine
op GDP : -> Modification .             *** bound to GDP
op GTP : -> Modification .             *** bound to GTP

Examples:  [Hras - GTP]  [Gab1 - Yphos]  [Gab1 - phos(Y 627)]


sorts LocName  CompName .
subsort CompName < LocName .
op CLm : -> LocName [ctor metadata "((definition \"Plasma Membrane\"))"] .
op CLi : -> LocName
    [ctor metadata "((definition \"Stuck to the inside of the plasma membrane\"))"] .
op EgfRC : -> CompName [ctor] .    *** EgfR complex
```

Examples: Occurrences

An occurrence models an entity such as a protein with specific modifications, in a particular location. (These will become places in a petri net representation.)

```
sorts SimpleThing Composite Complex Thing .
subsort Protein  Composite Chemical < SimpleThing .
subsort Complex SimpleThing < Thing .

sorts Occ Occs .
subsort Occ < Occs .   **** multiset of Occ
op <_,_> : Thing Loc -> Occ .


sort Dish .
op PD : Occs -> Dish [ctor] .

HrasDish:  experimental setup to study activation of Hras
PD(< Egf,XOut > < EgfR,EgfRC >
     < Gab1,CLc > < Gab2,CLc > < Gnai1,CLc > < Gnai3, CLc >
     < Grb2,CLc > < Pi3k,CLc > < Rasa1,CLc > < RasGrp3,CLc >
     < Shp2,CLc > < Sos1,CLc > < Src,CLc >
    <[Hras - GDP],CLi >)
```

Rule Knowledge Base (RKB) — A Rewrite Theory

- PL rules describe local change and specify required context

rl[529.Hras.irt.Egf]:

< Egf : [EgfR - Yphos], EgfRC > < [gab:GabS - Yphos], EgfRC >

< [hrasgef:HrasGEF - Yphos], EgfRC > < Pi3k, EgfRC > < [Shp2 - Yphos], EgfRC >

< [Hras - GDP], CLi >

=>

< Egf : [EgfR - Yphos], EgfRC > < [gab:GabS - Yphos], EgfRC >

< [hrasgef:HrasGEF - Yphos], EgfRC > < Pi3k, EgfRC > < [Shp2 - Yphos], EgfRC >

< [Hras - GTP], CLi >

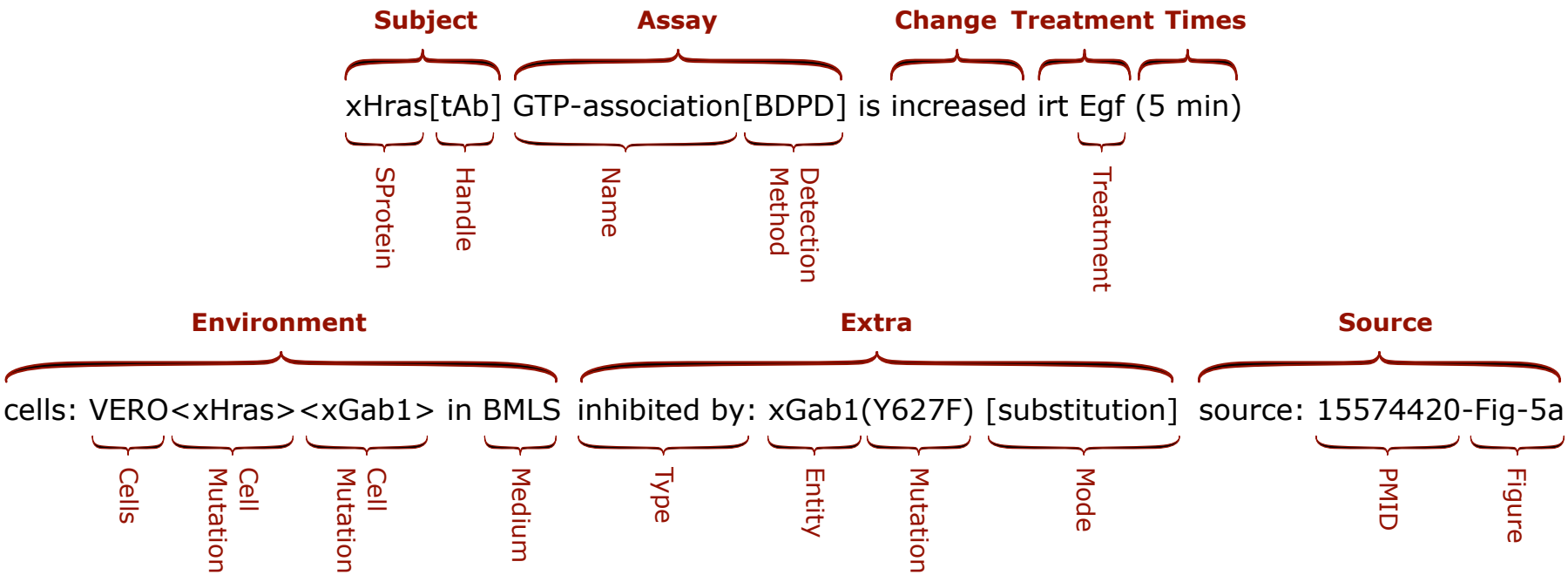
*** ~/evidence/Egf-Evidence/Hras.irt.Egf.529.txt

- Symbolic rules represent a family of rules using sorted variables
- EgfRC is the location of the Egf Receptor complex, it is populated in response to the Egf signal. CLi is the membrane interior
- gab:GabS is a variable standing for Gab1 or Gab2, hrasgef:HrasGEF is a variable for any of several HrasGEFs (enzymes to exchange GDP for GTP)

Where do rules come from?

- They are inferred from experimental findings.
- These are collected using a formal data structure call datums
  - datums are available in text (readable) or json (computable)
- The datum below says that the amount of GTP bound to Hras is increased after addition of Egf (Epidermal Growth Factor) to VERO cells for 5 minutes.

The Elements of a Datum



Inferring the Hras rule: the basic pattern

The `first line' of the previous Hras datum:

xHras[tAb] GTP-association[BDPD] is increased irt Egf (5 min)

can be represented by a rule pattern:

EgfTC C < [G - gmods act], Lg > < [Hras - GDP pmods], CLi >

=>

EgfTC C < [G - gmods act], Lg > < [Hras - GTP pmods], CLi >

- EgfTC is the treatment complex formed when Egf binds to the Egf Receptor
- G is a variable ranging over Hras GEFs, representing the general knowledge that exchange of GDP for GTP requires a GEF (Guanine exchange factor).
- gmods, pmods are variables indicating that we don't know the exact state of G or Hras.
- C is a variable standing for currently unknown requirements

Inferring the requirement for a Gab

The extra from the Hras datum

inhibited by: xGab1(Y627F) [substitution]

says that some function of Gab1 that relies on Y627 is required.
A plausible conjecture is that phosphorylation on Y627, or simple Yphos is required. I.e. [Gab1 - Yphos] is in the context C.

Inferring that Sos1 is a candidate GEF

The datum

rHras GDP-dissociation[3H-GDP] is increased by xSos1[tAb]IP
cells: none
IPfrom: HEK293 in BMS
source: 15039778-Fig-2c

reports direct GEF action of Sos1 in a test tube,

while the datum

xHras[tAb]IP GTP-association[TLC] is increased itpo xSos1
cells: HEK293 in BMS
source: 10896938-Fig-1c

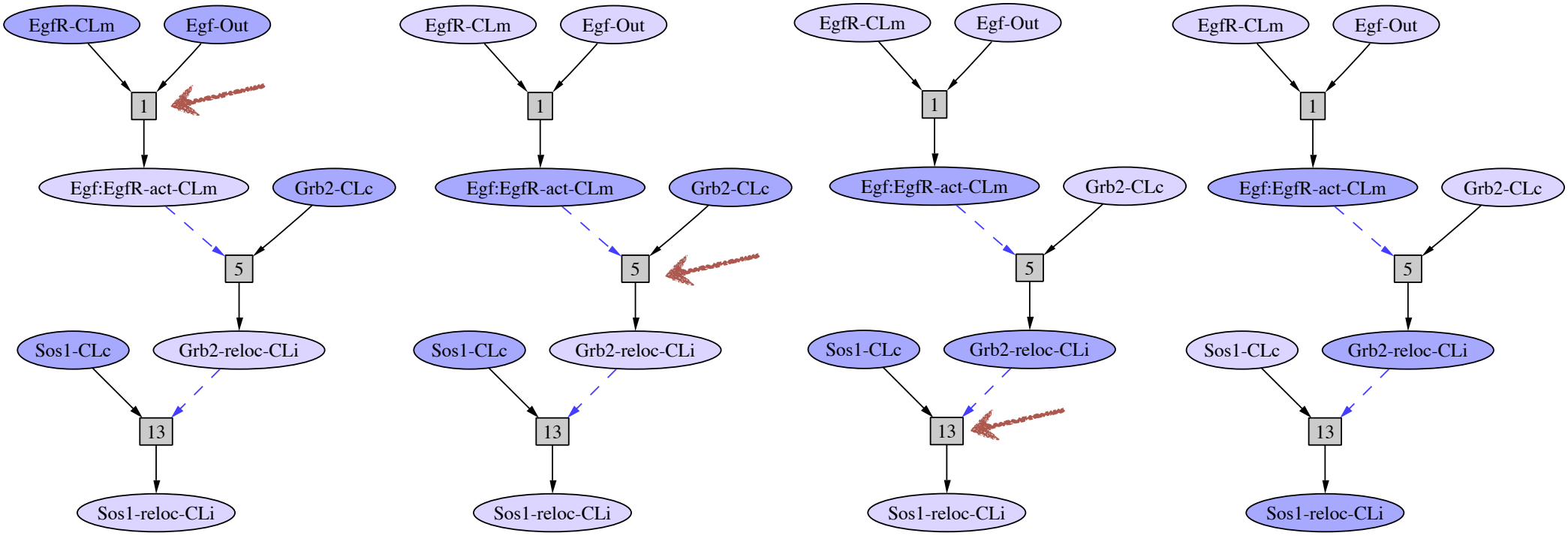
reports interaction in a live cell.

The combination tells us that Sos1 is a candidate GEF for Hras.

Where do the graphs come from?
What do they mean?

The Pathway Logic Assistant
(PLA)

PLA represents (concrete) networks as Petri Nets. They can be executed and model checked.



Sos1Dish =rule1=> Sos1Dish1 =rule5=> Sos1Dish2 =rule13=> Sos1Dish3

Ovals are occurrences -- biomolecules in locations (aka places).
Dark ovals are present in the current state (marked).
Squares are rules (aka transitions).
Dashed edges connect components that are not changed.

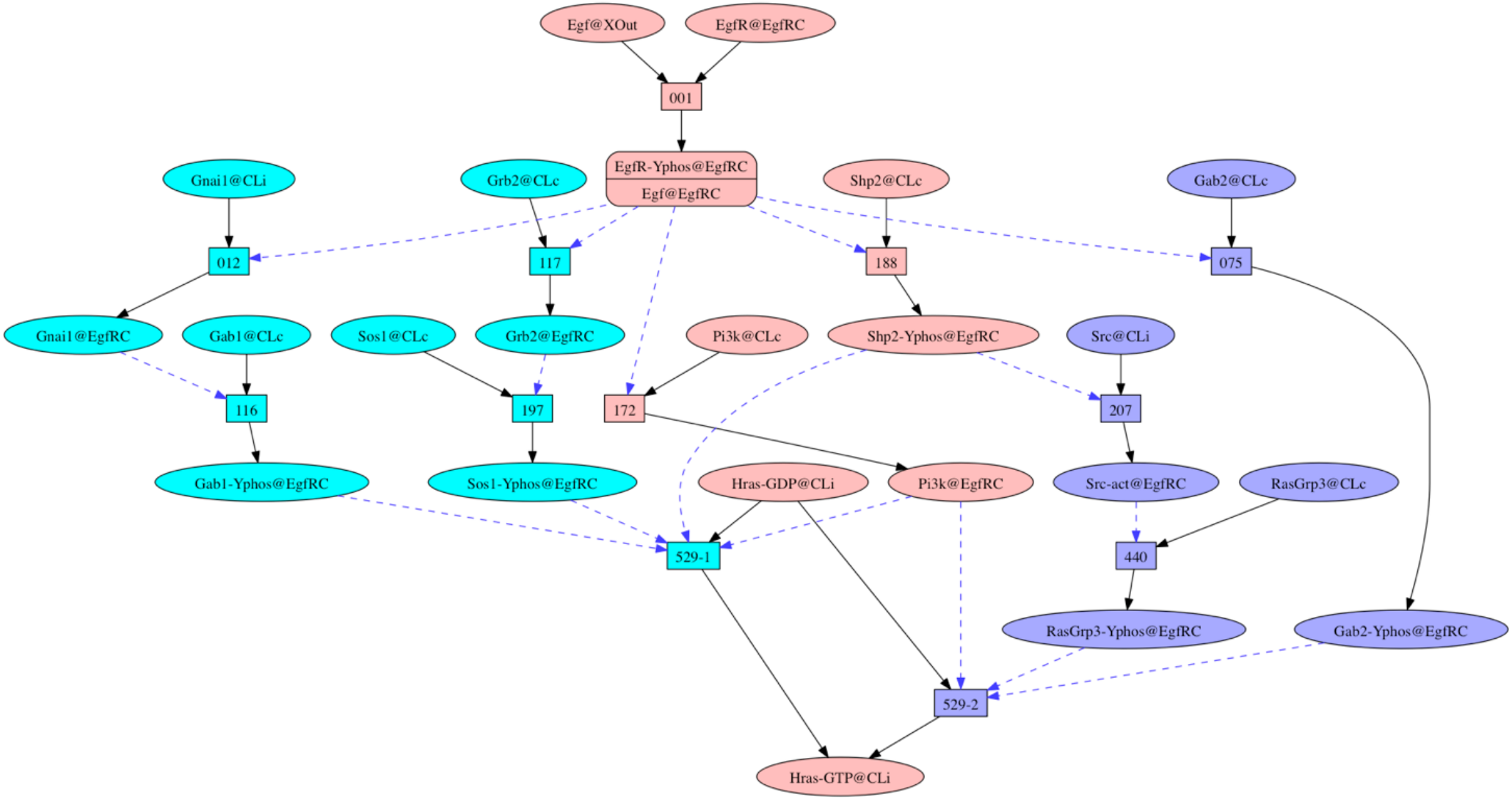
Pathway Logic Assistant: Features

- Derive the network of rules modeling response to a stimulus
  - STM Model: CD40 Csf1 **Egf** Igf1 IL1 Ins LPS Tgfb Tnf and more
- Find Erk activated in the picture (where is waldo?)
- Compute (and display) subnet activating Erk,
- Compare subnets
- Find an execution Path
- Explore -- what is upstream/downstream of X
- All Paths analysis — essential rules, knockouts, uses

PLA uses Symbolic analysis

- Forward reasoning generates a model of Egf signaling, the network of rules reachable from the EfgDish
- Backward reasoning generates the network of all execution paths leading to activation of Erks in the Egf model
- Model checking finds a specific execution pathway activating Erks in the Egf model, by asserting that there is no path and finding a counter example.
- AllPaths analysis (specific to Petri Nets) finds 6 ways to reach [Hras - GTP] and thousands of paths leading to activation of Erks!

Comparing two pathways in the Hras subnet, found by model checking.

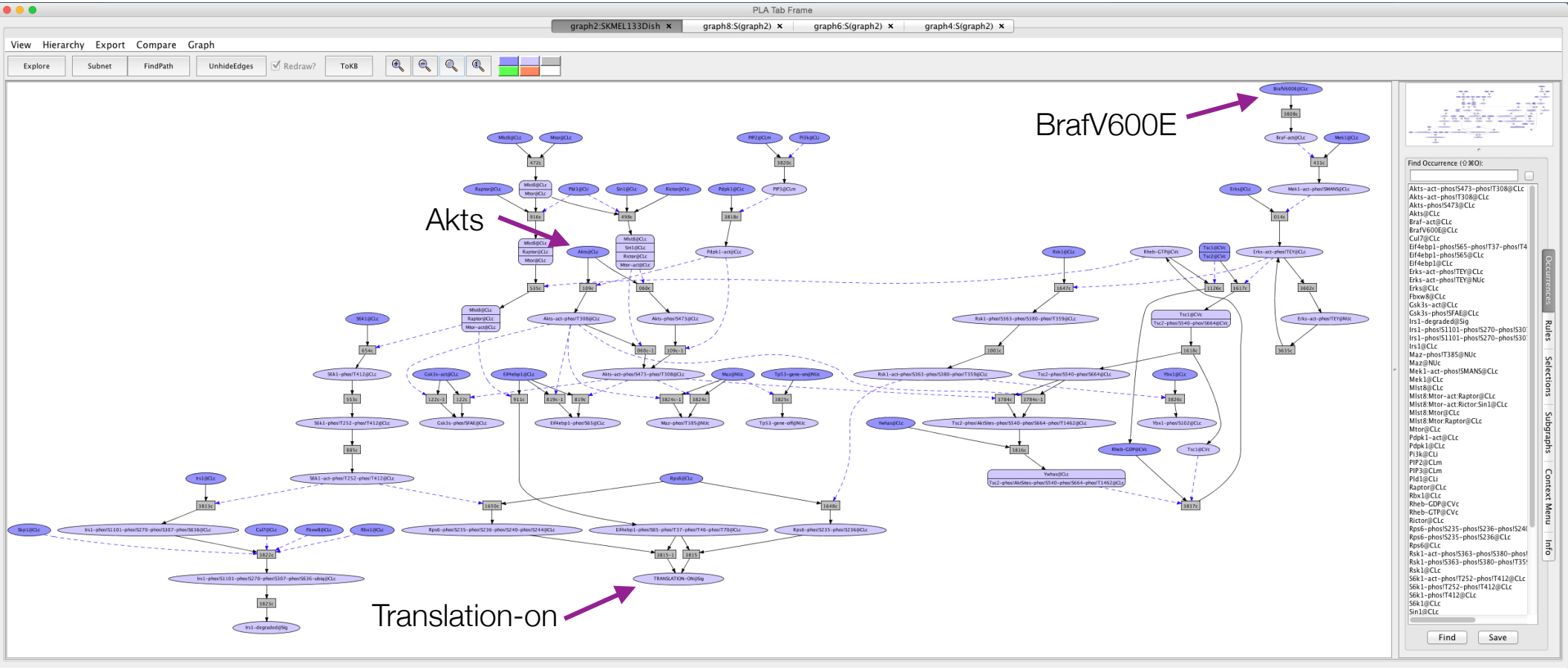


Pink: both pathways, cyan. blue different pathways.

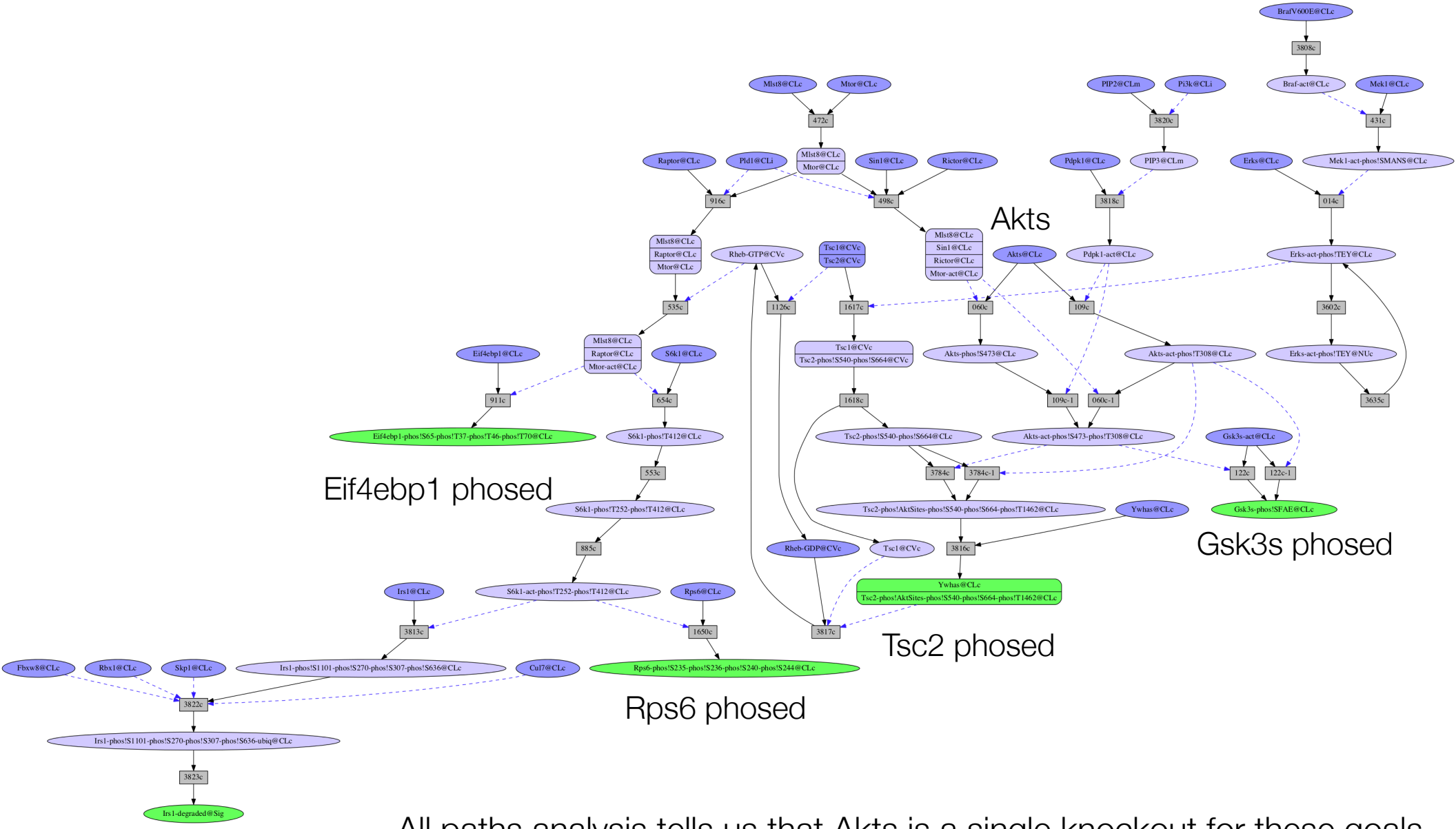
Explaining drug effects

SKMEL133 cells treated with an Akt inhibitor show decrease in specific phosphorylated forms of Eif4ebp1, Gsk3s, Tsc2, Rps6 and increase in Irs1.

PL model of aspects of SKMEL133 cells



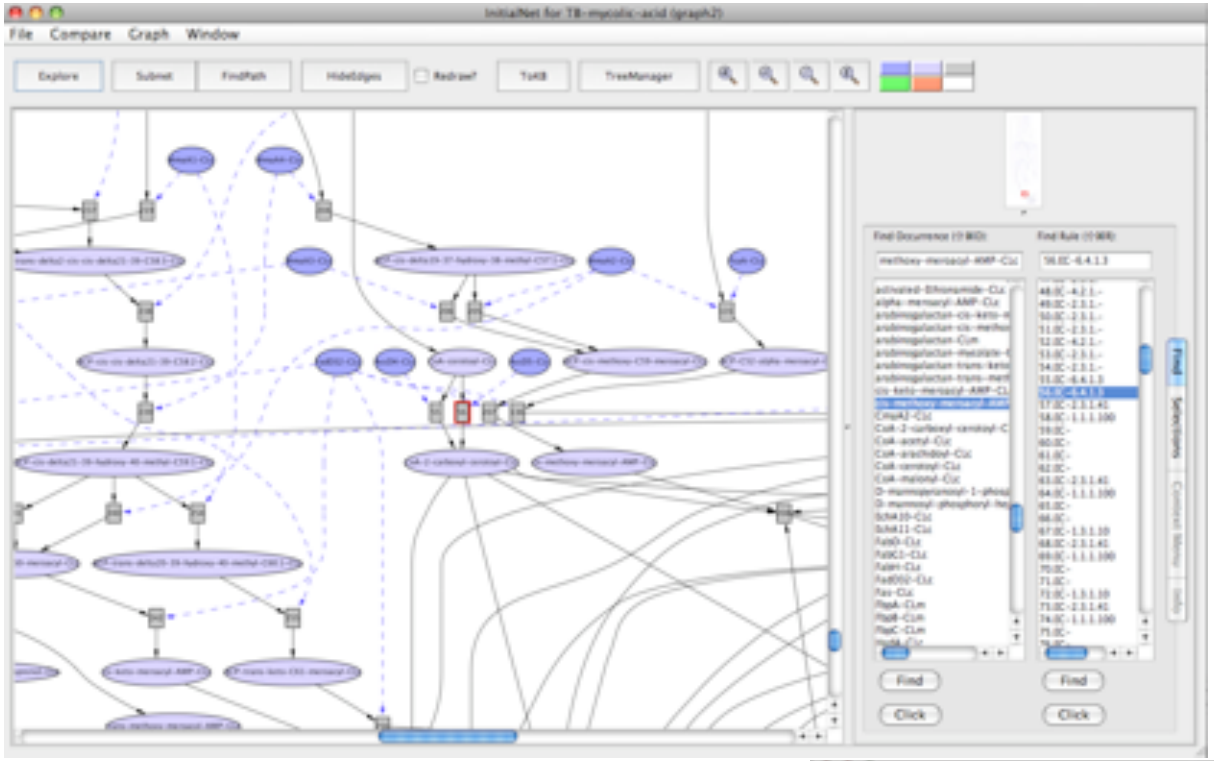
Subnet for occurrences down regulated by an Akt inhibitor



All paths analysis tells us that Akts is a single knockout for these goals

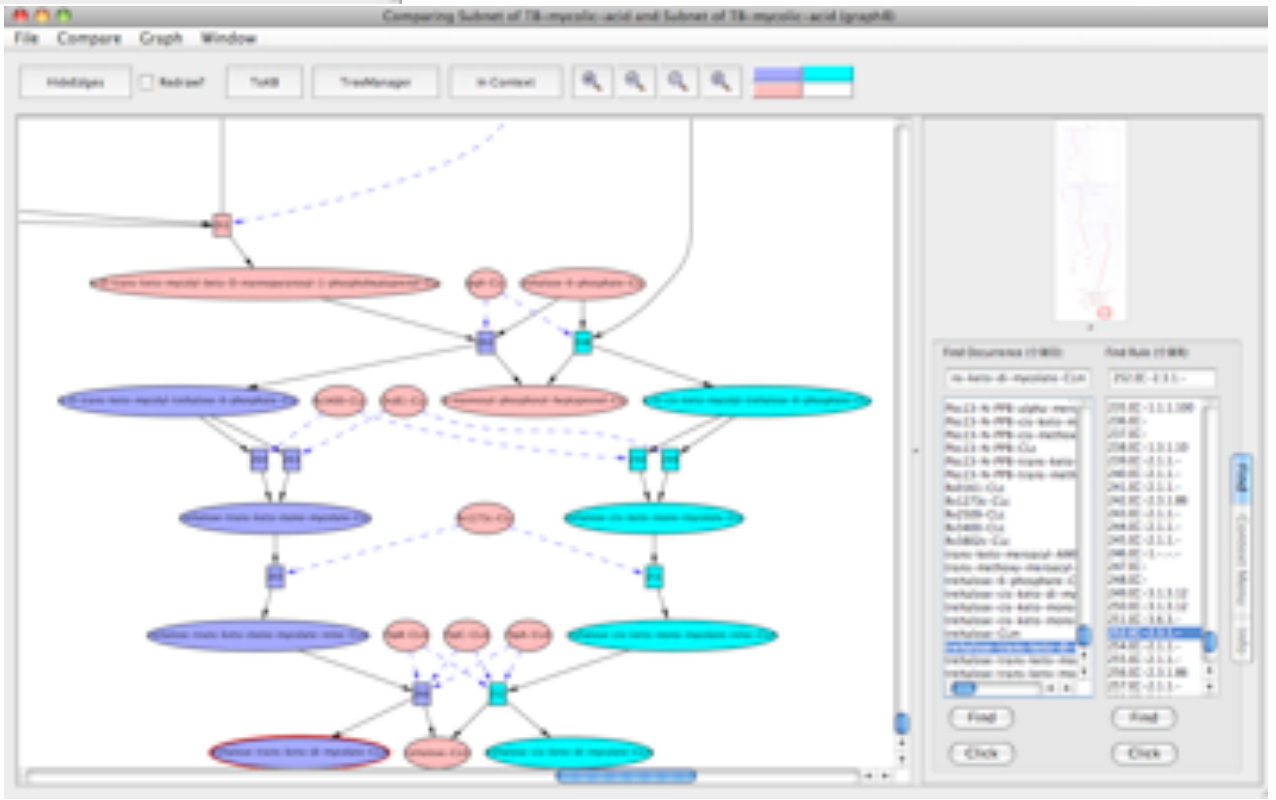
Beyond STM

- Metabolism -- Mycolic acid synthesis
- Glycosylation
- Immune system

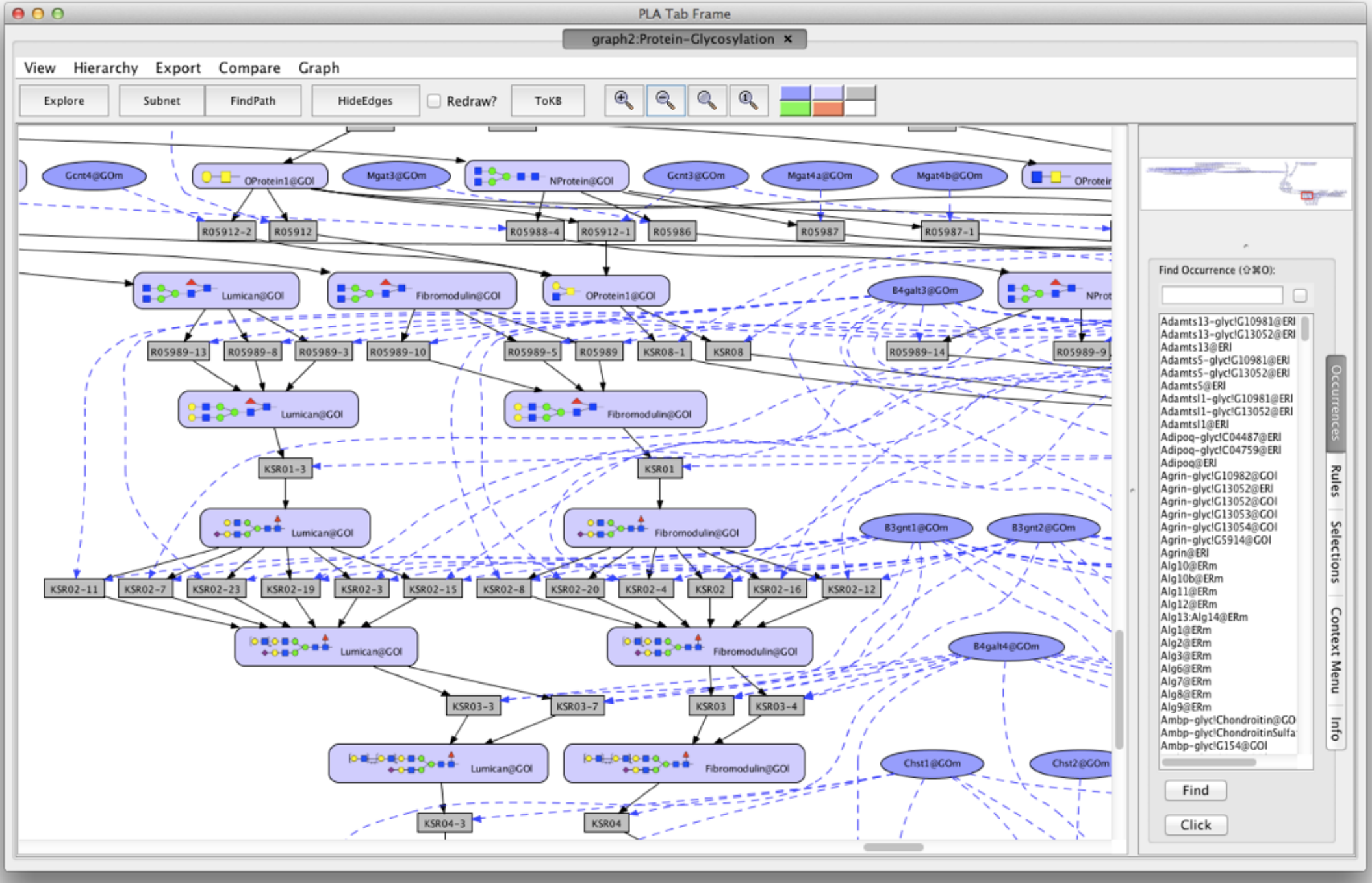


TB-mycolic-acid synthesis (Malabika)

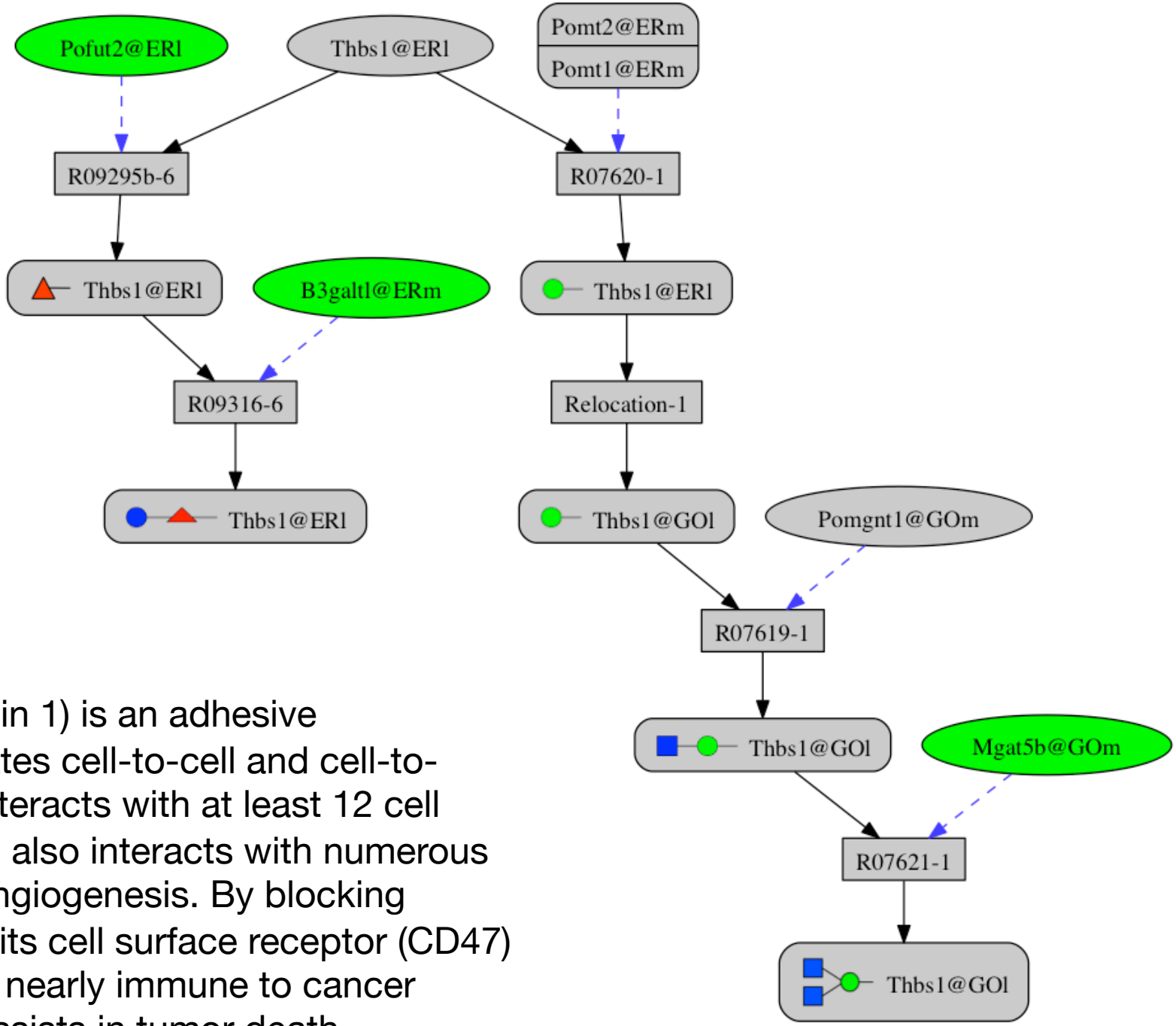
Multiple synthesis routes



Protein Gycosylation
KEGG + Uniprot



Exploring Gycosylation of TBS1

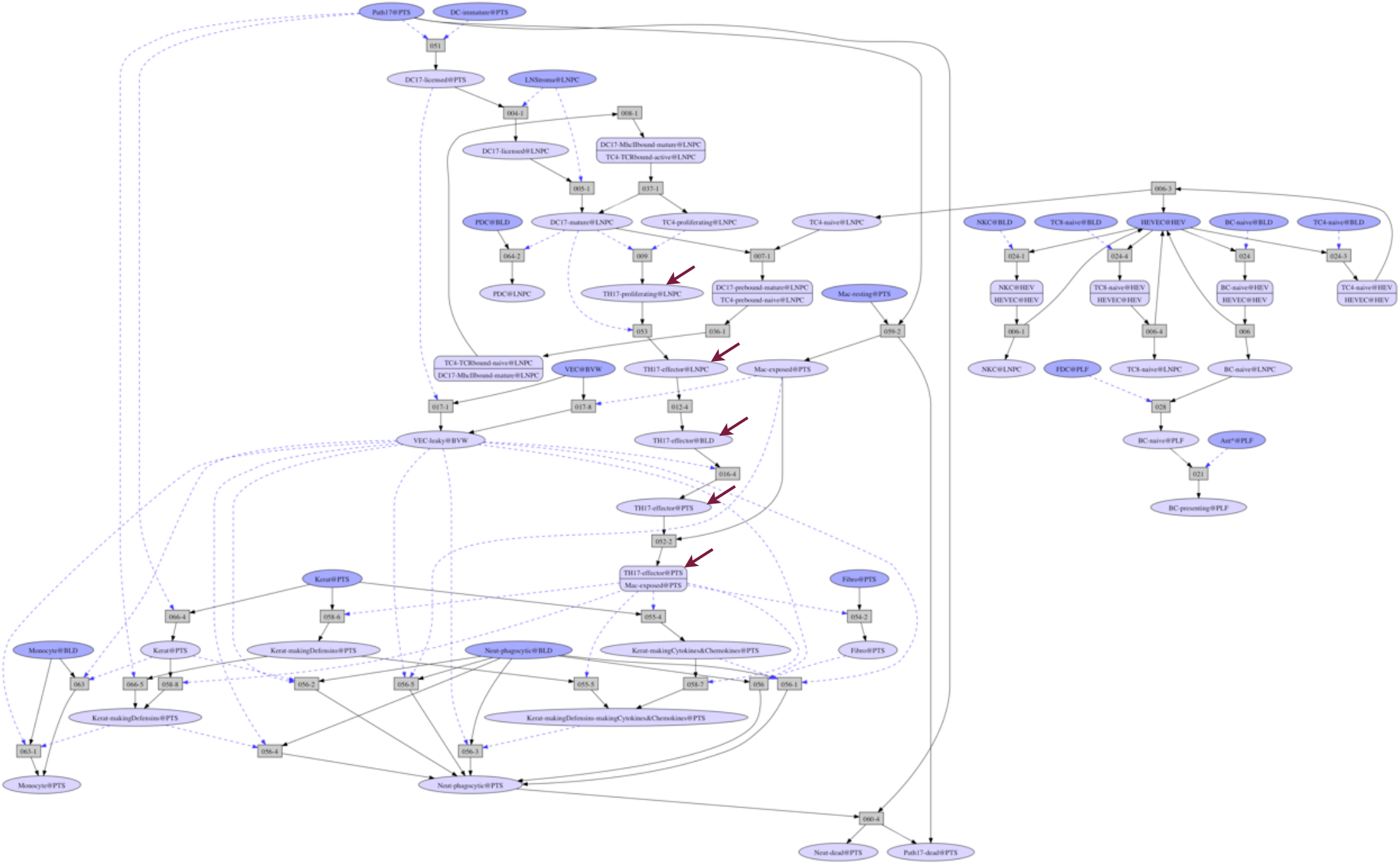


THSP1 (Thrombospondin 1) is an adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. It interacts with at least 12 cell adhesion receptors and also interacts with numerous proteases involved in angiogenesis. By blocking THSP1 from binding to its cell surface receptor (CD47) normal tissue becomes nearly immune to cancer radiation therapy and assists in tumor death.

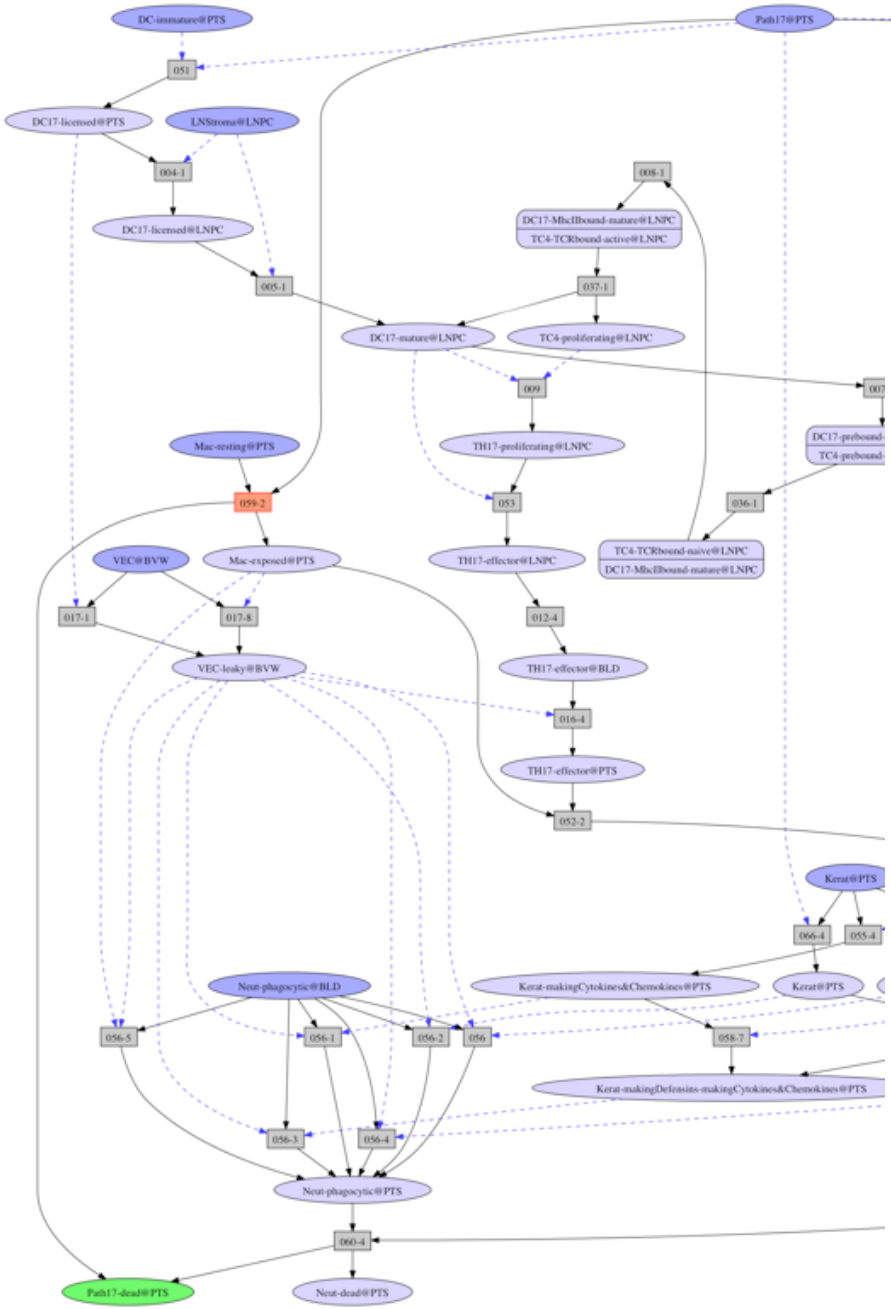
Model of the Immune System ala Janeway

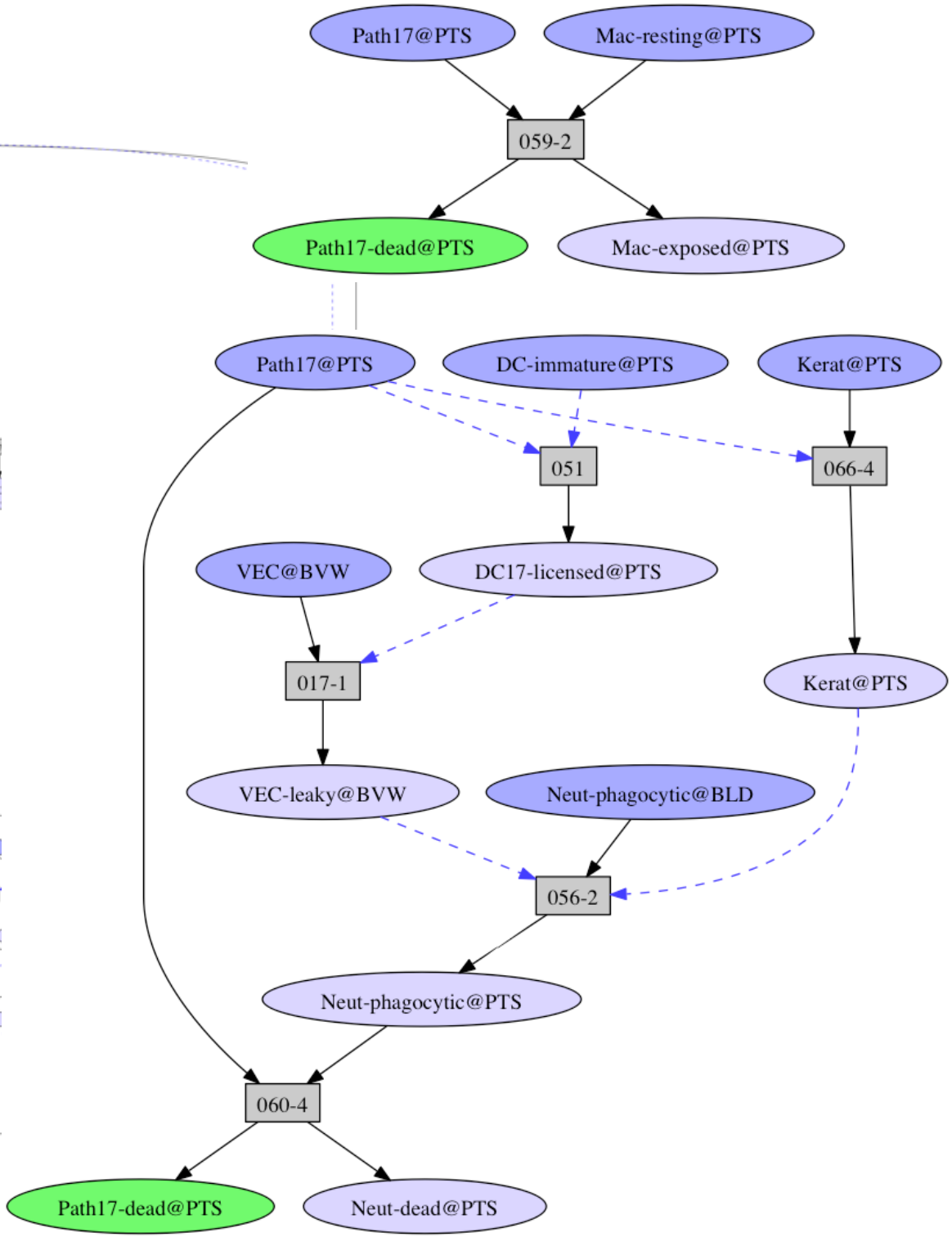
- Different classes of challenge
  - Listeria: Intracellular bacterium Listeria monocytogenes
  - Path1: Pathogens that lead to TH1
  - Path2: Pathogens that lead to TH2
  - Path8: Pathogens killed by Cd8 cytotoxic cells
  - Path17: Pathogens that lead to TH17
  - PathFH: Pathogens that lead to THfh
  - Virus
  - NoPathogen
- 258 rules

Response to Path17



Killing Path17





Summary

- Pathway Logic features include
  - formal representation of experimental evidence
  - formal links from symbols representing biomolecules to external reference resources
  - links from rules to supporting evidence.
  - assembly of models and pathways by symbolic reasoning
- Challenges
  - scaling — curation, analysis, visualization
  - rates and quantities
  - real multi-scale models

PL extended team (past and present)*

- Robin Donaldson (Glasgow/Stanford)
- Steven Eker
- Merrill Knapp
- Keith Laderoute
- Pat Lincoln
- Ian Mason
- Jose Meseguer
- Huaiyu Mi (USC, PANTHER)
- Anupama Panikkar (SV)
- Andy Poggio
- Malabika Sarker (DC)
- Carolyn Talcott
- Maneesh Yadav

*Alphabetical order

Thats all folks!