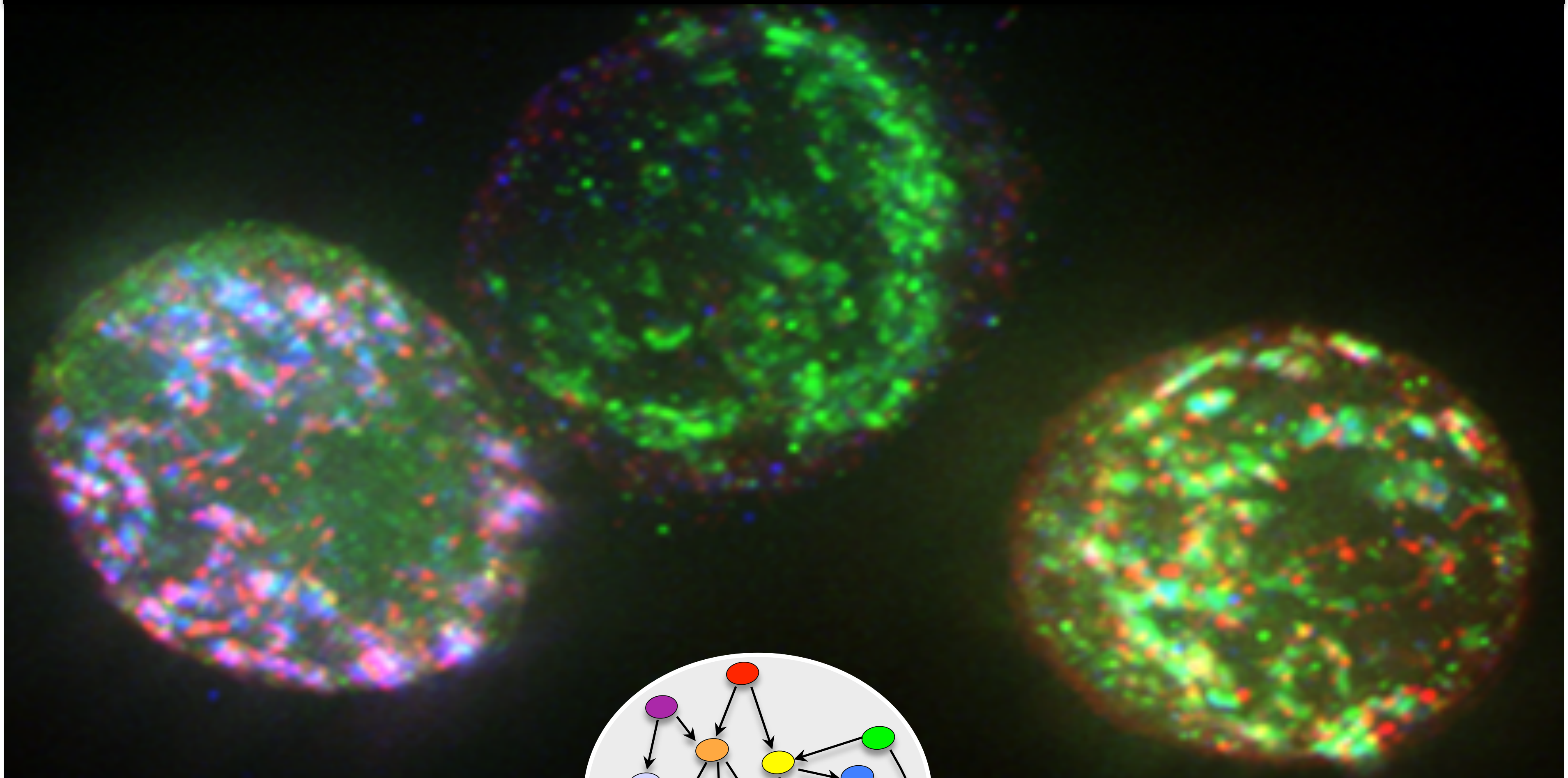
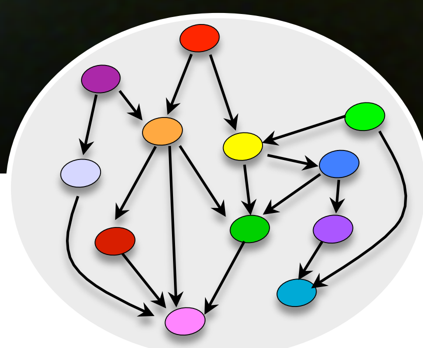
Causal learning of biomolecular networks

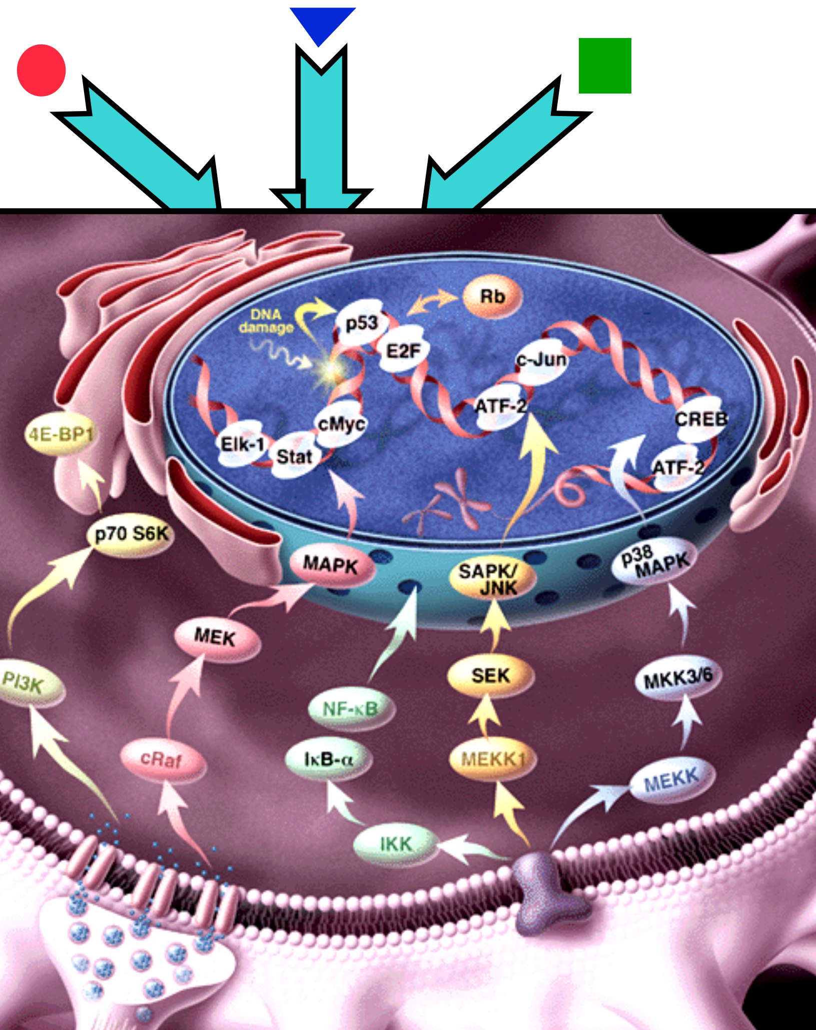




Karen Sachs

Stanford University

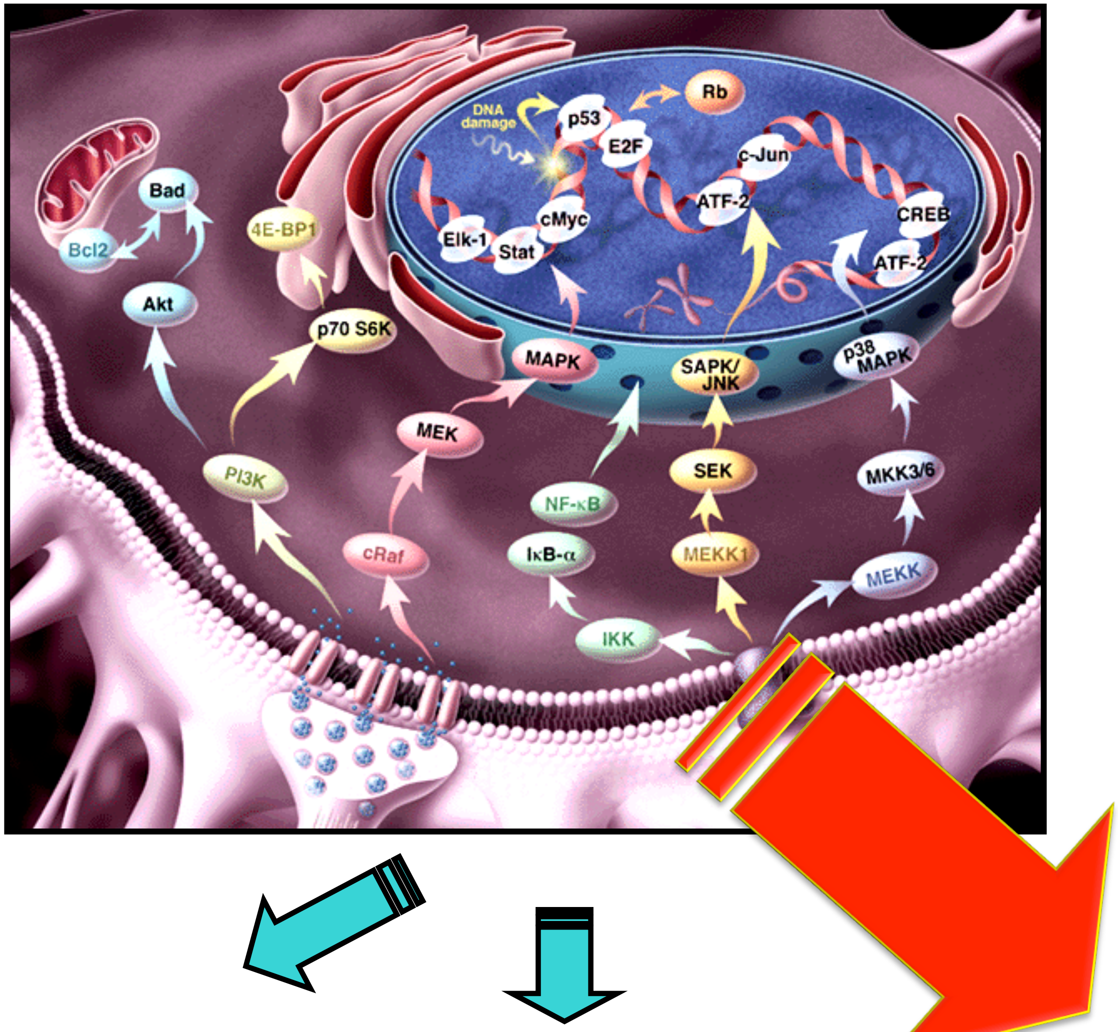
Cells respond to their environment



Secrete cytokines | Cell death | Proliferation

Inside each cell is a molecular network

..which breaks down in disease states

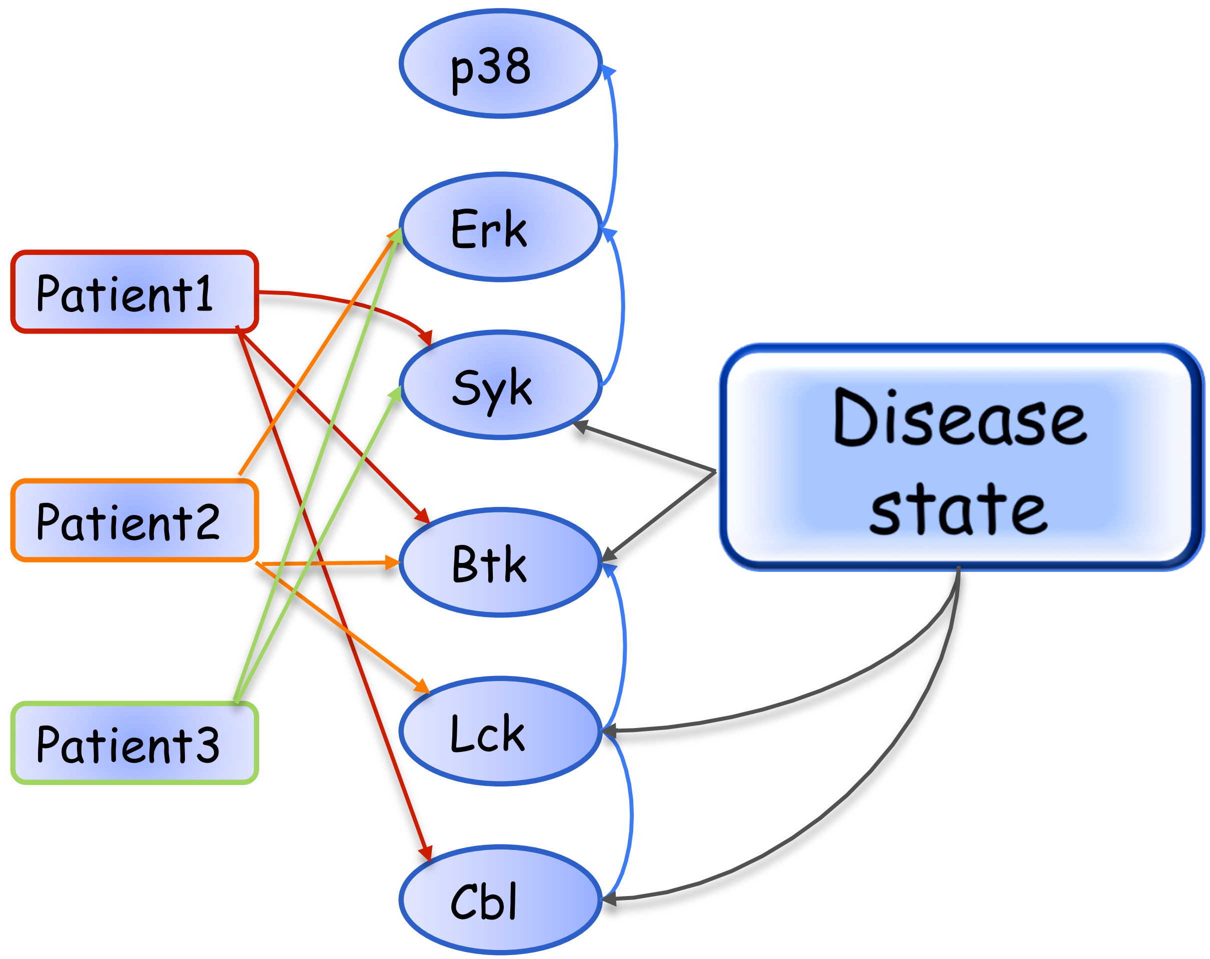


Secrete signals

Cell death

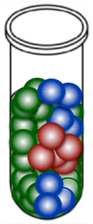
Proliferation

Motive: Characterize normal, disease, drug..



Sachs et al. IEEE

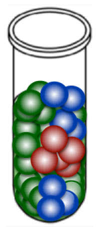
Causal learning in signaling



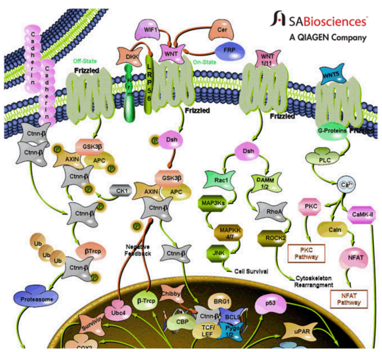
Where does data come from?

- Technology

Causal learning in signaling



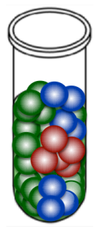
Where does data come from?



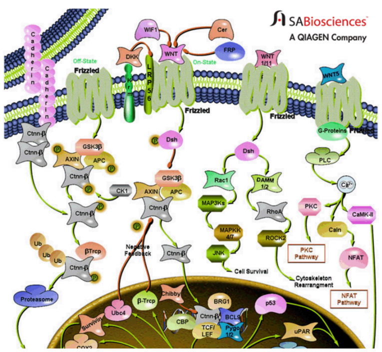
What causal connections appear?

- What happens?
- What can we see?

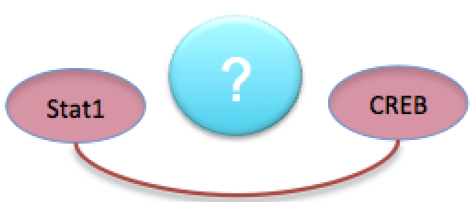
Causal learning in signaling



Where does data come from?



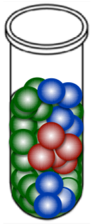
What causal connections appear?



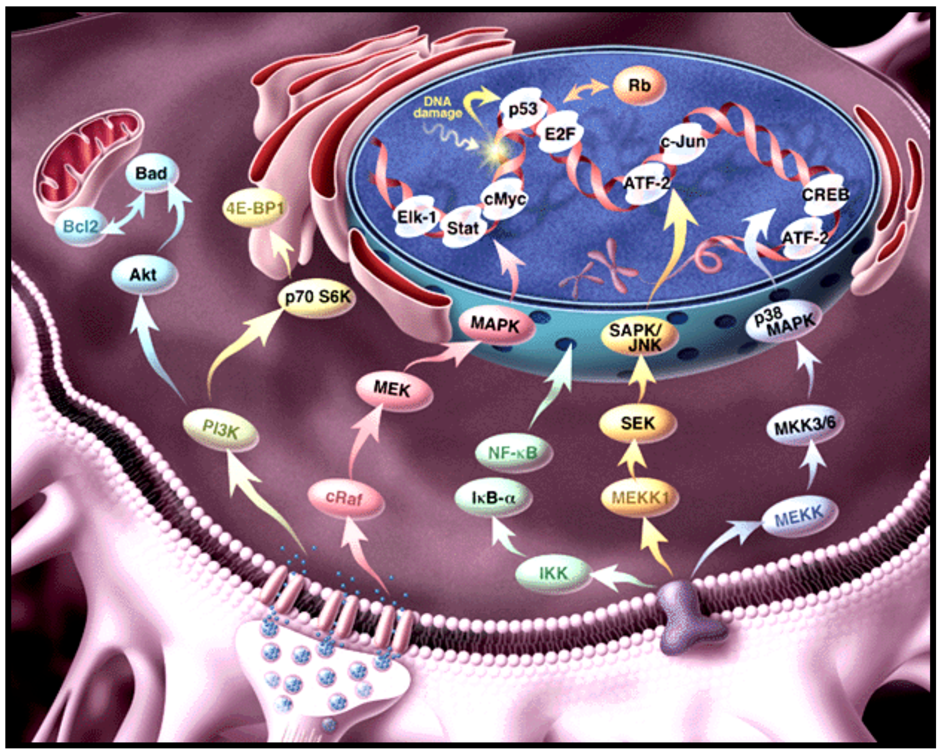
What is needed for causal learning?

- Outstanding challenges

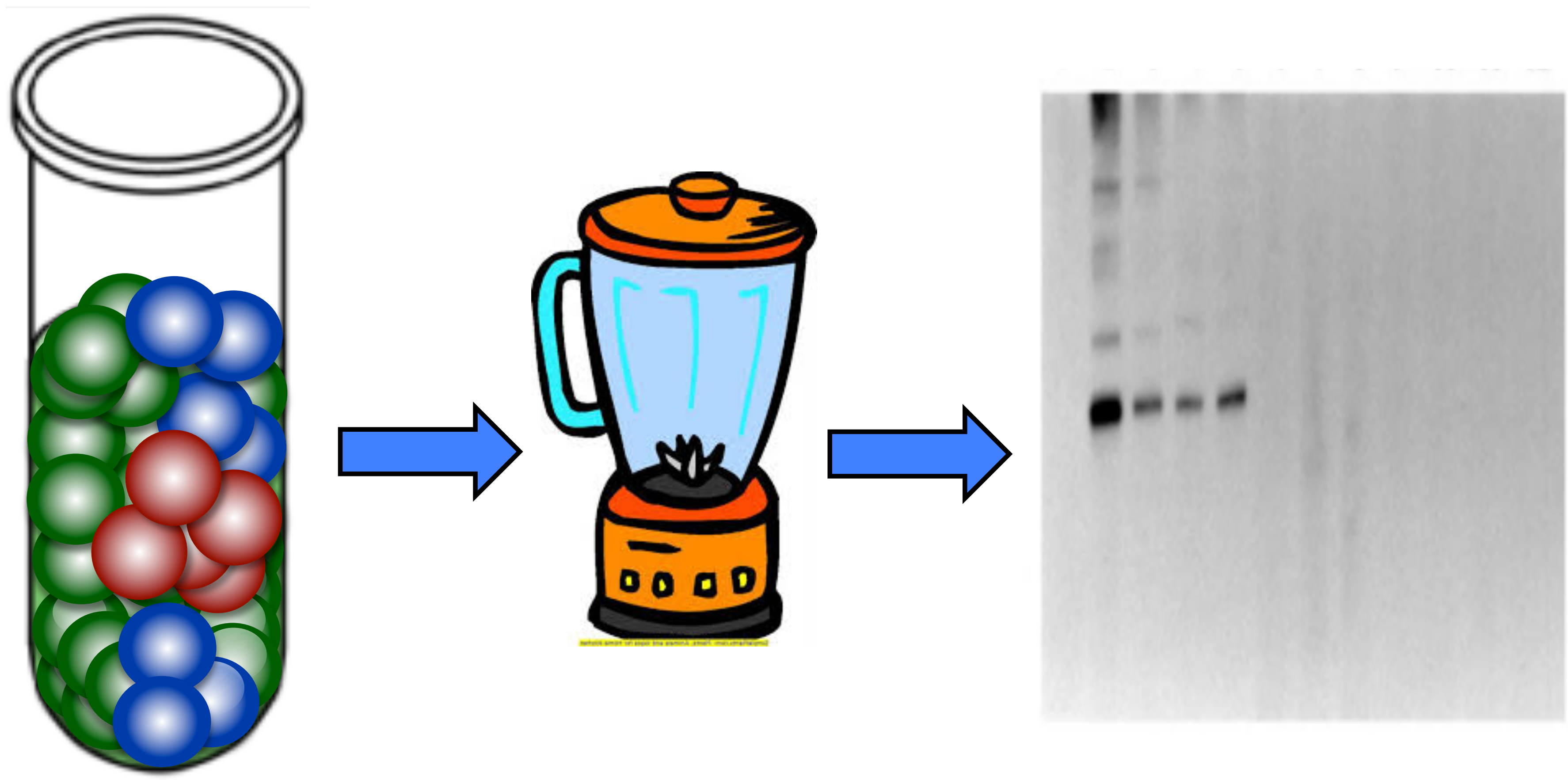
Causal learning in signaling



1. Where does data come from?



Samples are blended routinely



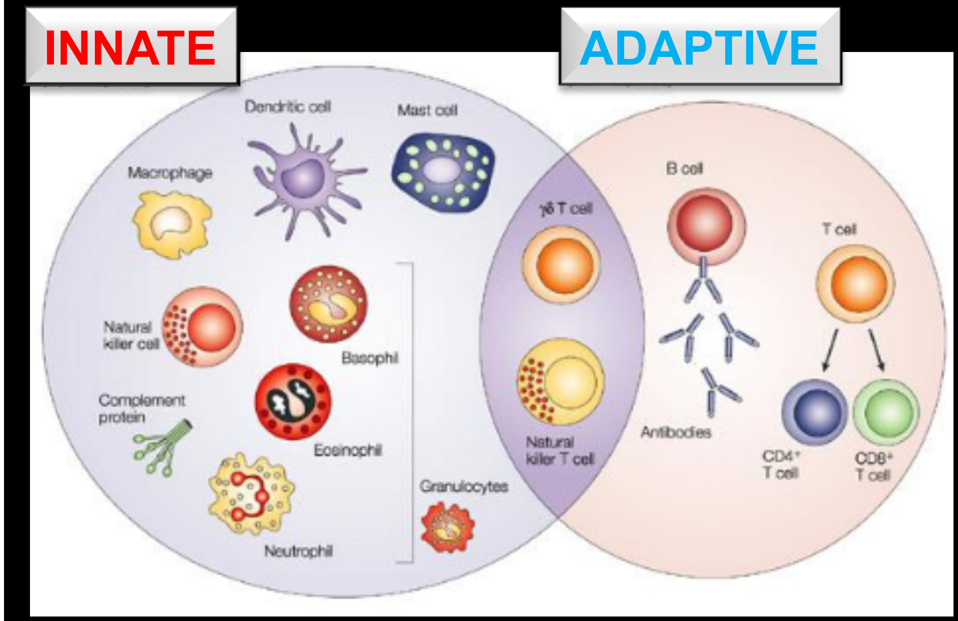
Tumor sample

Lab Blender!

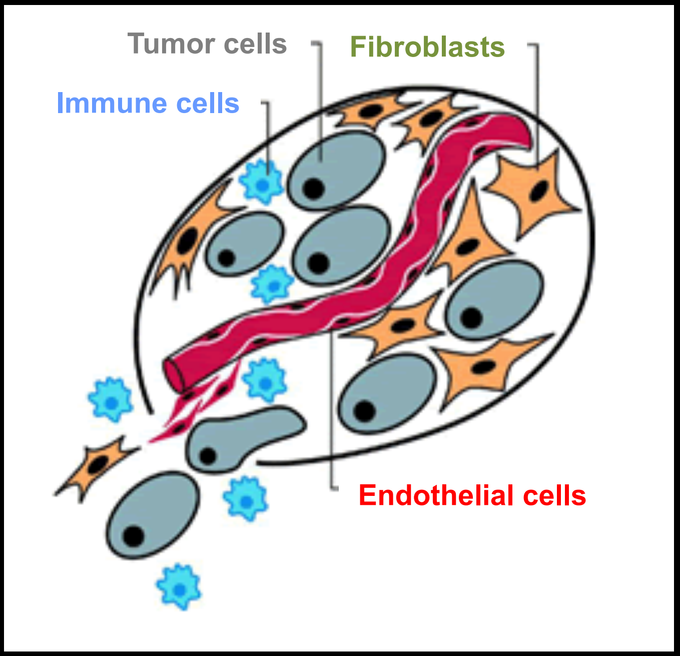
Biological measurements

Why single cell? (Biology perspective)

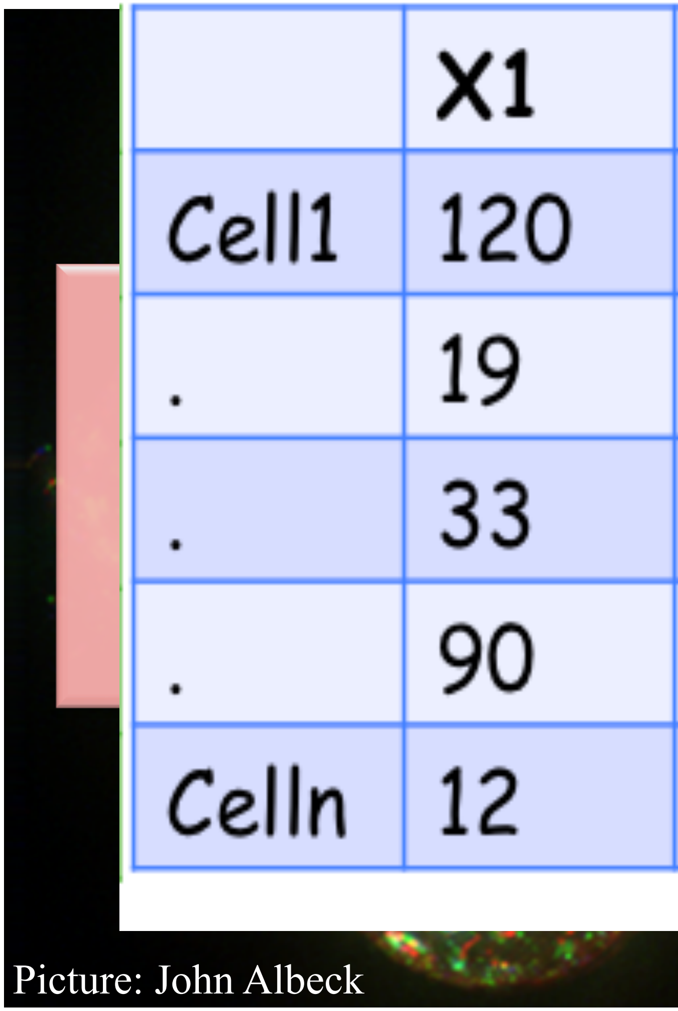
Innate and **adaptive** branches of the Immune **system** communicate with each other to mount an effective immune response



Cancer is a complex **system** with defined interdependent compartments



Why single cell? (Stats perspective)

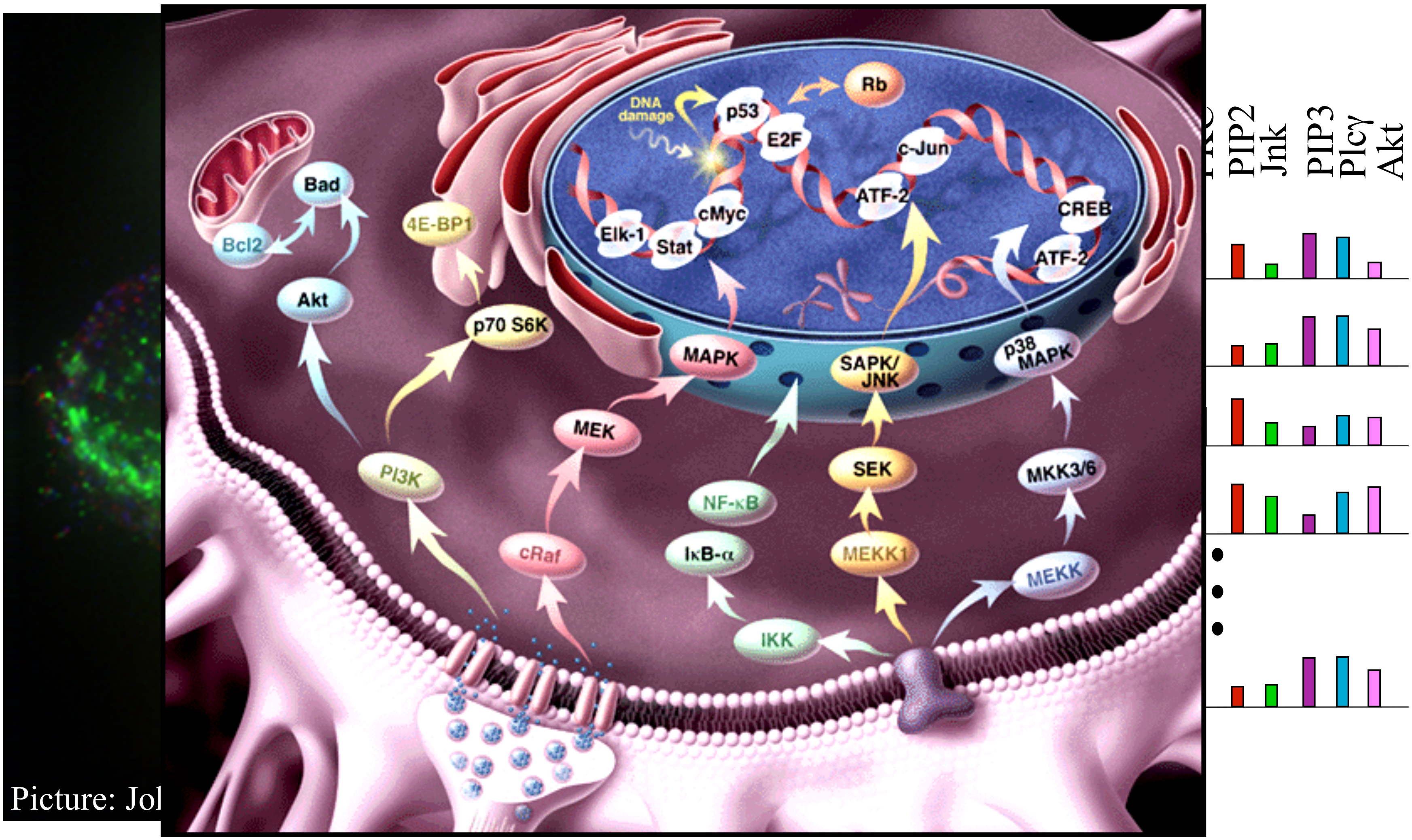


	X1	X2	X3	...		Xp
Cell1	120	...				77
.	19					98
.	33					3
.	90					40
Celln	12	...				34

Picture: John Albeck

High throughput data

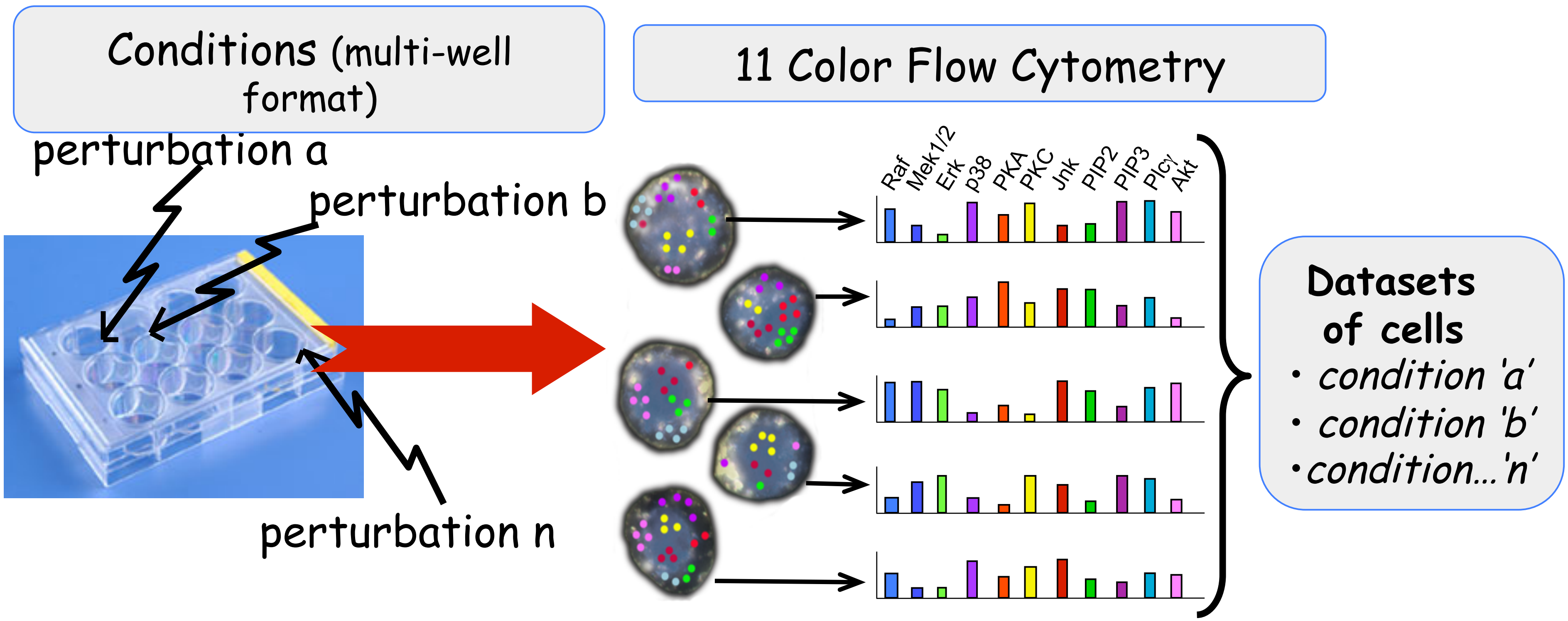
From Phospho-molecular profiling to Signaling pathways



High-throughput data

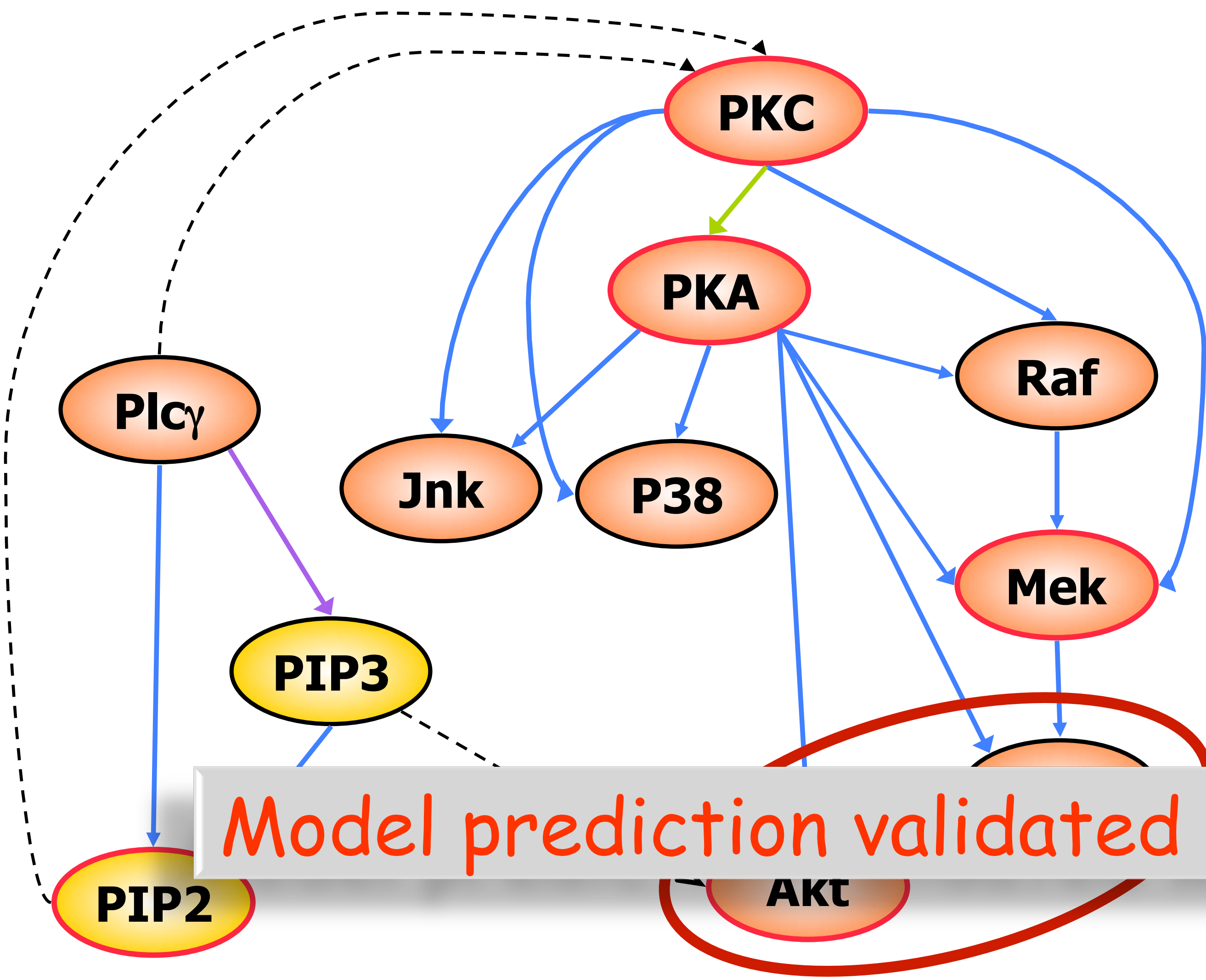
Signaling Pathways

T-Lymphocyte Data



- Primary human T-Cells
- 9 conditions
  - (6 **Specific** interventions)
- 9 phosphoproteins, 2 phospolipids
- 600 cells per condition
  - 5400 data-points

Accurate Network Inference



Phospho-Proteins
Phospho-Lipids
Perturbed in data

T Cells	15/17
Reported	17/17
Reversed	1
Missed	3

Model prediction validated with siRNA

[Sachs *et al*, *Science* 2005]

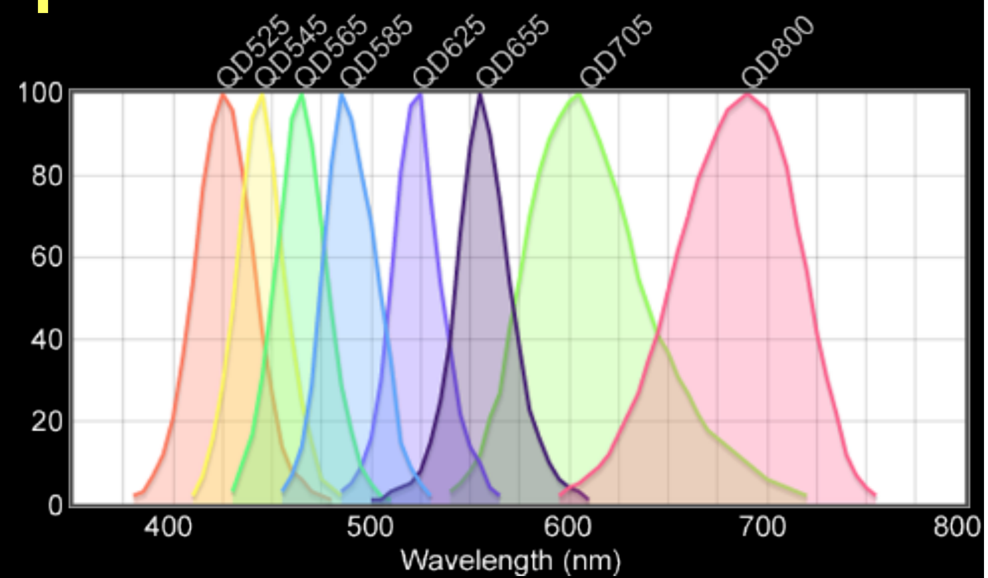
Flow Cytometry: *Single Cell Analysis*

Bendal et al, Science 2011

The Fluorescence Spectrum is Crowded

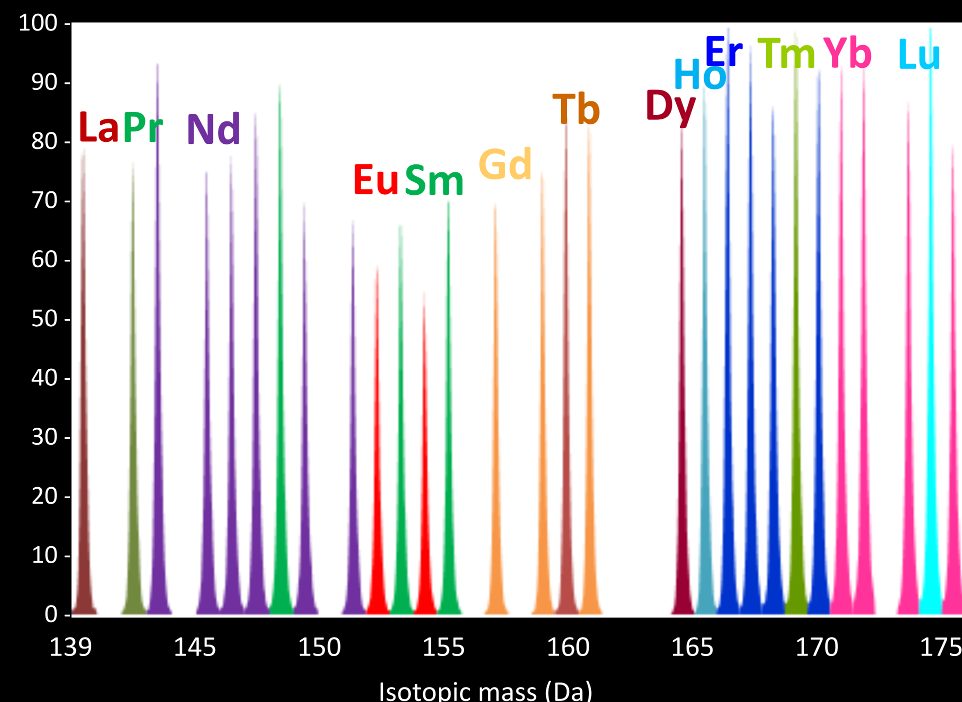
Fluorescent cytometry

- 6-8 parameters is “routine”
- 17 parameters has been reported
- Autofluorescence
- High background

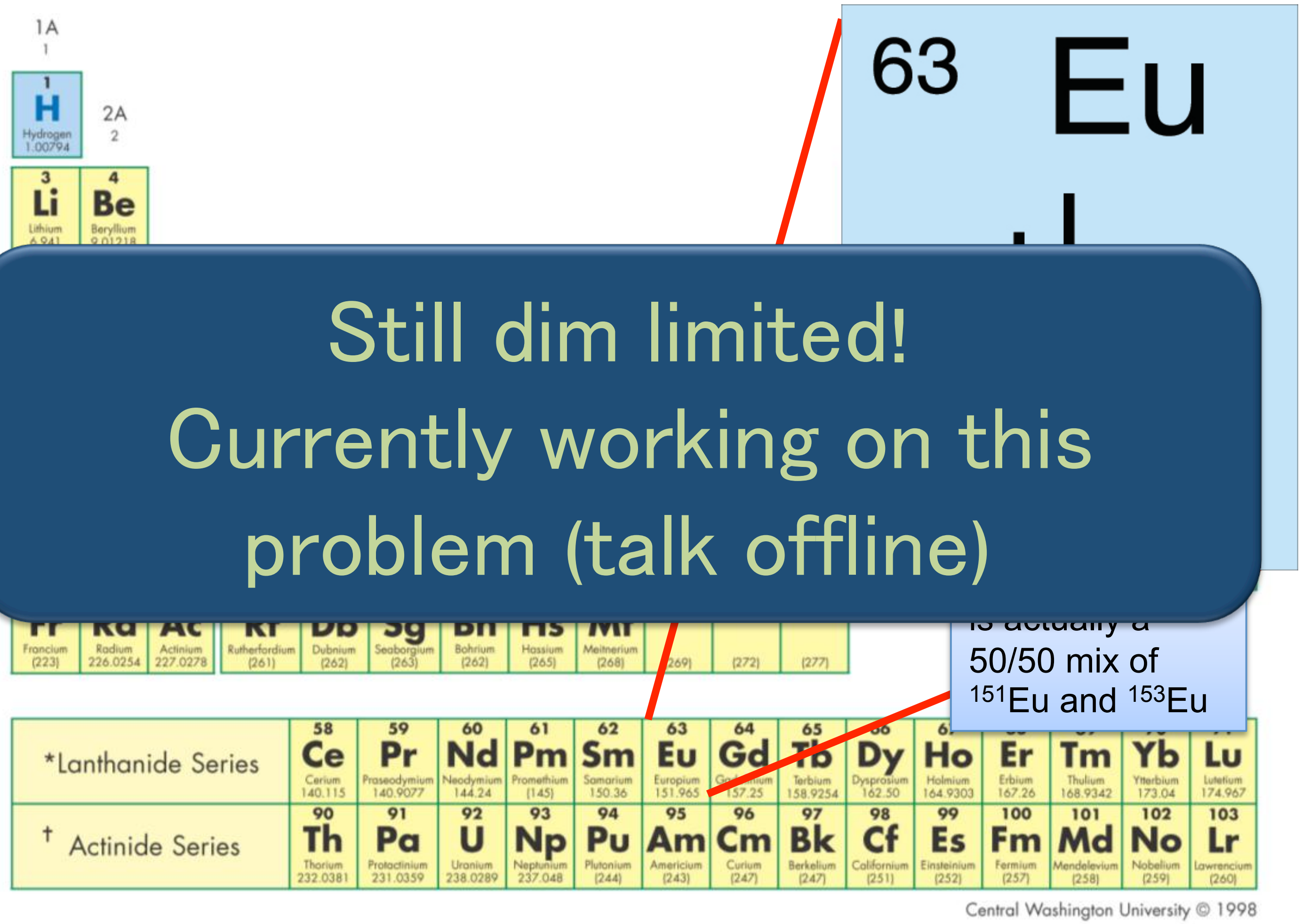


Mass cytometry

- 100 discrete mass channels
- 38 parameters easily (58 soon)
- No compensation required
- Zero background



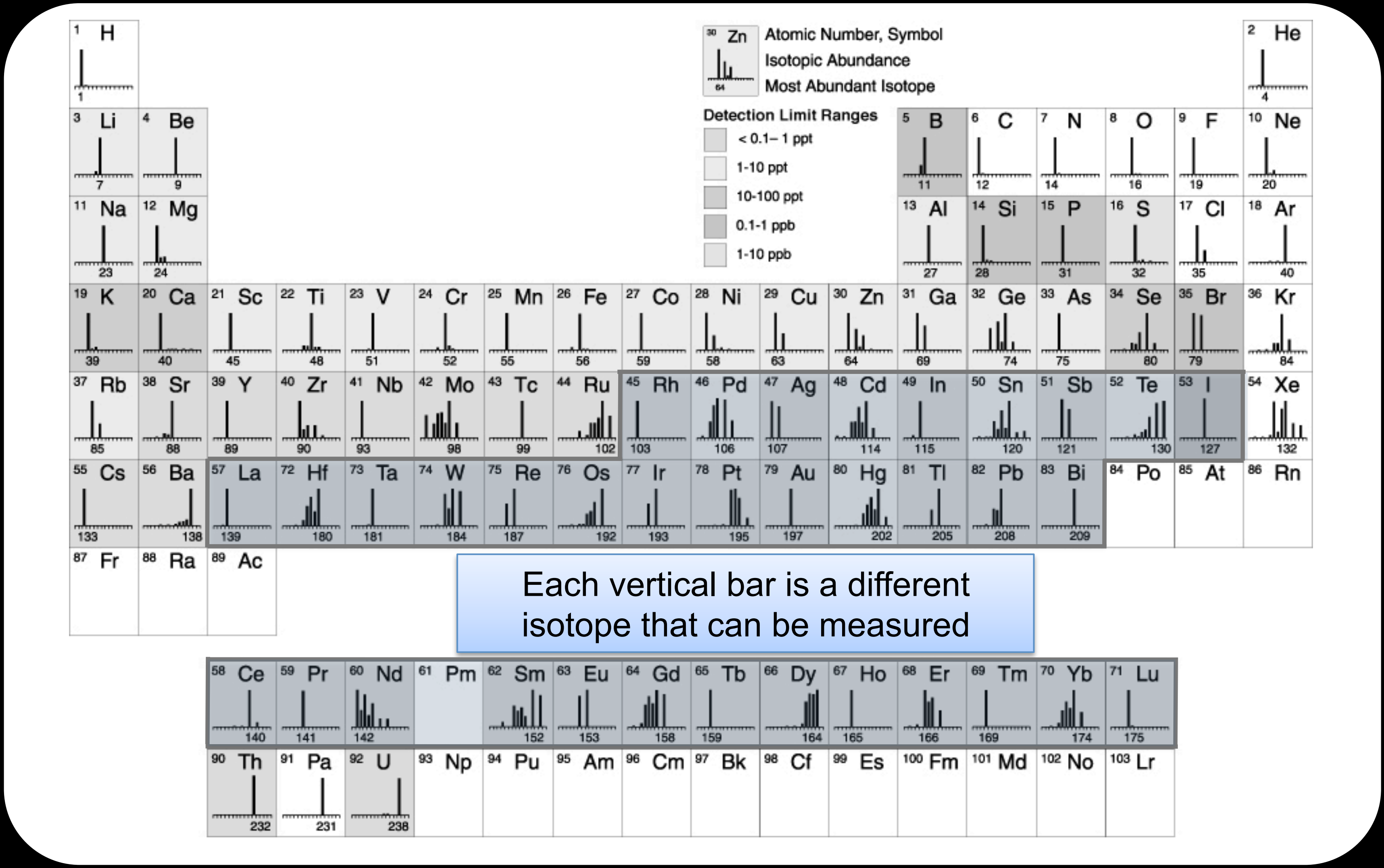
Antibody labels: isotopes of elements



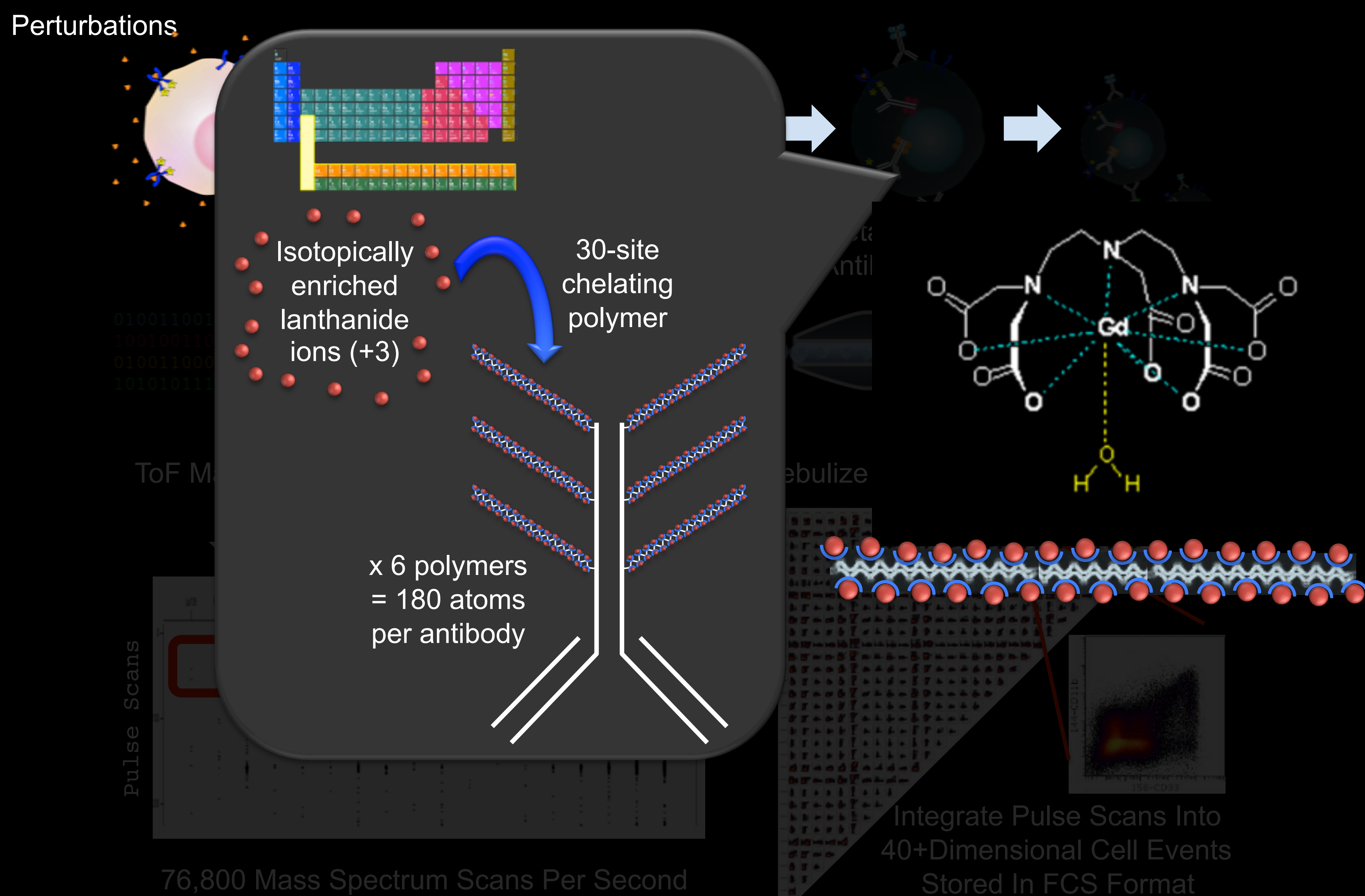
Still dim limited!

Currently working on this problem (talk offline)

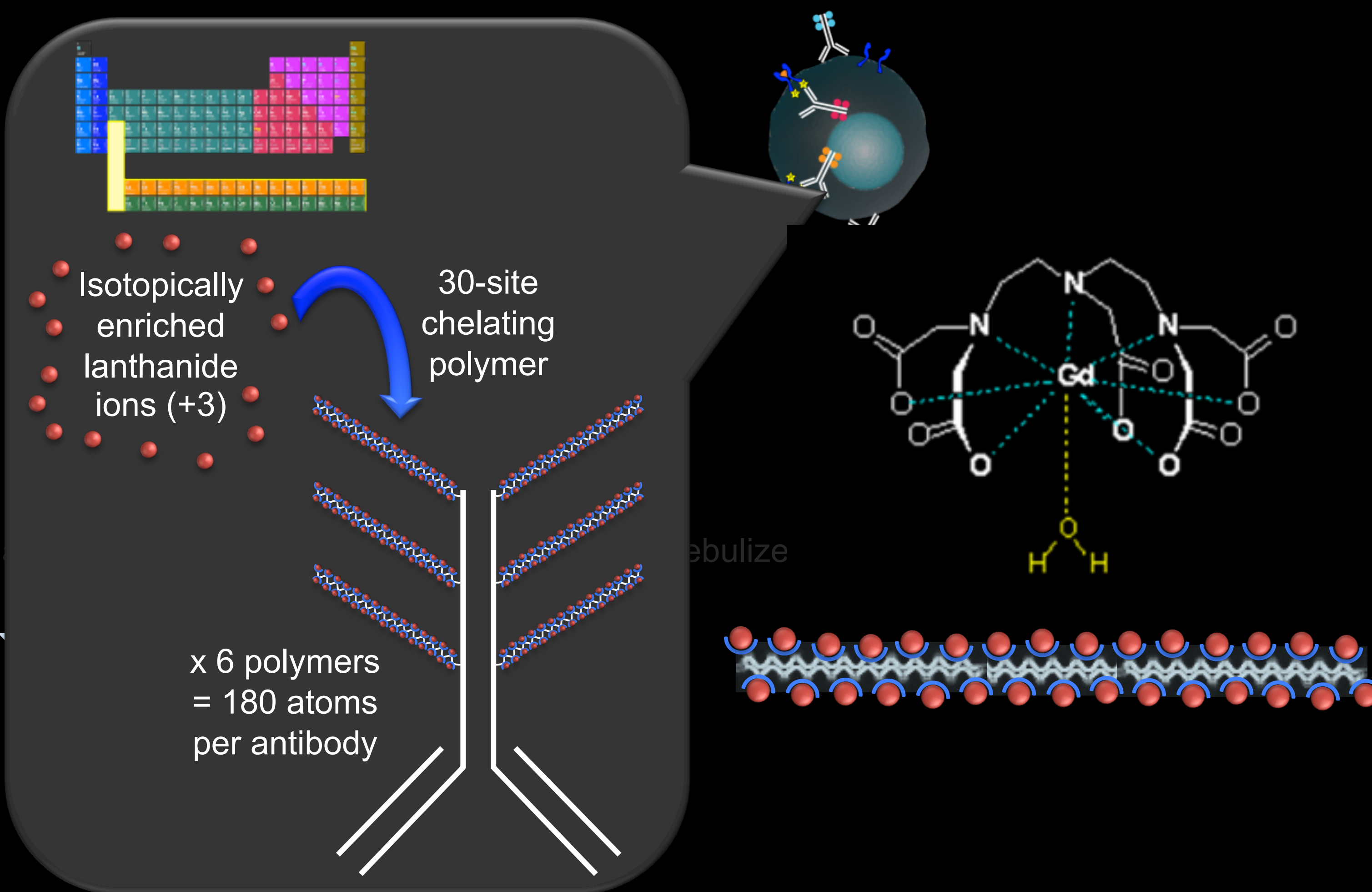
How Do You Get 100 Channels From 35 Elements?



45-dimensional Single Cell Mass Cytometry

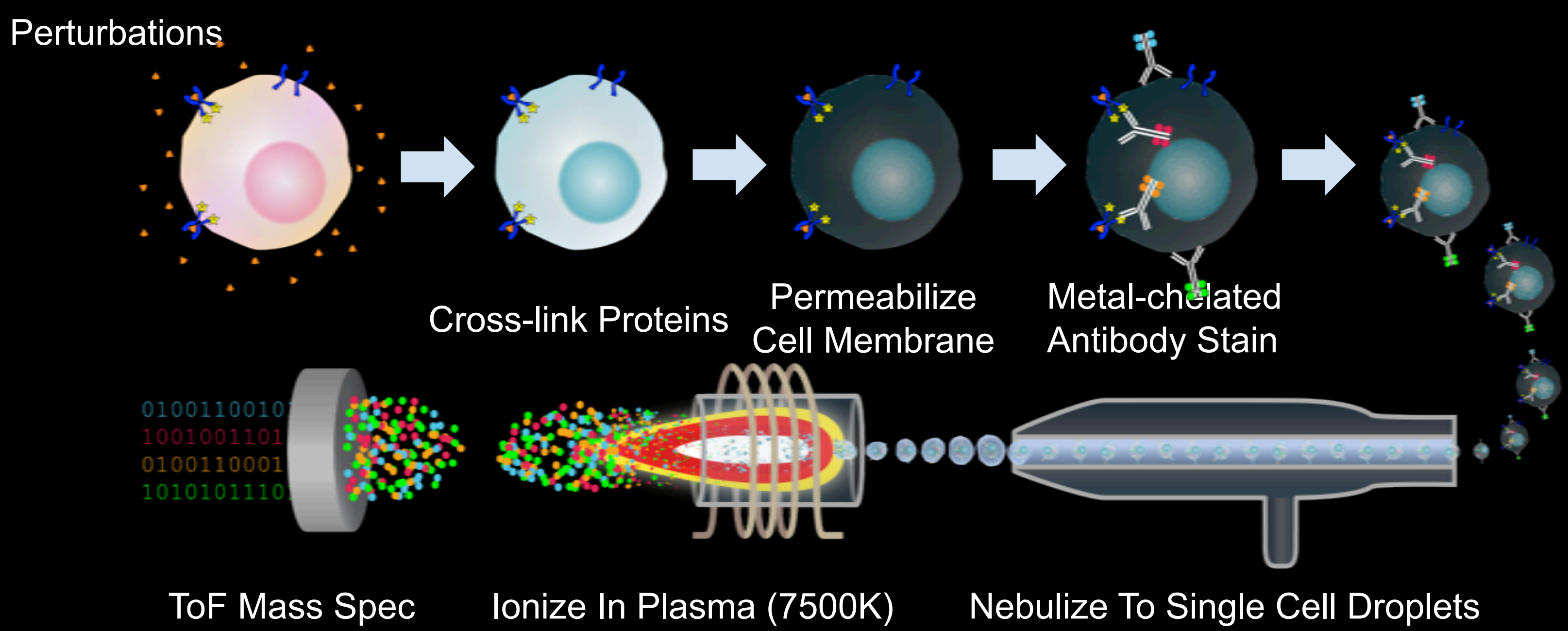


Tags

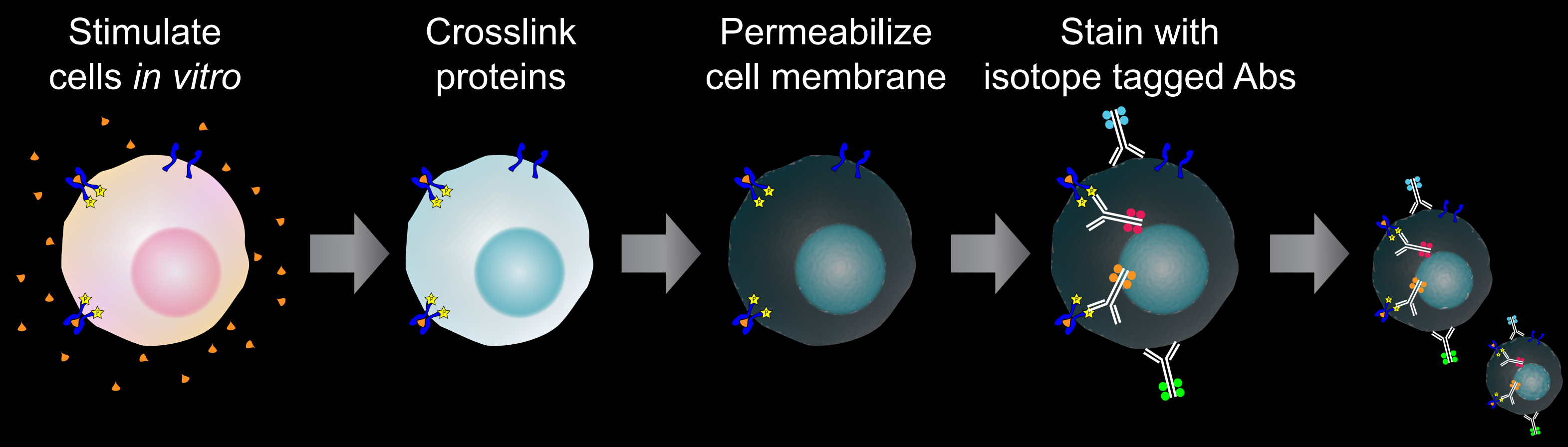


Isotopically enriched lanthanide ions (+3)
30-site chelating polymer
x 6 polymers = 180 atoms per antibody
Gd

Workflow

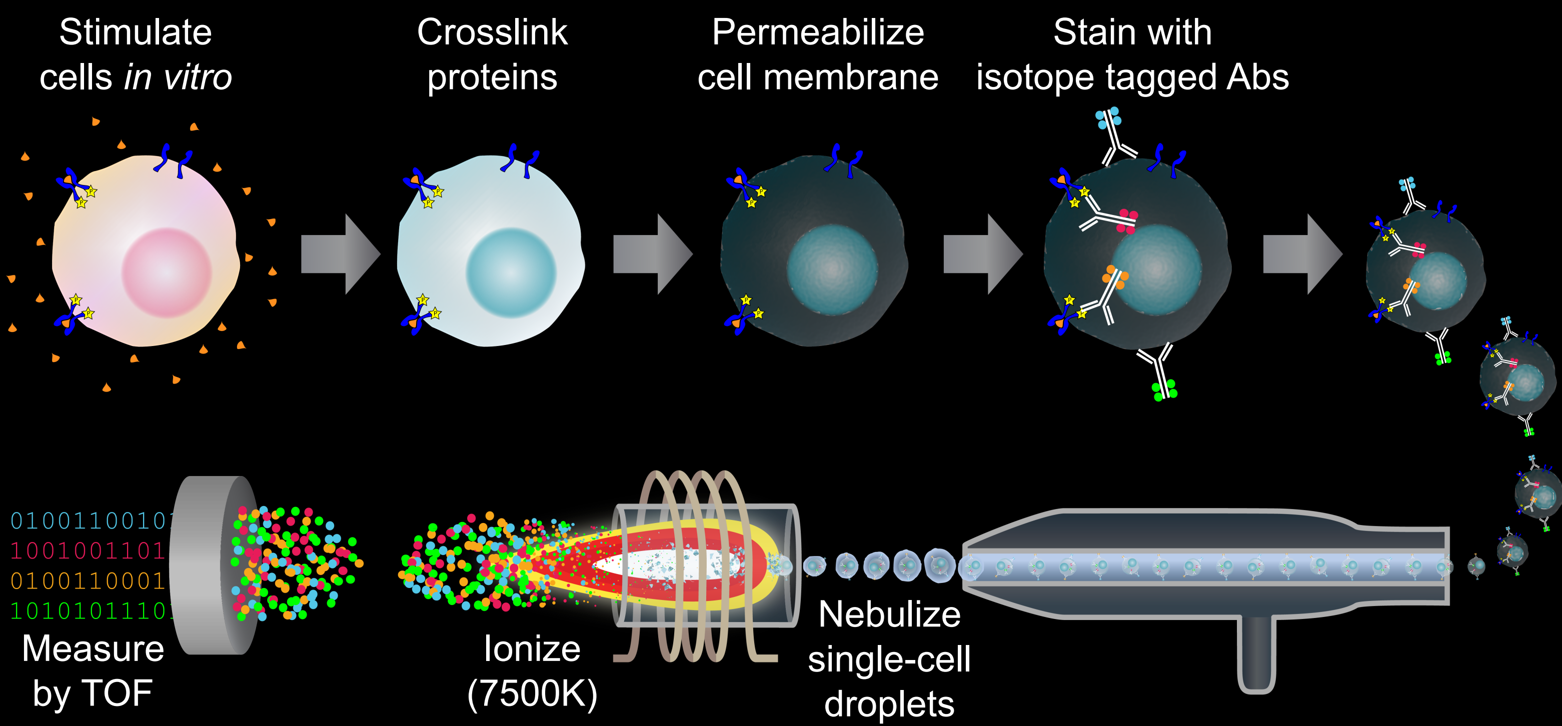

Perturbations
Cross-link Proteins
Permeabilize
Cell Membrane
Metal-chelated
Antibody Stain
ToF Mass Spec
Ionize In Plasma (7500K)
Nebulize To Single Cell Droplets

Workflow: Measuring signaling by mass cytometry

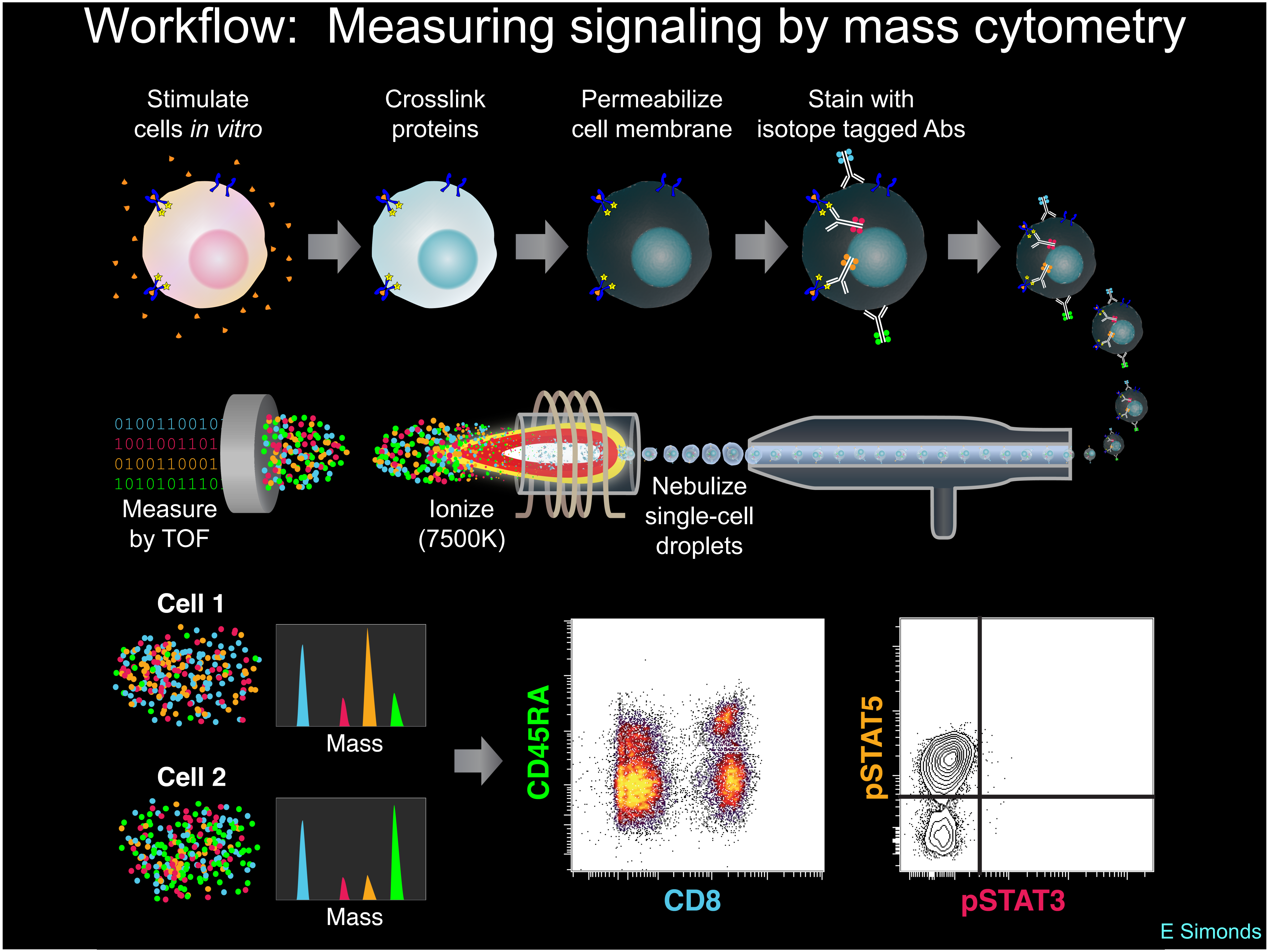


E Simonds

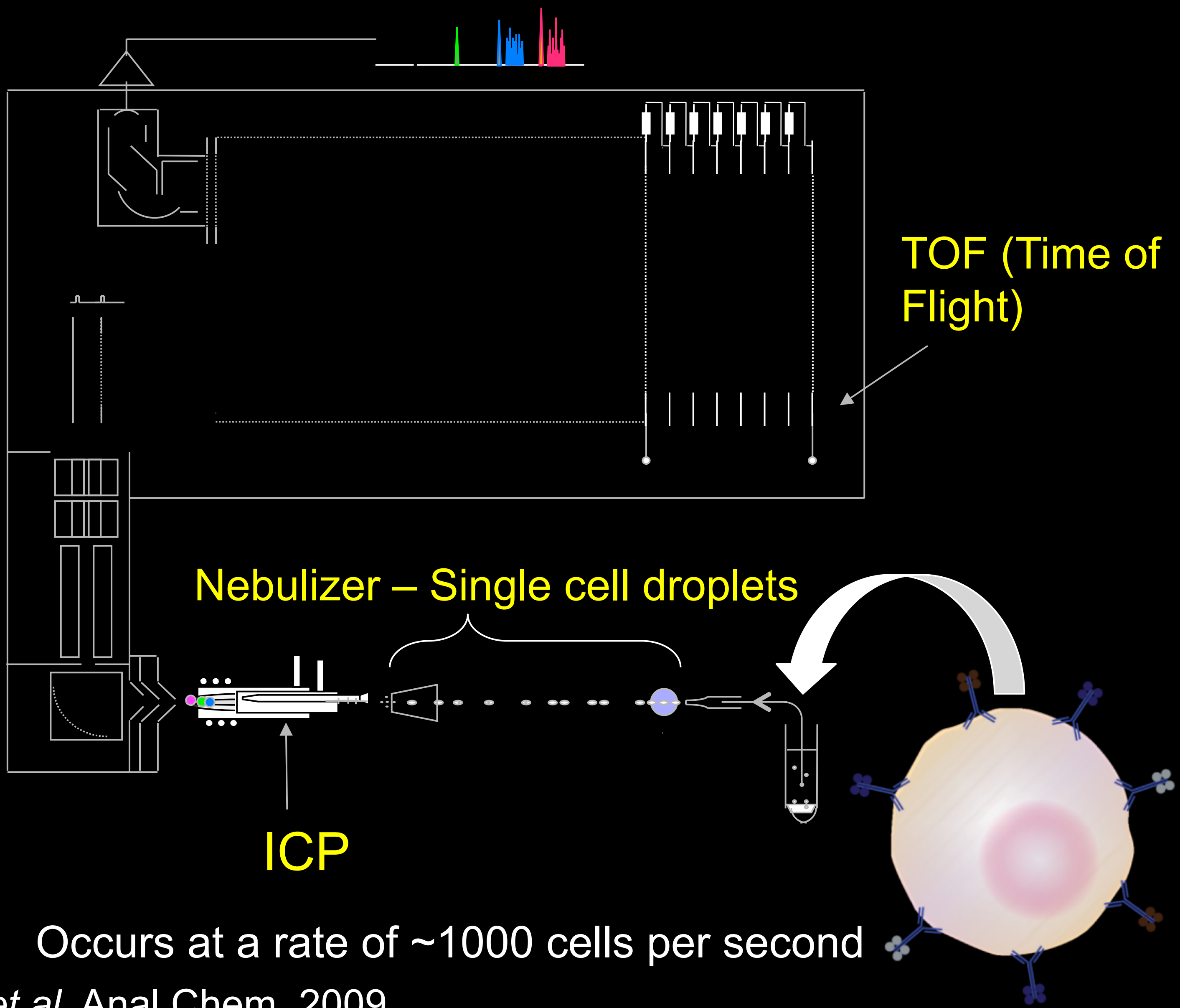
Workflow: Measuring signaling by mass cytometry



E Simonds


Workflow: Measuring signaling by mass cytometry
Stimulate cells *in vitro*
Crosslink proteins
Permeabilize cell membrane
Stain with isotope tagged Abs
0100110010
1001001101
0100110001
1010101110
Measure by TOF
Ionize (7500K)
Nebulize single-cell droplets
Cell 1
Mass
Cell 2
Mass
CD45RA
CD8
pSTAT5
pSTAT3
E Simonds

CyTOF: A prototype schematic

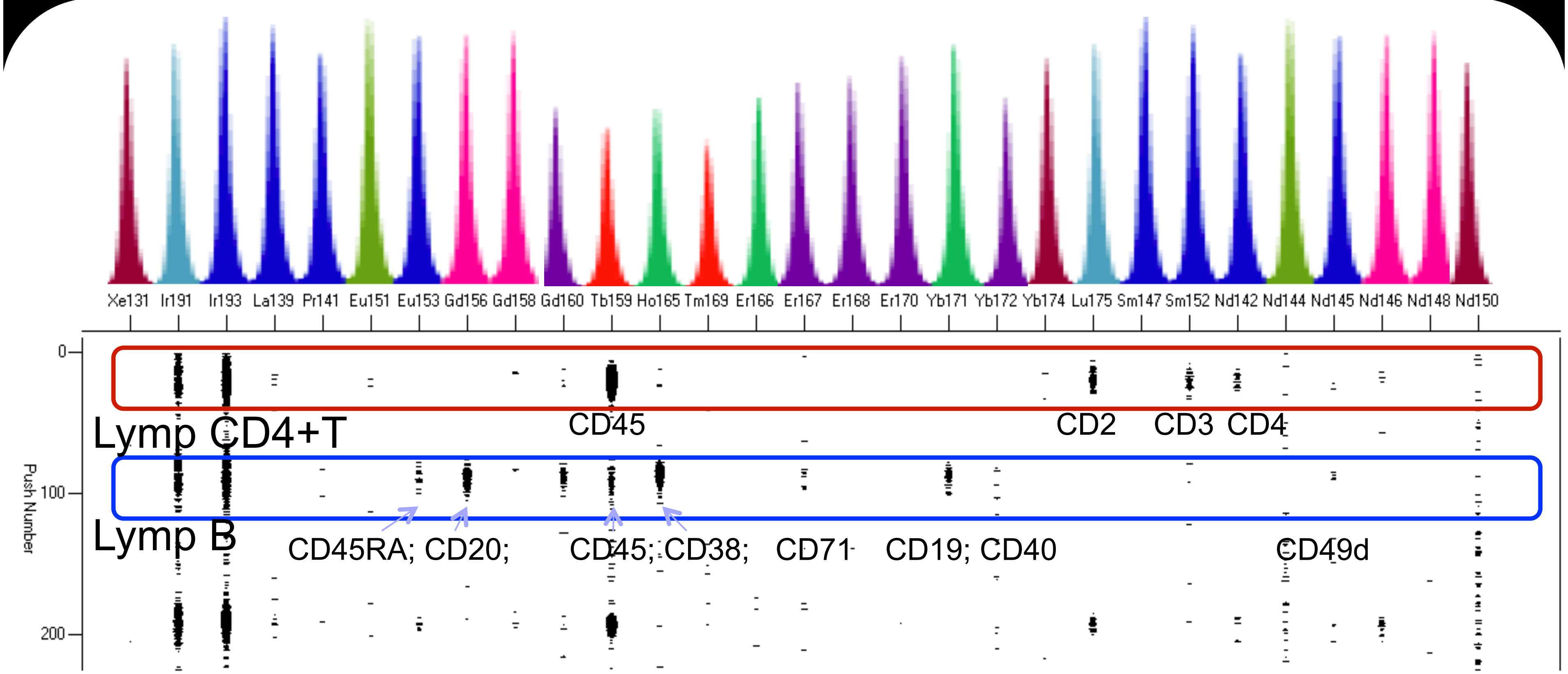


Occurs at a rate of ~1000 cells per second

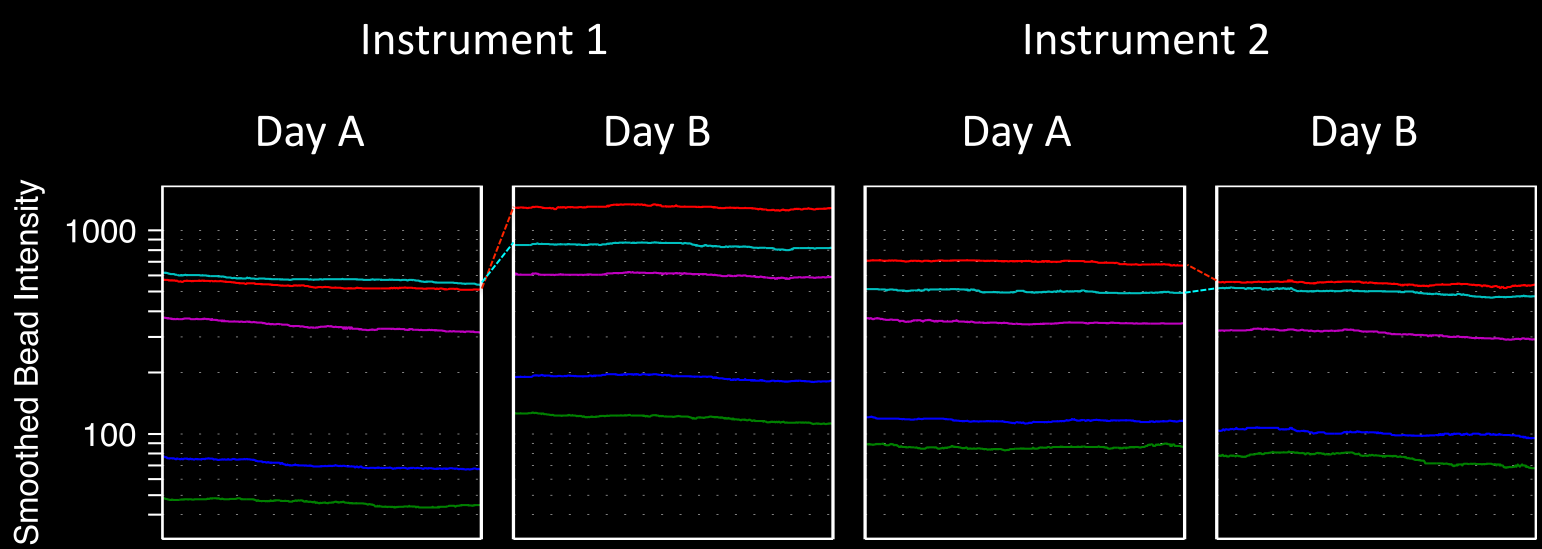
Bandura D, *et al*. Anal Chem. 2009

Fresh PBMc stained with 27 markers (mix I):

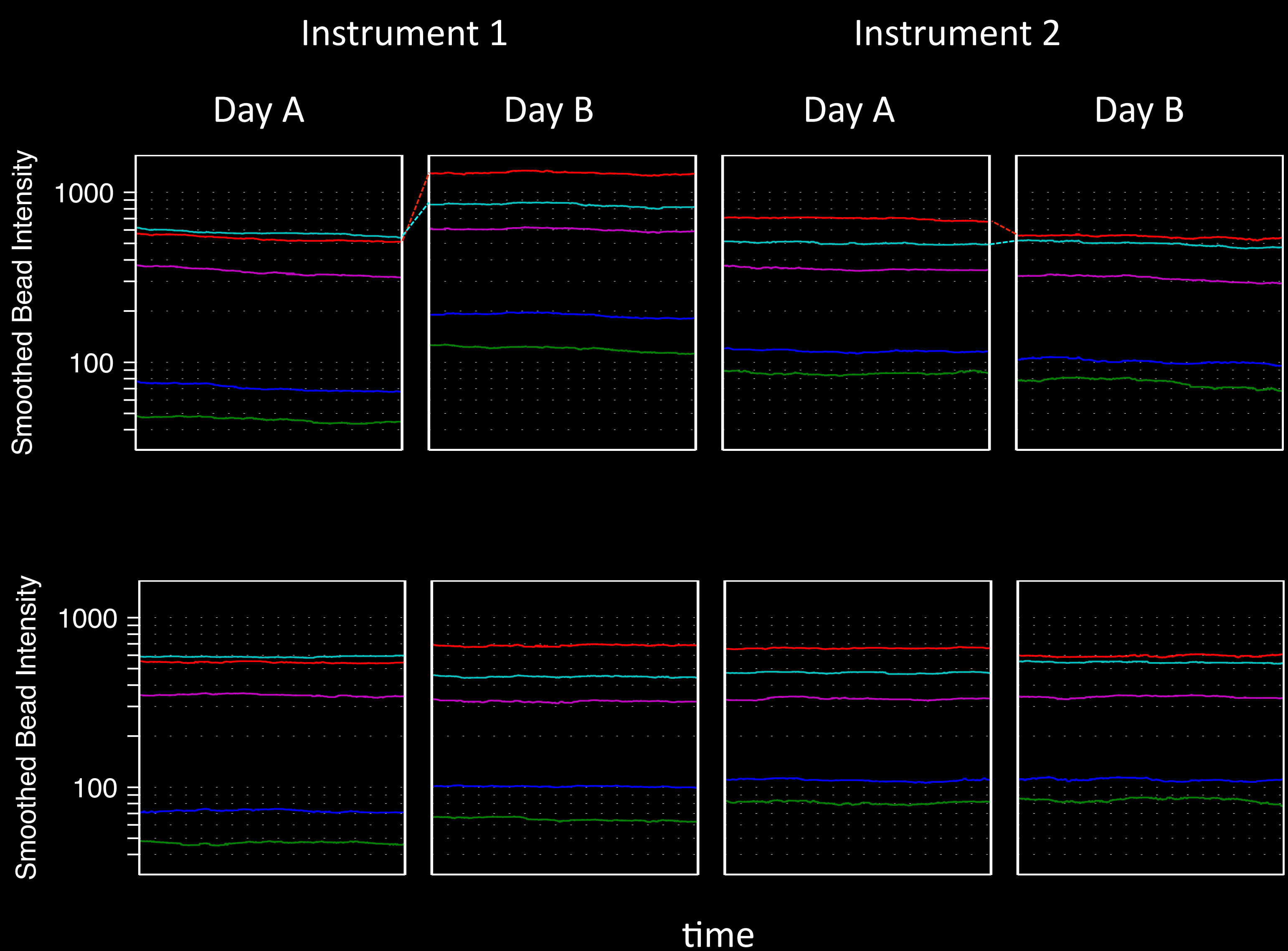
CD2	^{175}Lu	CD13	^{166}Er	CD36	^{150}Nd	CD56	^{176}Yb
CD3	^{152}Sm	CD15	^{170}Er	CD38	^{165}Ho	CD64	^{148}Nd
CD4	^{142}Nd	CD19	^{171}Yb	CD40	^{172}Yb	CD71	^{167}Er
CD7	^{139}La	CD20	^{156}Gd	CD44	^{151}Eu	CD90	^{174}Yb
CD8	^{146}Nd	CD31	^{144}Nd	CD45	^{159}Tb	CD117	^{147}Sm
CD10	^{168}Er	CD33	^{141}Pr	CD45RA	^{153}Eu	HLA-DR	^{160}Gd
CD11b	^{158}Gd	CD34	^{169}Tm	CD49d	^{145}Nd		



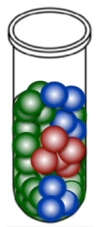
Variation Across Calibrations and Instruments



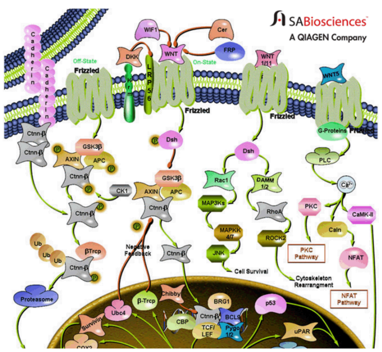
Bead normalization tames variation



Causal learning in signaling

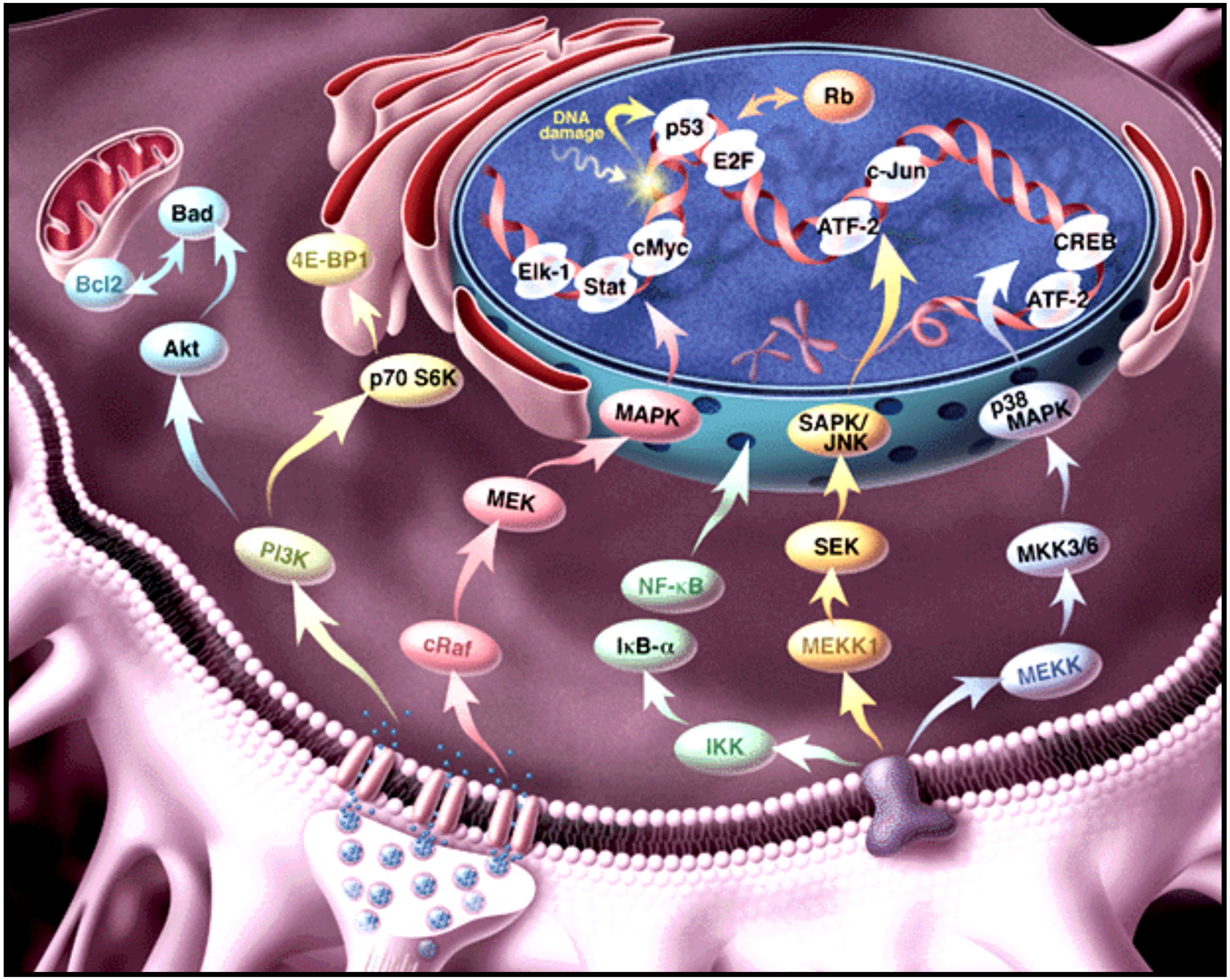


1. Where does data come from?

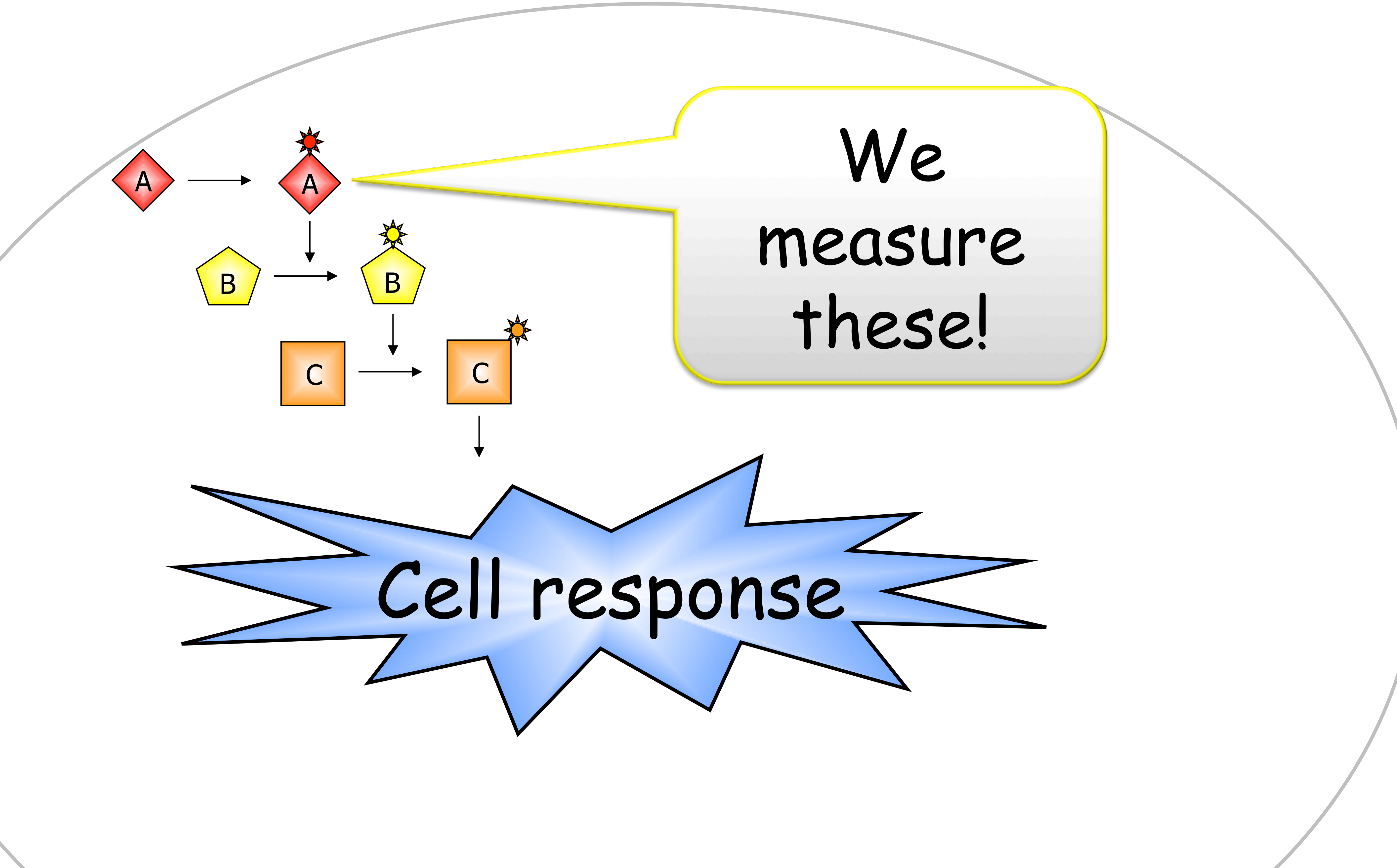


2. What causal connections appear?
   - What happens?
   - What can we see?

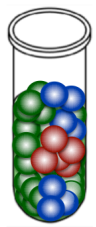
Signaling 101



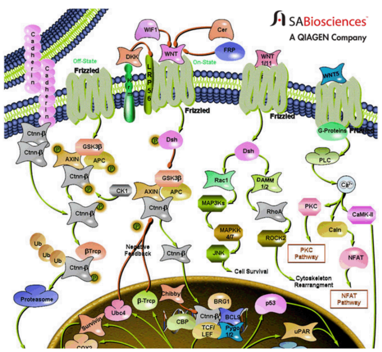
Signaling 101: Measure activated species



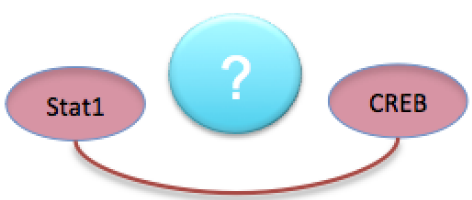
Causal learning in signaling



Where does data come from?



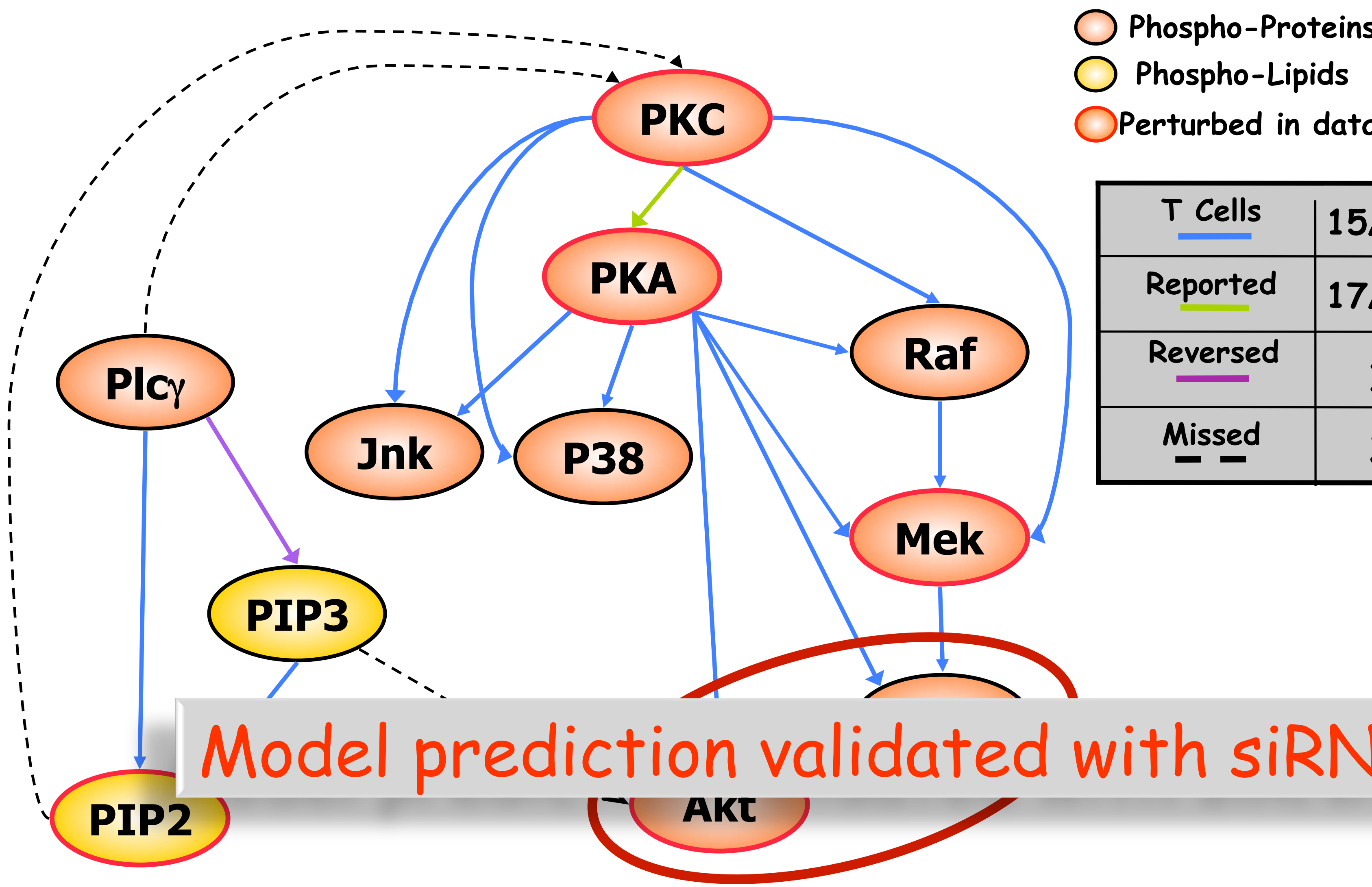
What causal connections appear?



What is needed for causal learning?

- Outstanding challenges

Accurate Network Inference



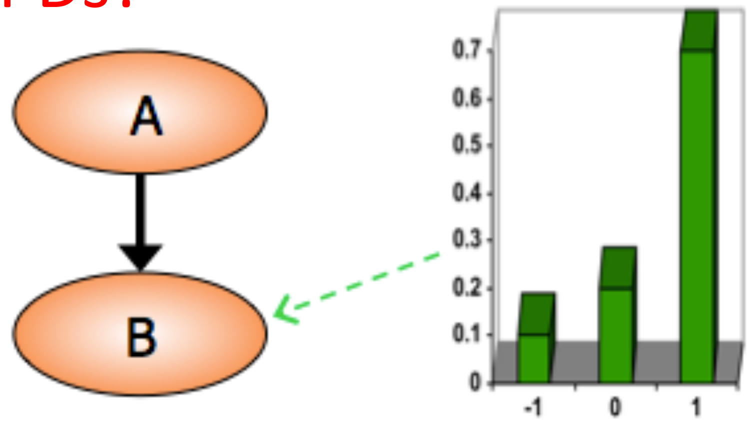
T Cells	15/17
Reported	17/17
Reversed	1
Missed	3

Model prediction validated with siRNA

[Sachs *et al*, *Science* 2005]

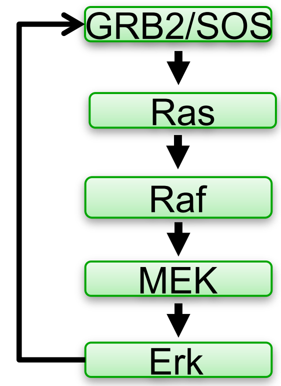
Wait! What about..

- CPDs?

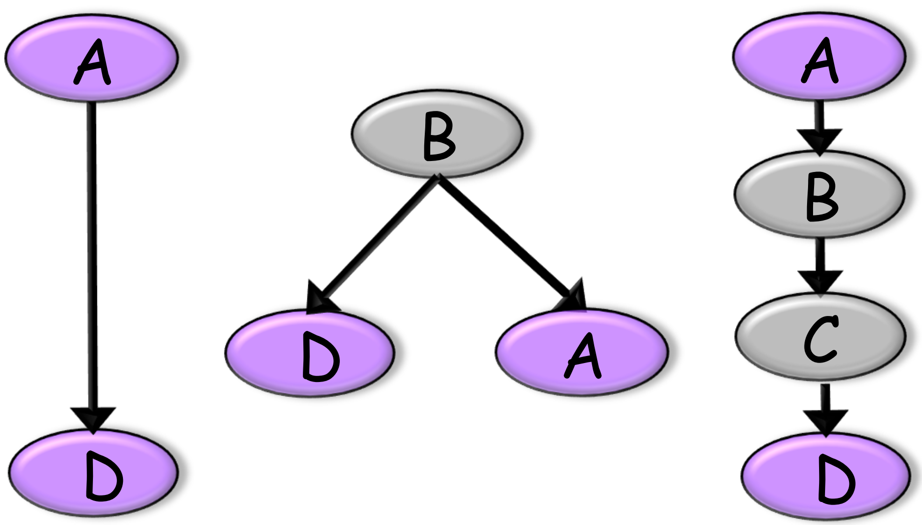


- Cycles?

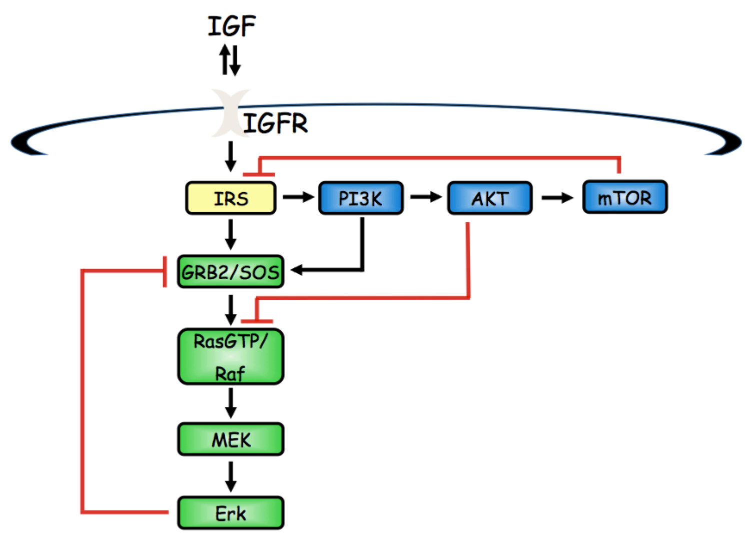
Sachs and Itani, IEEE 2008, PSB 2009, JMLR Proc 2008



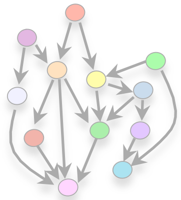
- Hidden variables?



- Dynamics?

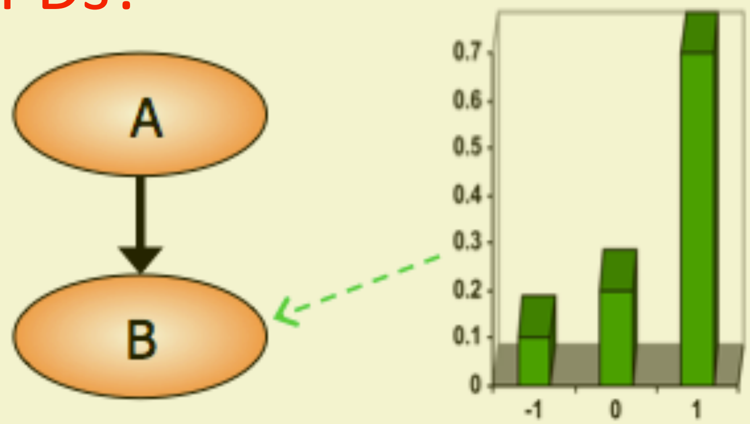


Sachs, Interface Focus 2013



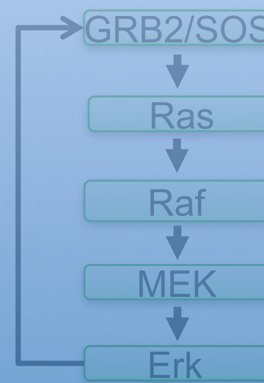
Remaining challenges

- CPDs?

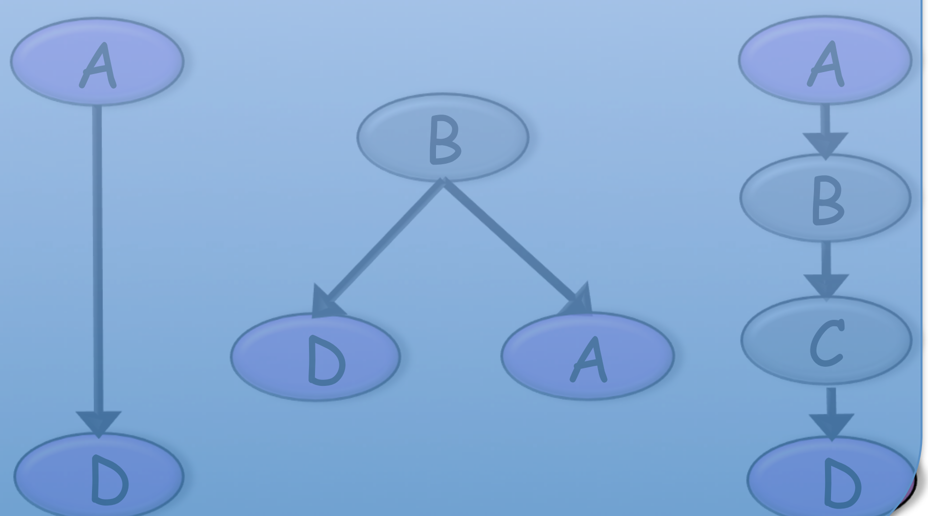


- Cycles?

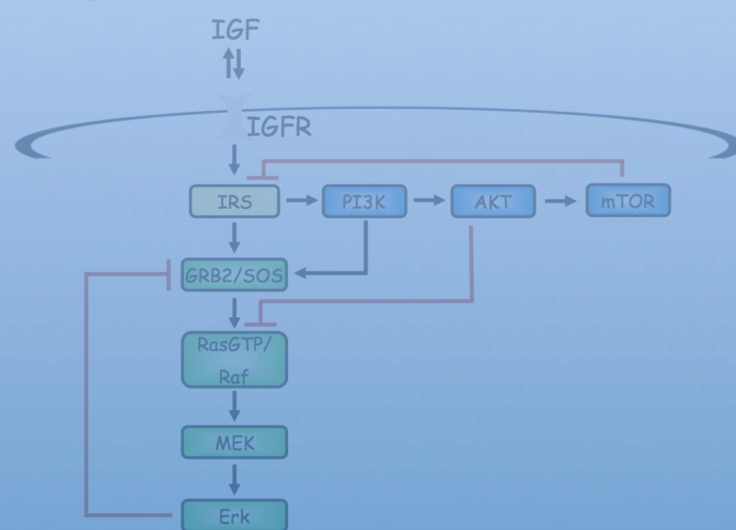
Sachs and Itani, IEEE 2008, PSB 2009, JMLR Proc 2008



- Hidden variables?

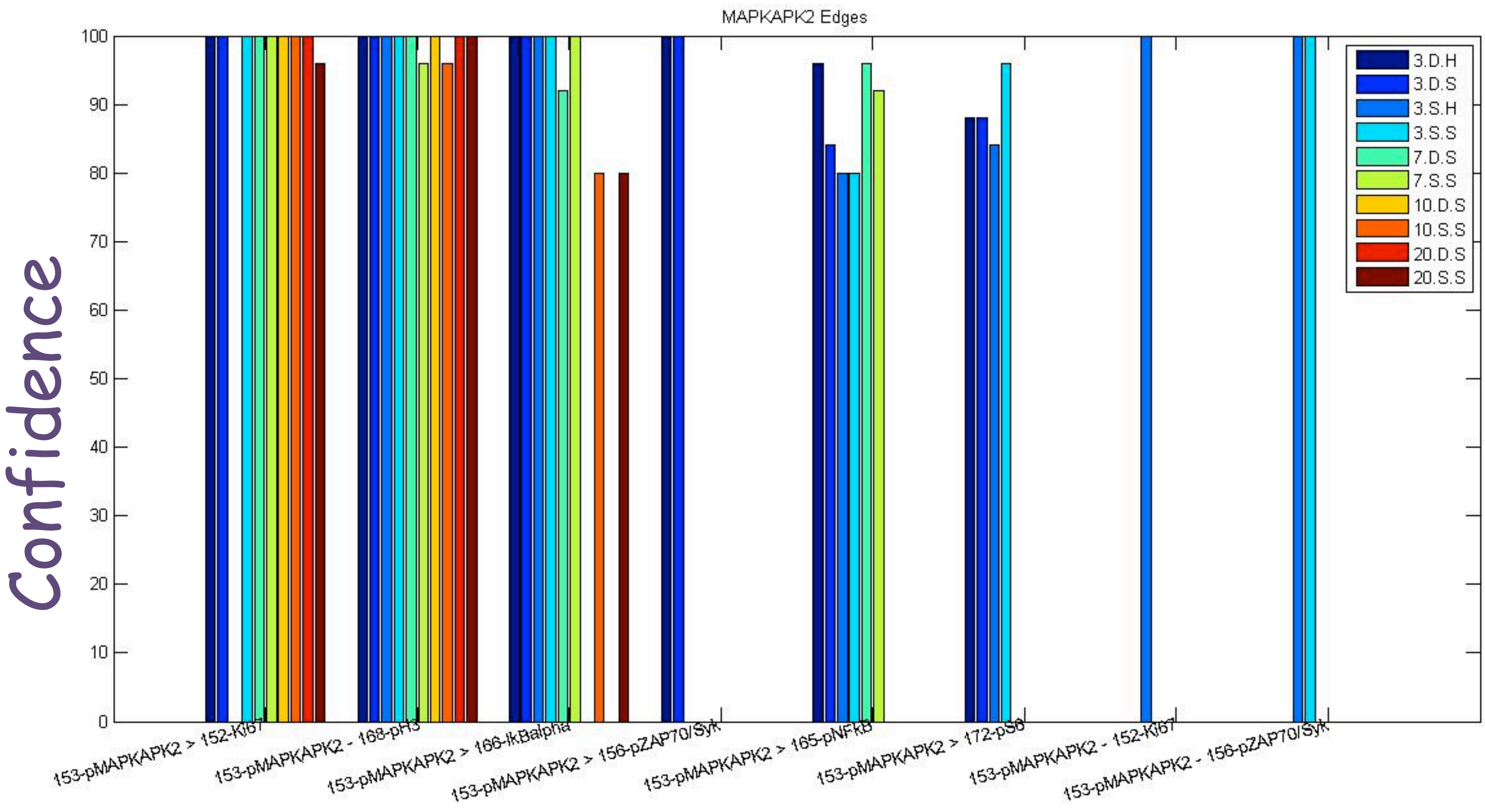


- Dynamics?

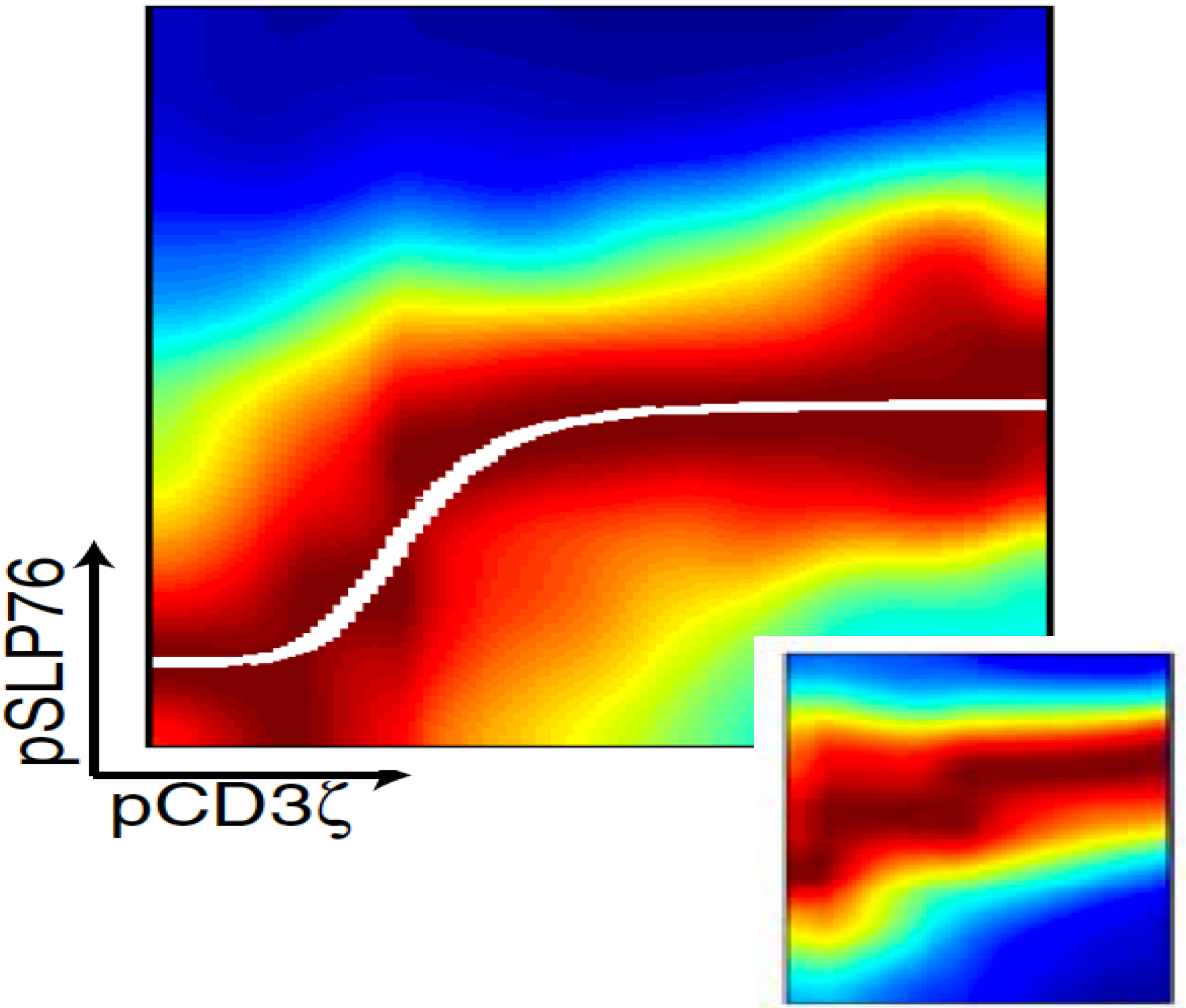


Sachs, Interface Focus 2013

Edges can be CPD dependent



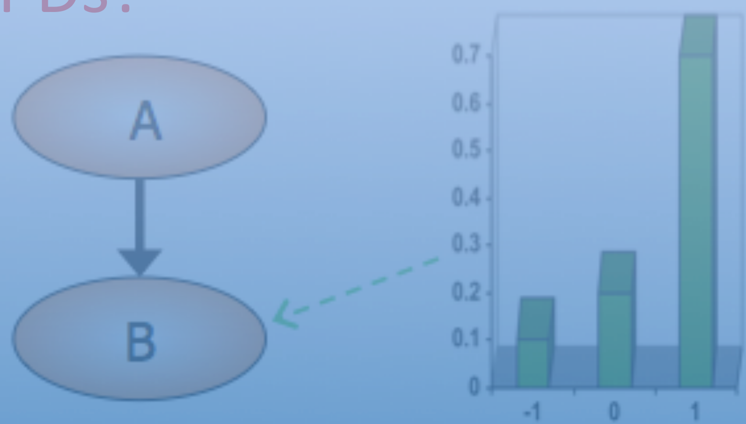
CPDs need expressive power



- Multinomial?
- Linear?
- Other?
- GP (J. Mooij)

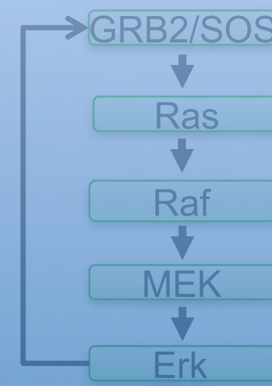
Remaining challenges

- CPDs?



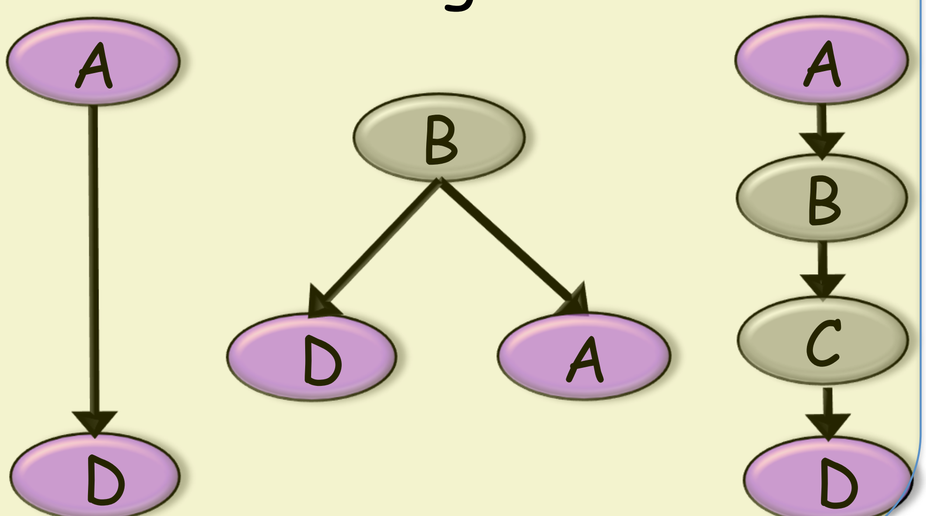
- Cycles?

Sachs and Itani, IEEE 2008, PSB 2009, JMLR Proc 2008

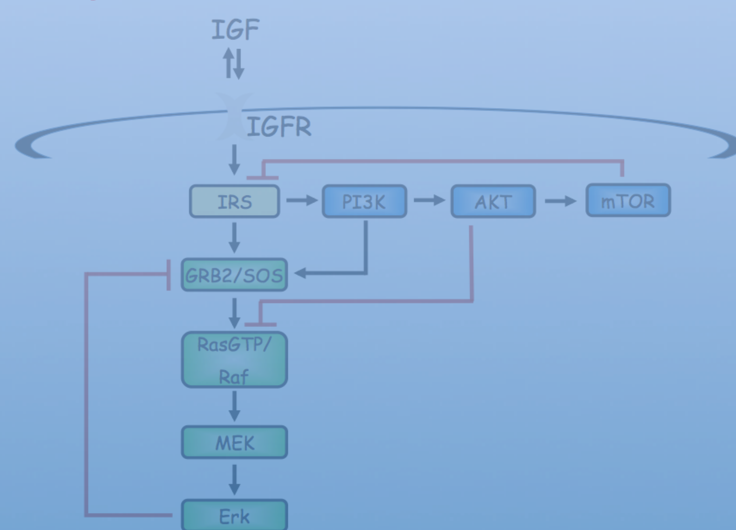


- Hidden variables?

Data integration

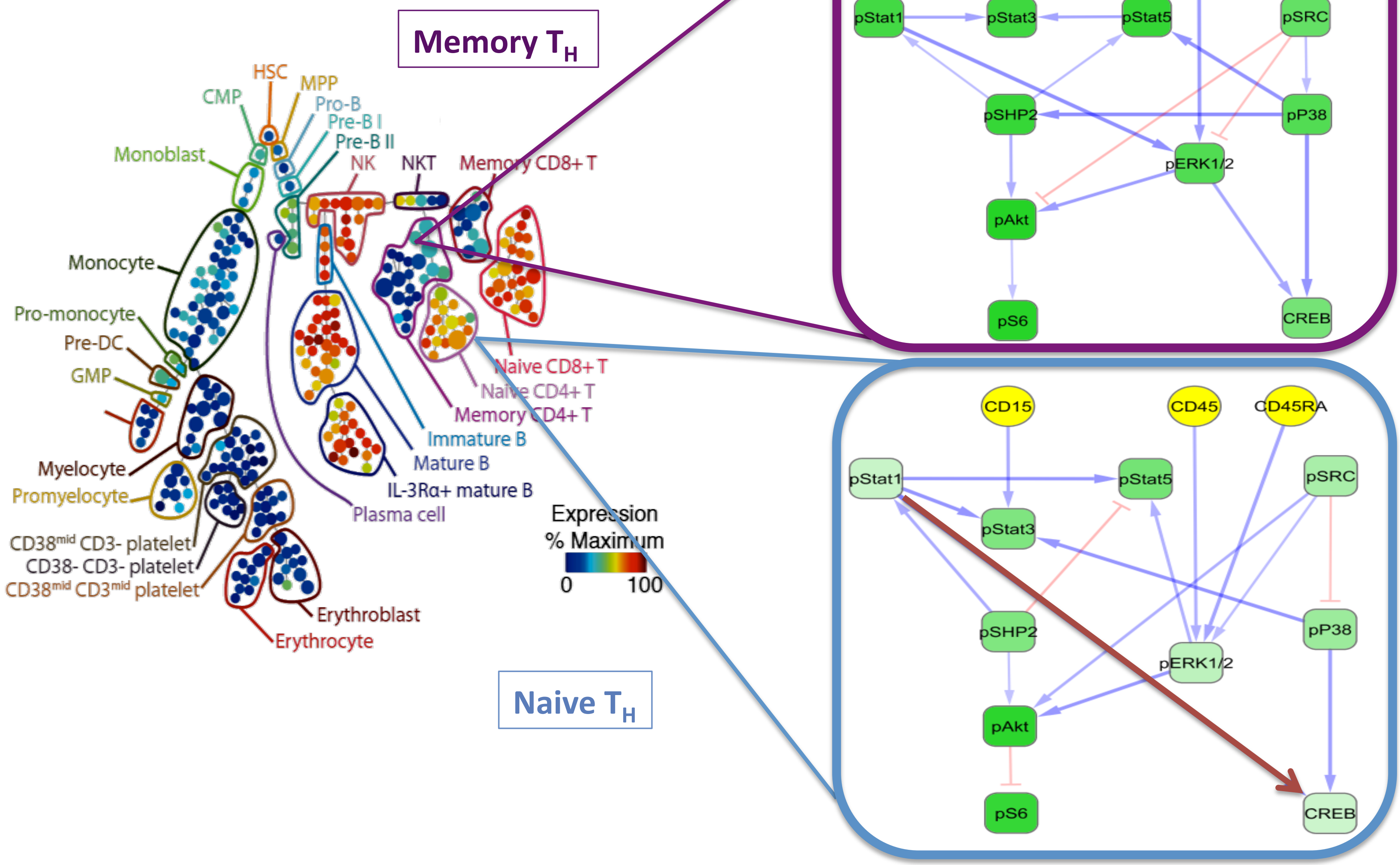


- Dynamics?

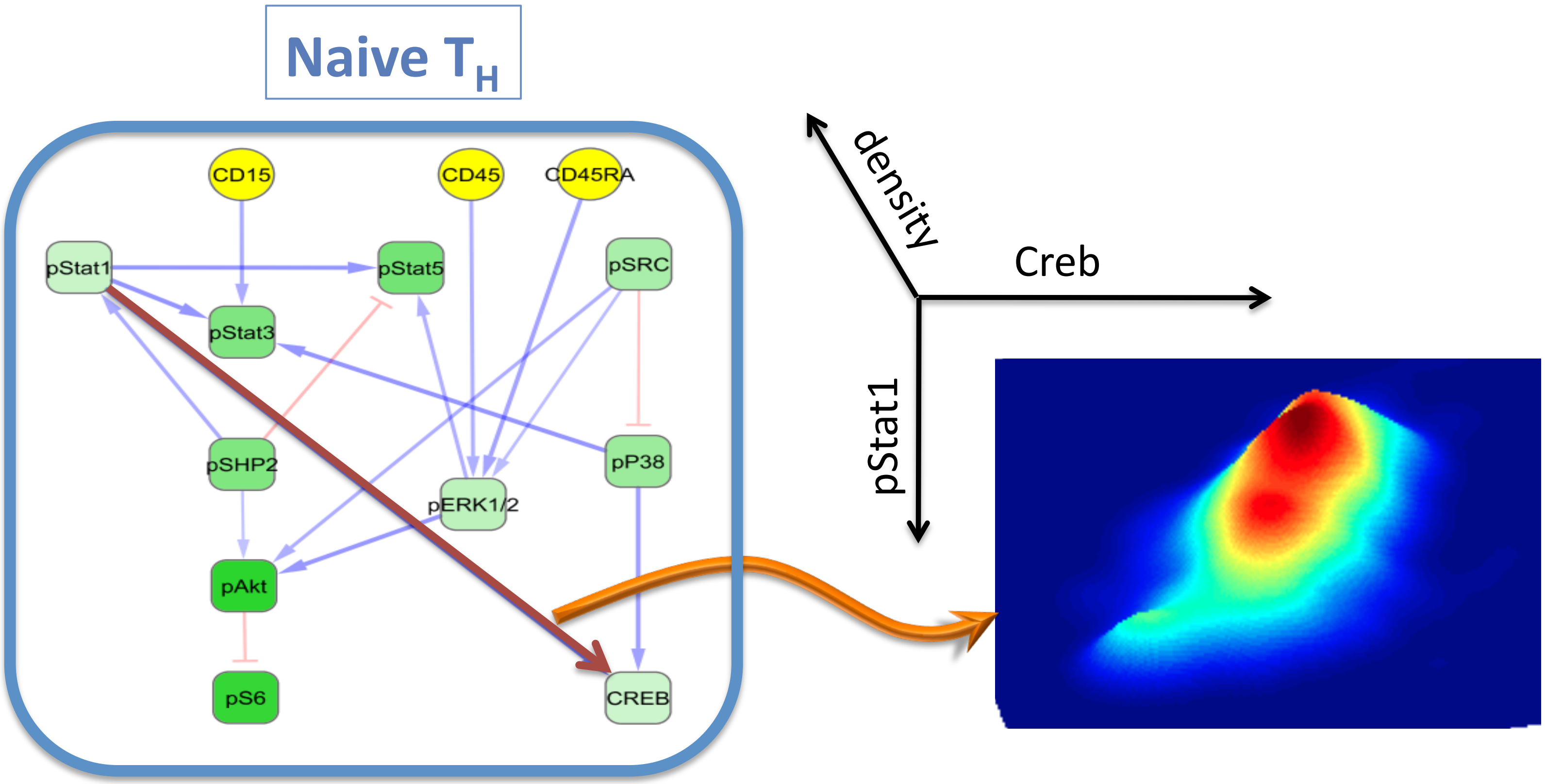


Sachs, Interface Focus 2013

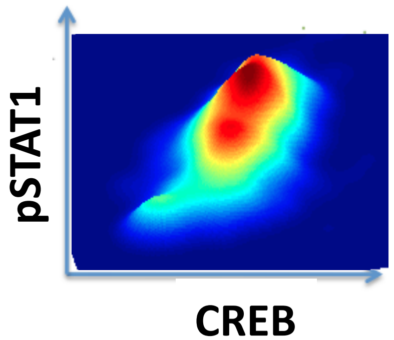
Comparative signaling



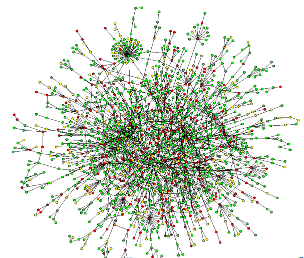
Elucidated edge is supported in the data



Workflow for adding hidden variables



Statistical inference



Literature & curated DBs

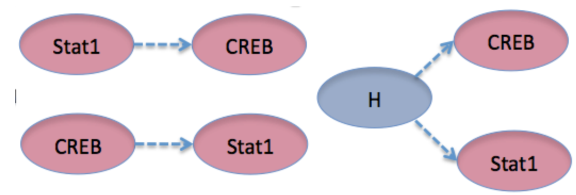
Scansite

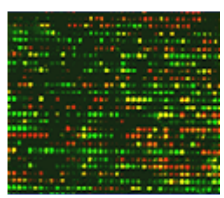
In silico predictions

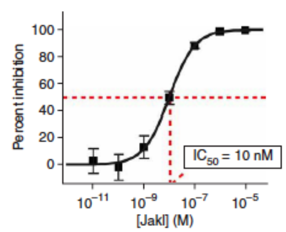
Extract potential underlying molecular paths

IGF1R(3480)(3480) -PHOS(Y1229)
(NETWORKIN)-> MUC1() -+PHOS(SNR)
(SRI)-> IKBA(4792) , IGF1R(3480)(3480) -
PHOS(Y719/Y934)(NETWORKIN)->
KIT(3815) -
+BINDING(KEGG_MAMMALIAN)->
STAT5A(6776)
IGF1R(3480)(3480) -PHOS(Y1229)
(NETWORKIN)-> MUC1() -+PHOS(SNR)
(SRI)-> IKBA(4792) , IGF1R(3480)(3480) -
PHOS(Y1008)(NETWORKIN)-> JAK2(3717)
-+YPHOS(SRI)-> STAT5A(6776)
CDK1()() -PHOS(S1227)(NETWORKIN)->
MUC1() -+PHOS(SNR)(SRI)-> IKBA(4792) ,
CDK1()() -#VALUE!
PHOS(SIGNED_KINOME)-> ABL1(25) -+

Putative ***directed paths*** and ***common ancestors***



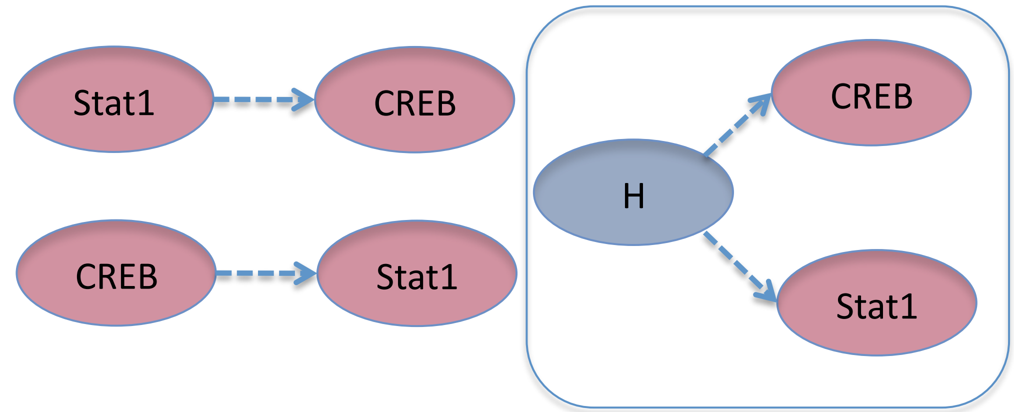


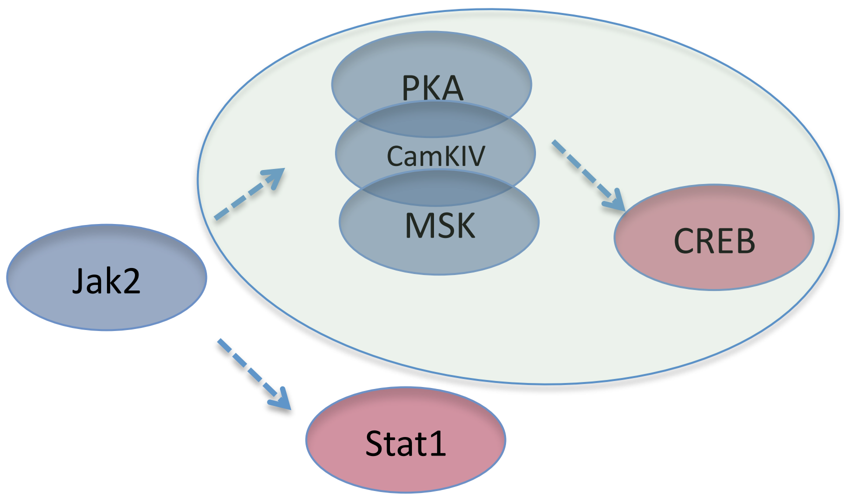


✗ A→B→C
✗ A→D→F→C
✓ A←G→C

Rank and select most probable path(s)

Uncovering underlying path

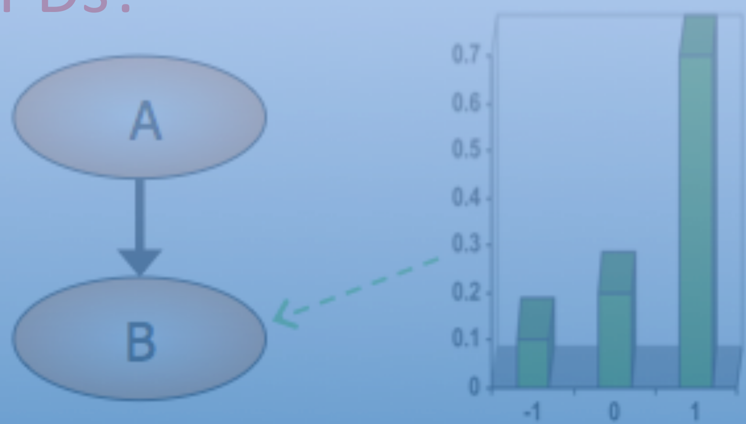




Branch active in Naïve
Not active in Memory

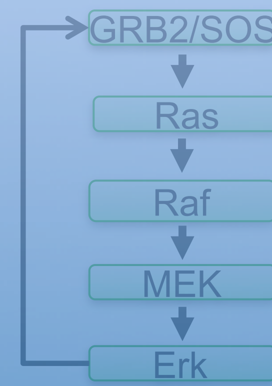
Remaining challenges

- CPDs?



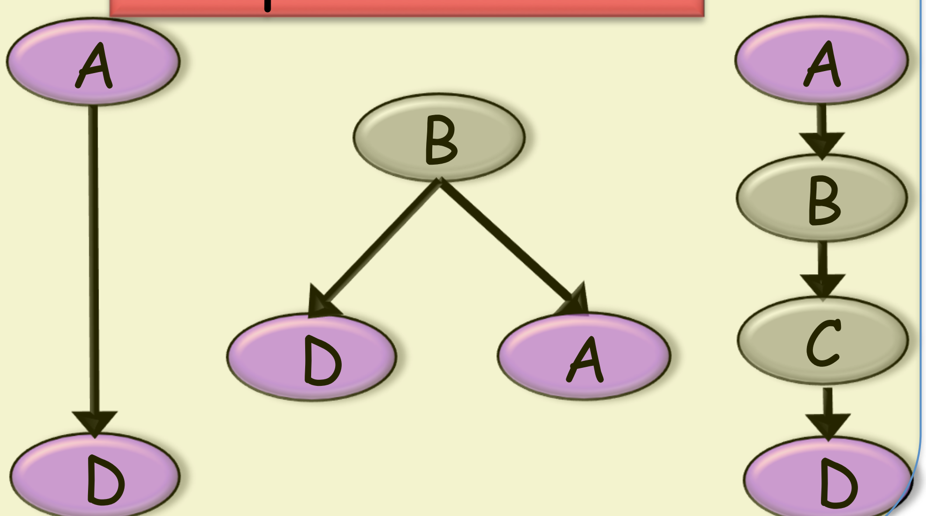
- Cycles?

Sachs and Itani, IEEE 2008, PSB 2009, JMLR Proc 2008

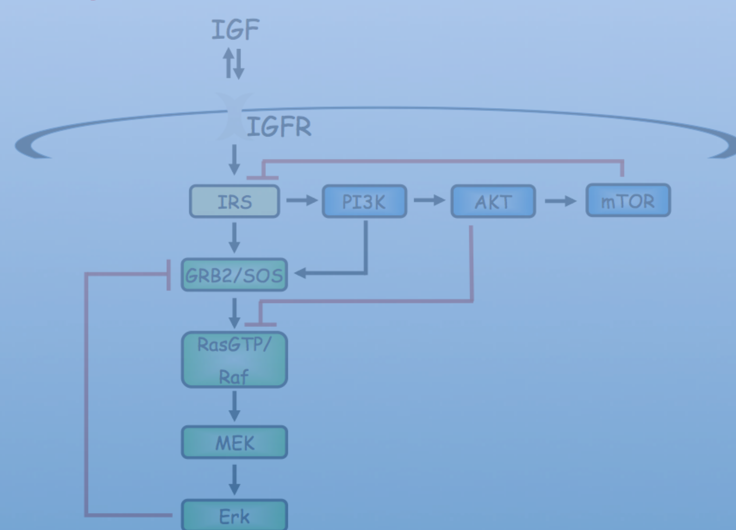


- Hidden variables?

Special case



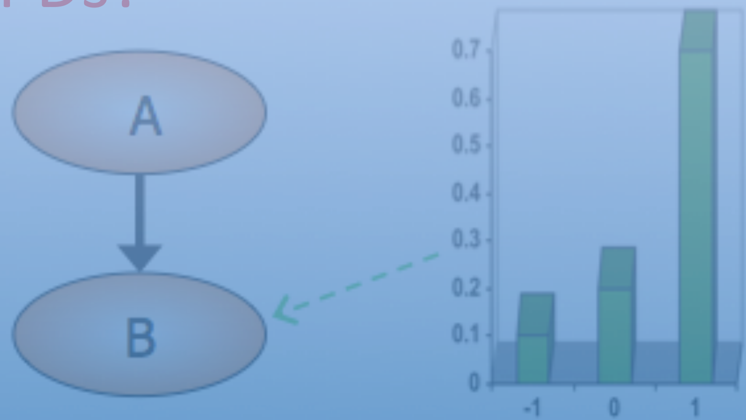
- Dynamics?



Sachs, Interface Focus 2013

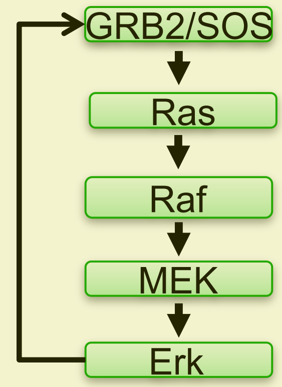
Remaining challenges

- CPDs?



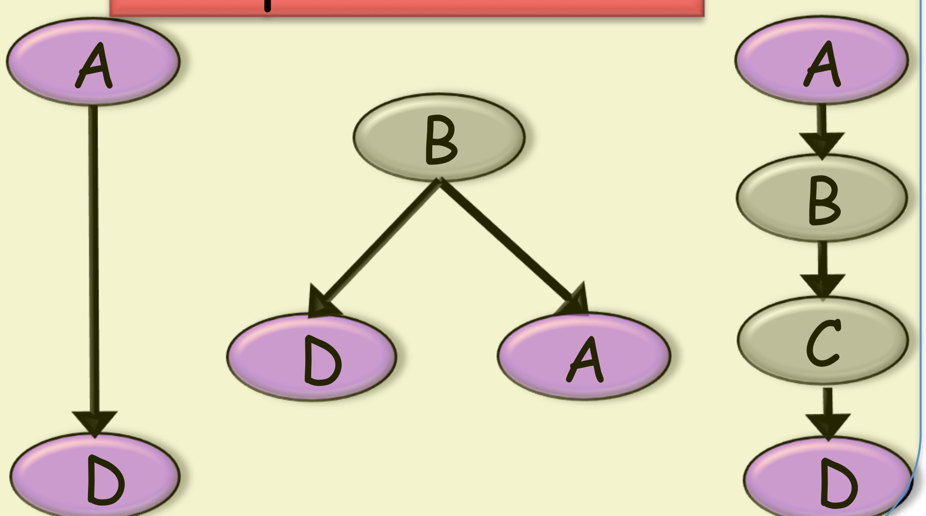
- Cycles?

Sachs and Itani,
IEEE 2008,
PSB 2009,
JMLR Proc 2008

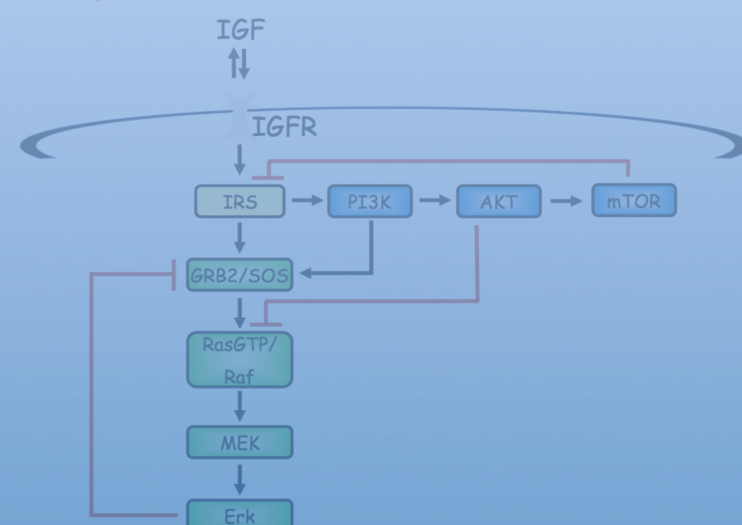


- Hidden variables?

Special case

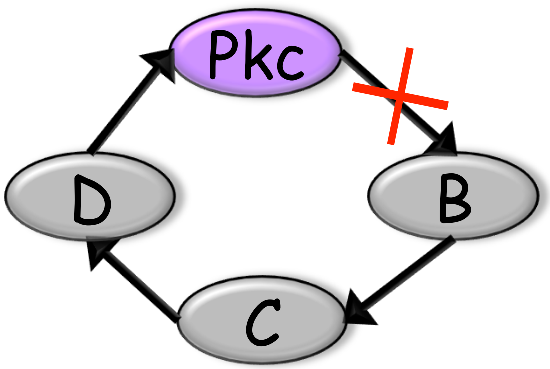


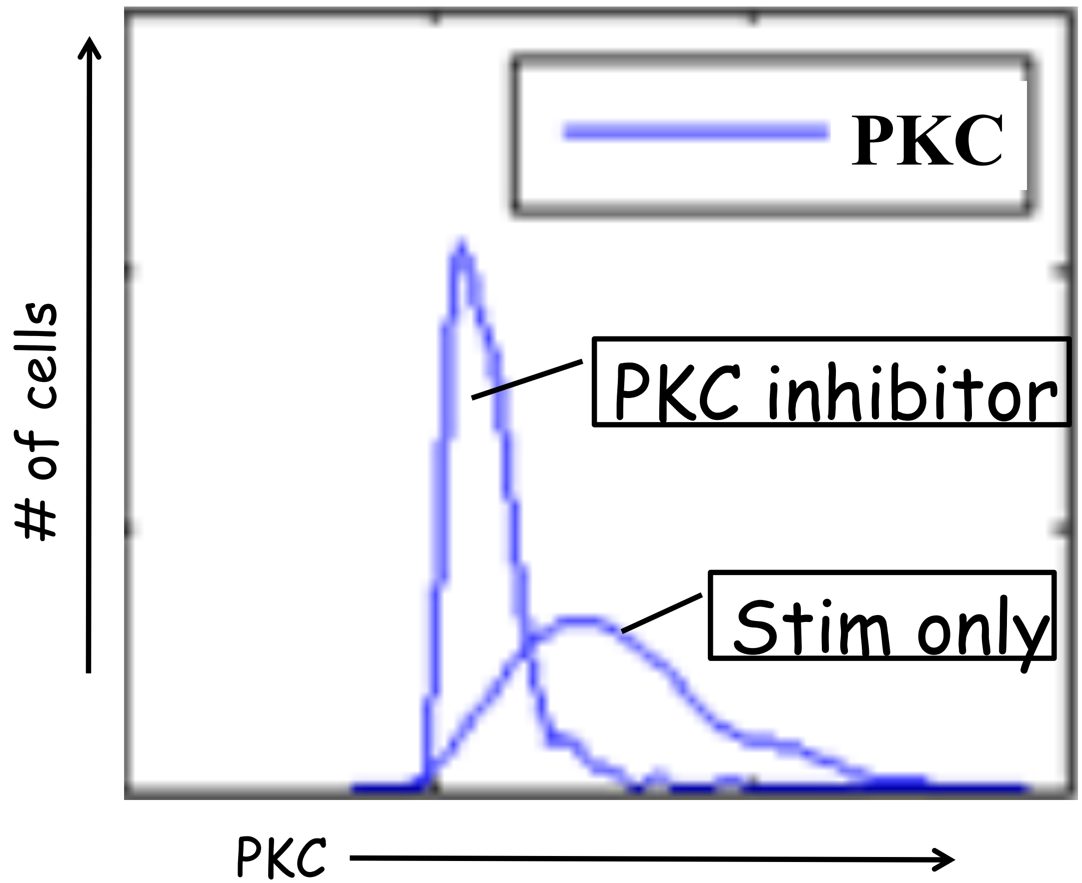
- Dynamics?



Sachs, Interface Focus 2013

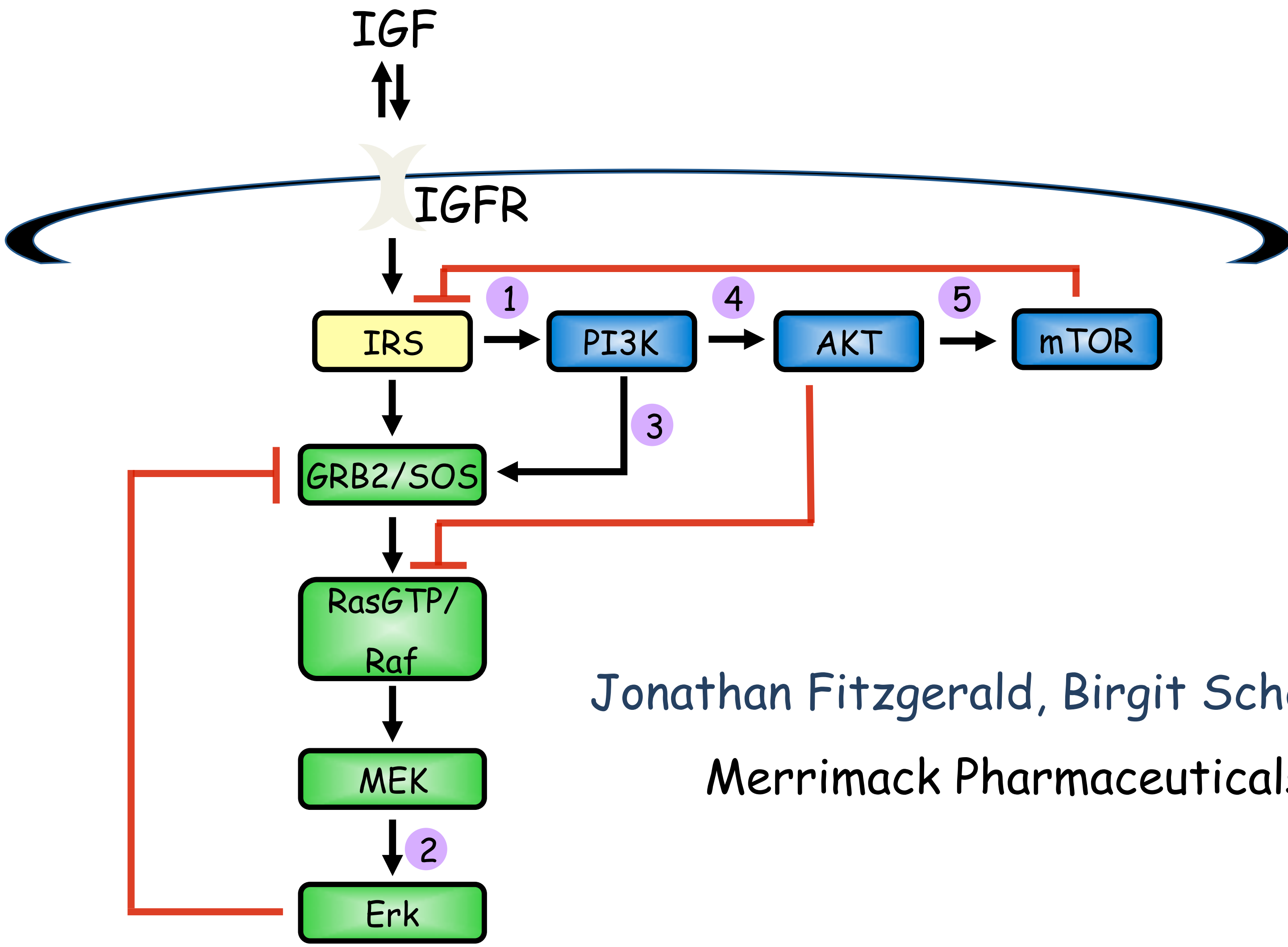
Cyclic structure learning





Itani and Sachs et al, JMLR Proc 2008

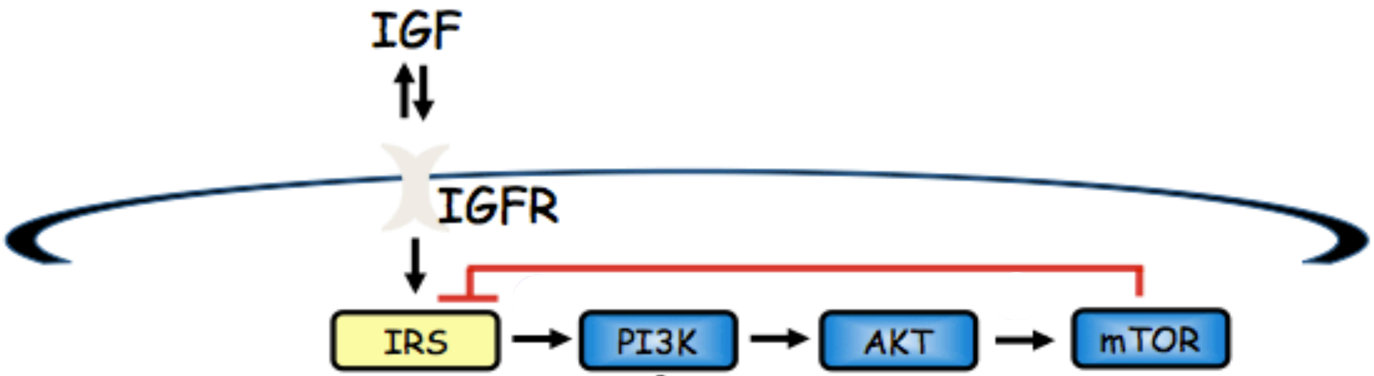
ODE model for realistic synthetic data



Jonathan Fitzgerald, Birgit Schoeberl

Merrimack Pharmaceuticals

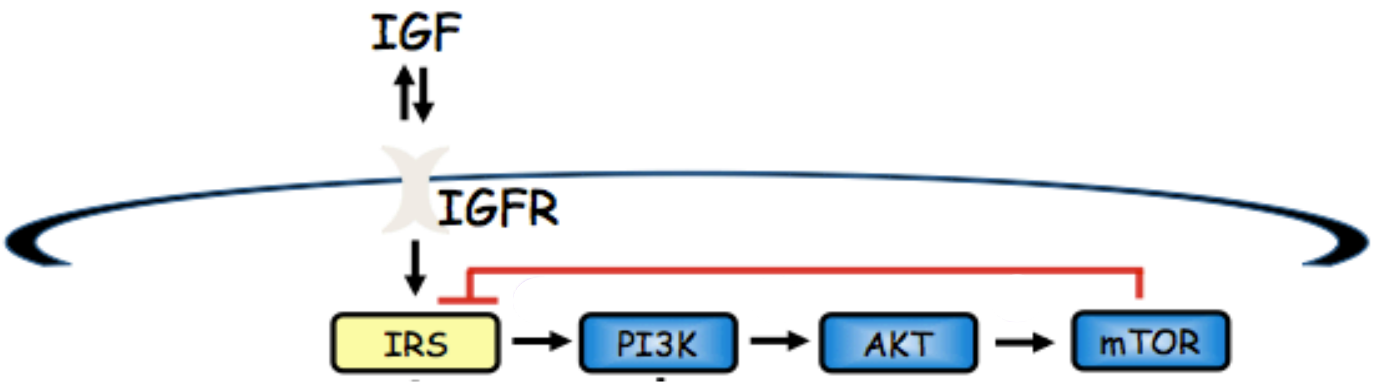
Extra edge?



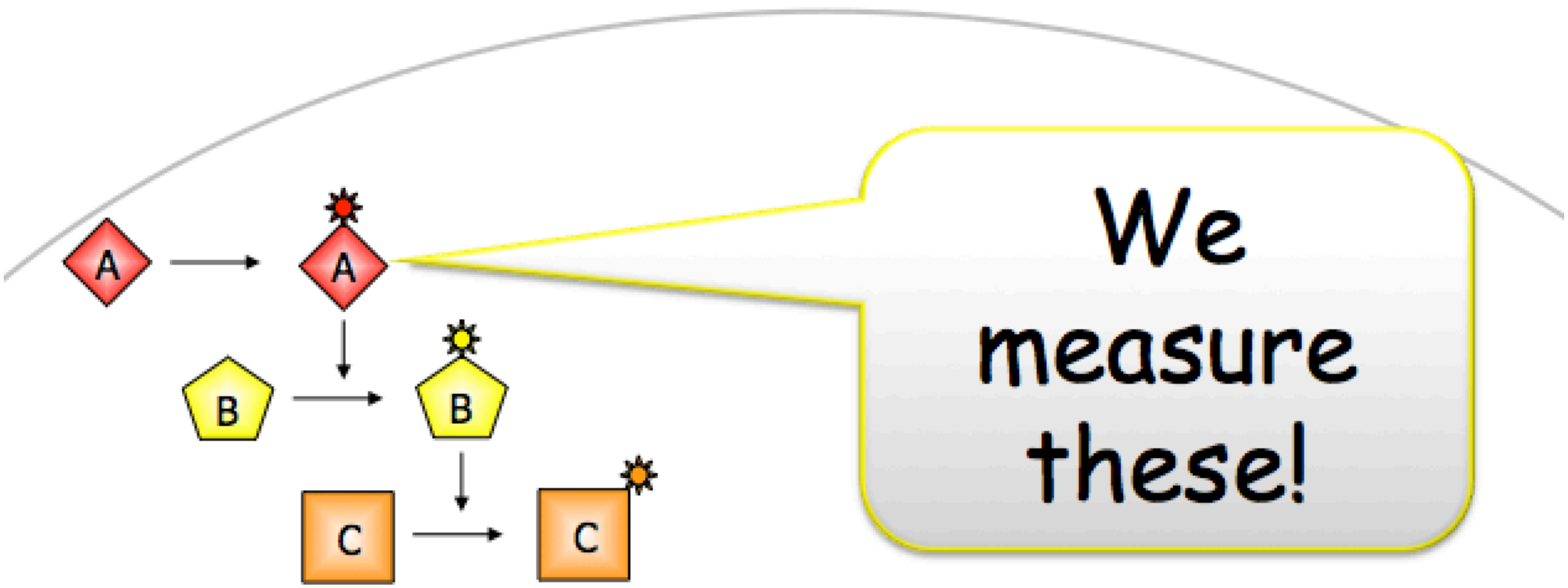
Ground truth

Model results

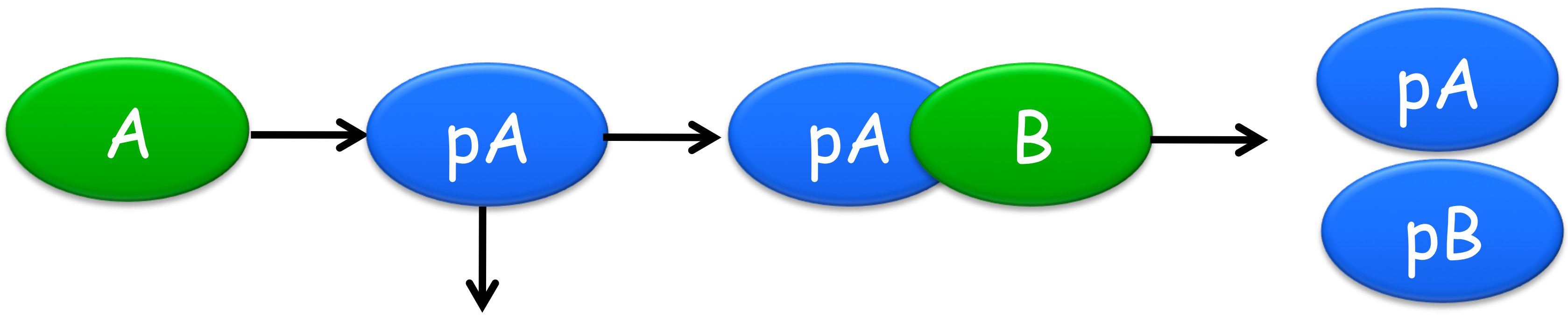
Zoom in on signaling

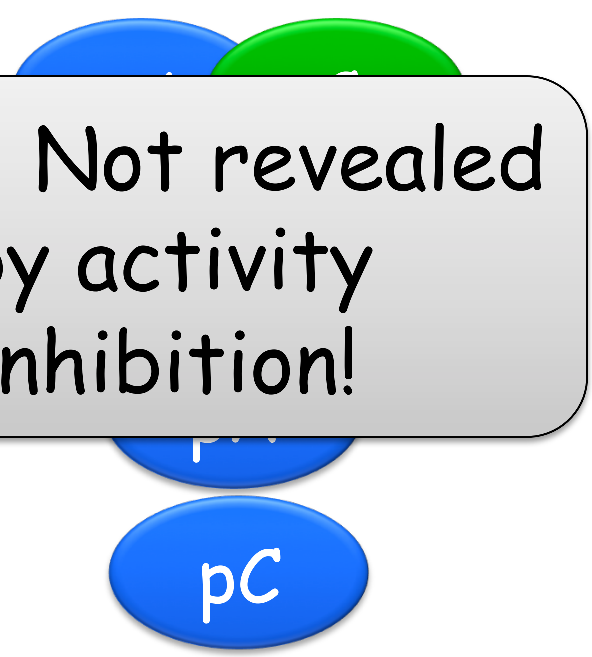


Ground truth



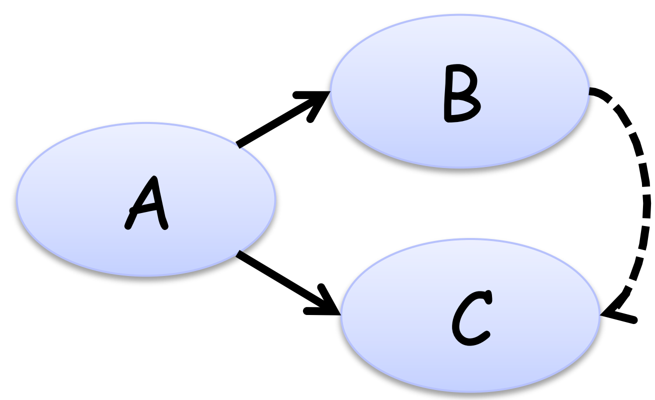
Zoom in on signaling



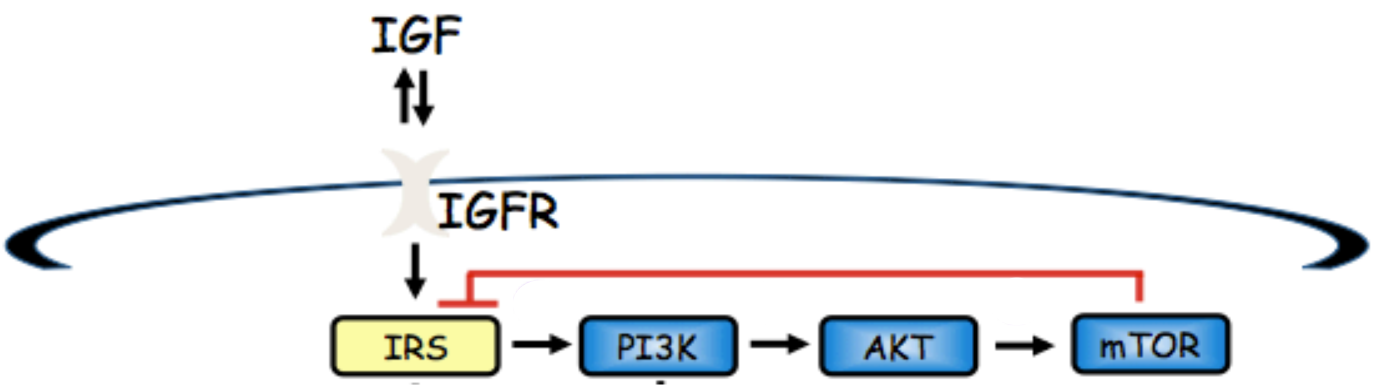


Note: Not revealed by activity inhibition!

What happens when [A] is limiting?



Extra edge: causal via competition

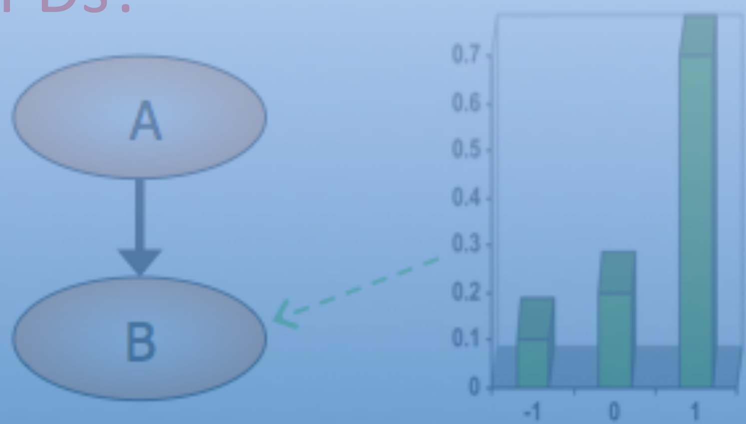


Ground truth

Model results

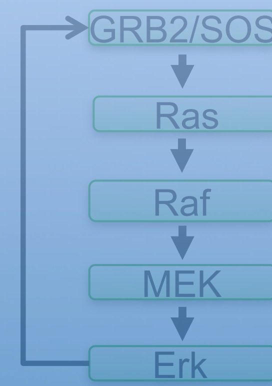
Remaining challenges

- CPDs?

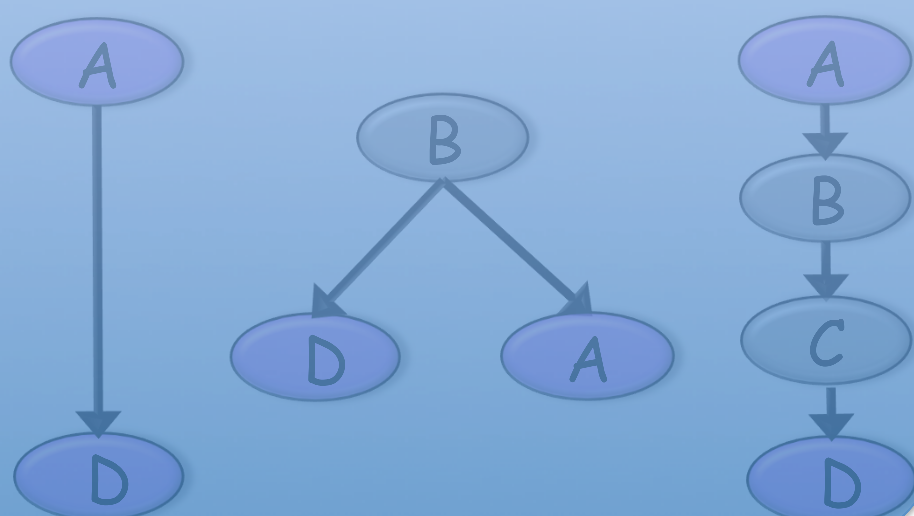


- Cycles?

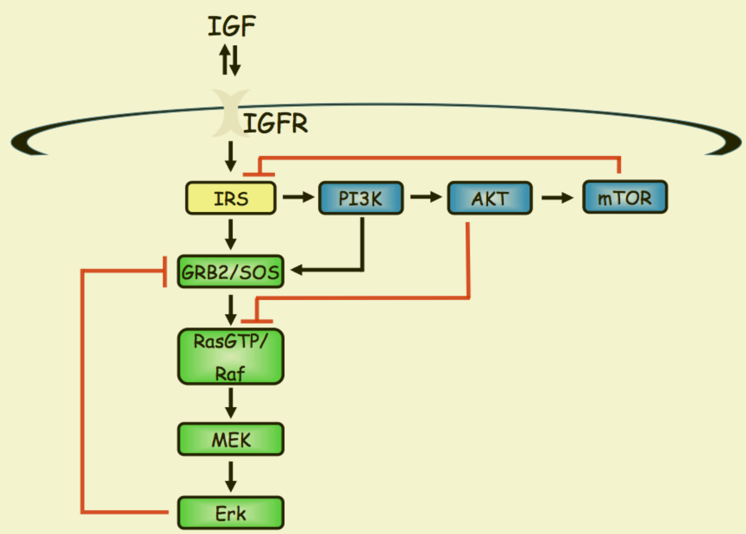
Sachs and Itani, IEEE 2008, PSB 2009, JMLR Proc 2008



- Hidden variables?

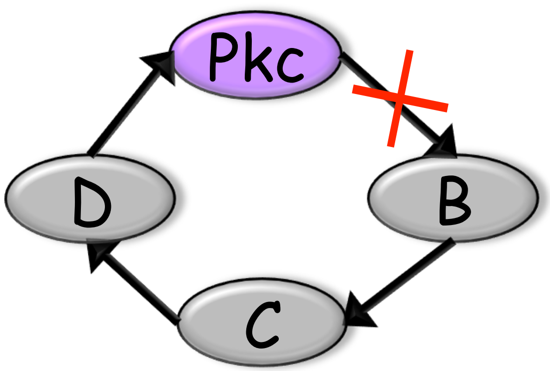


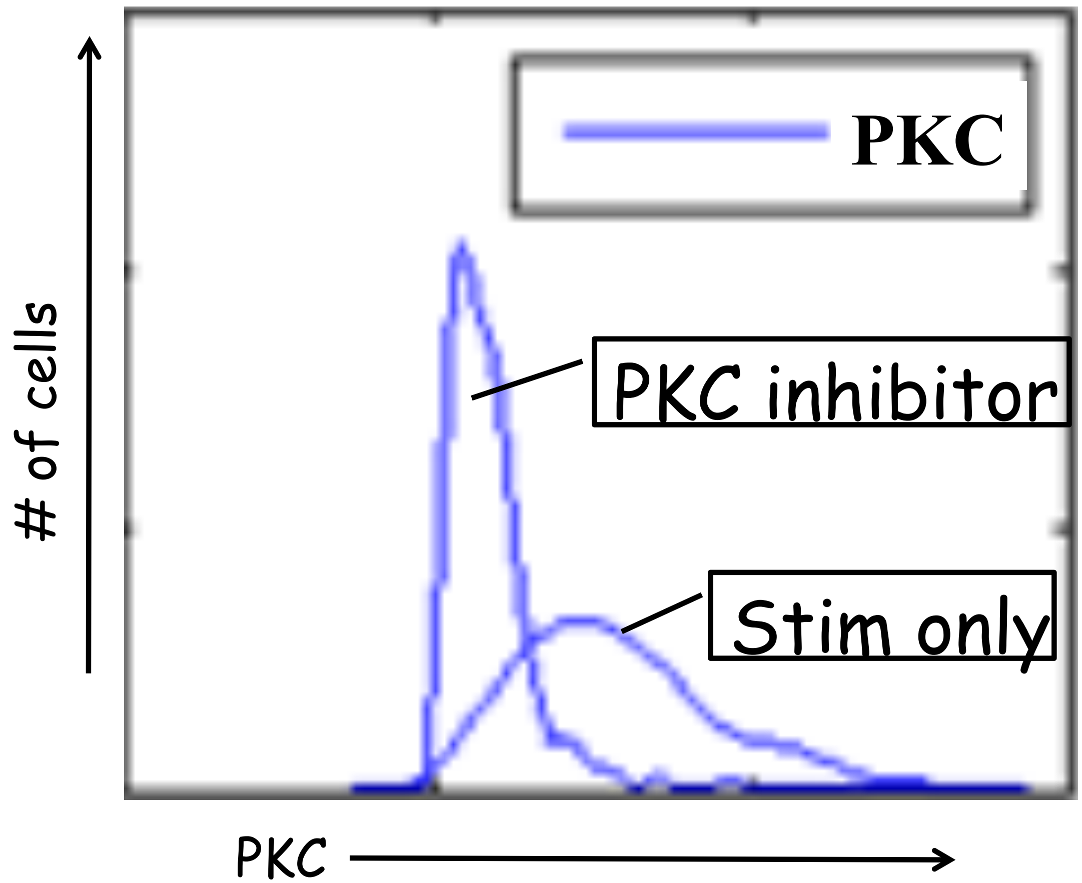
- Dynamics?



Sachs, Interface Focus 2013

Cyclic structure learning





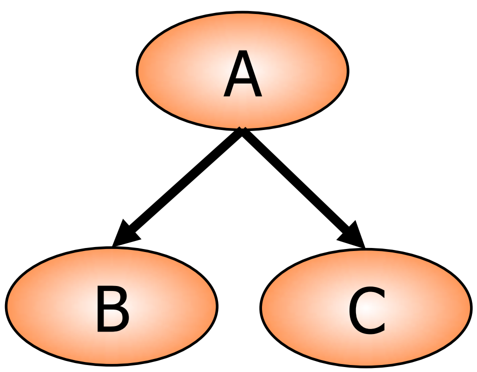
Itani and Sachs et al, JMLR Proc 2008

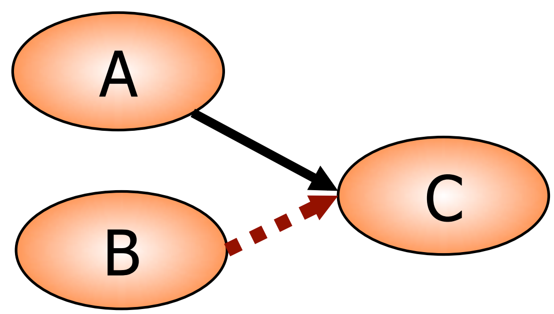
Dynamics can confound causality: Example

$A(t)$ randomly reset to 0 or 1

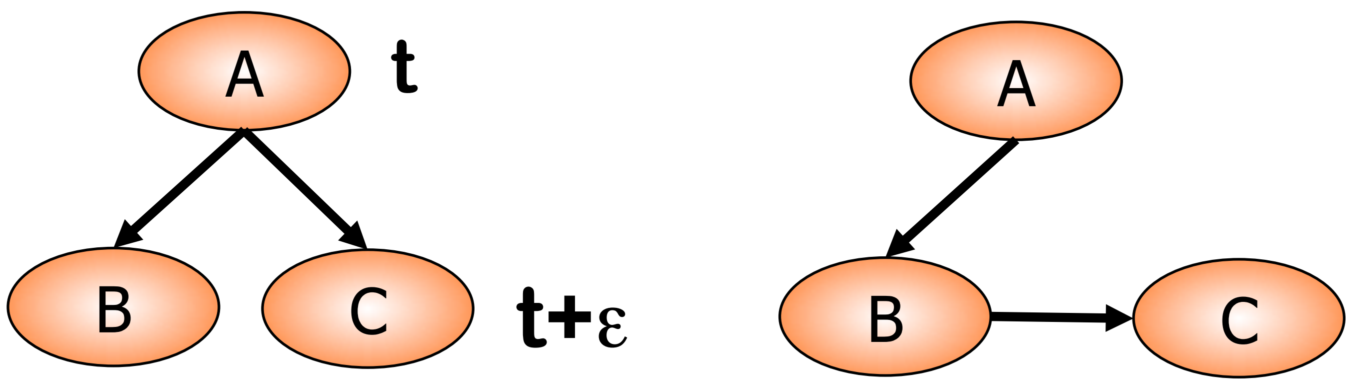
$$B=A(t-1),$$

$$C(t)=A(t) \vee A(t-1)$$

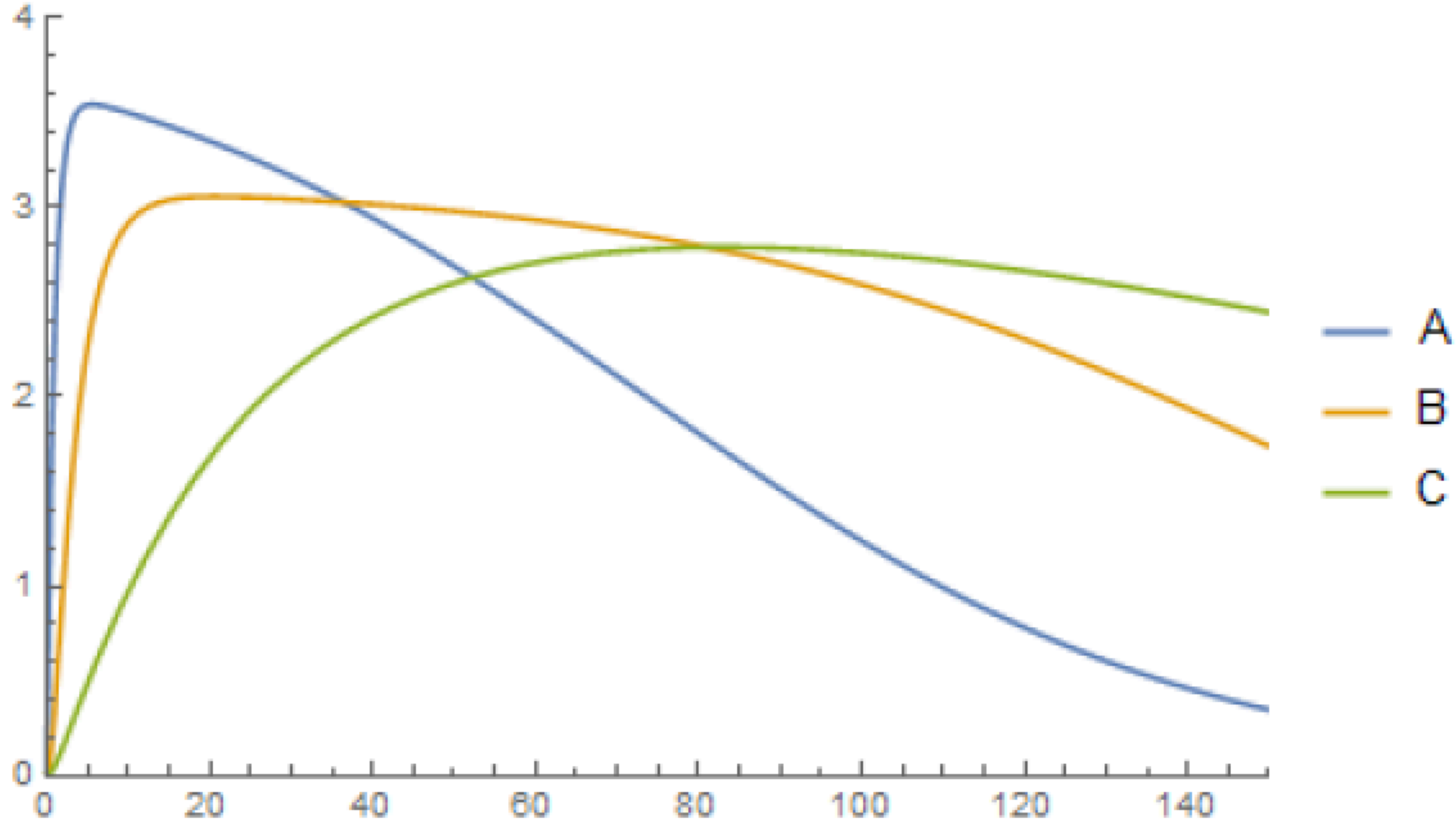


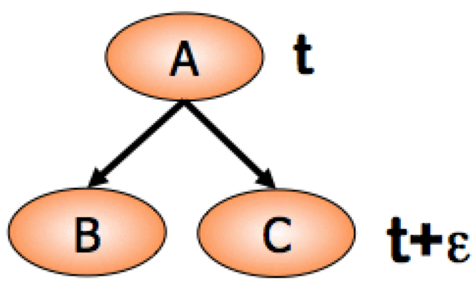


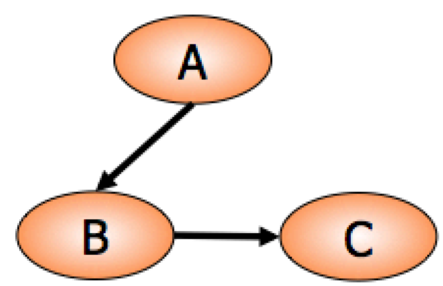
Bio Example

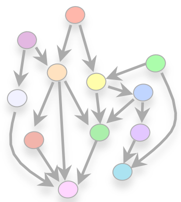


Bio Example: C depends on the *history* of A

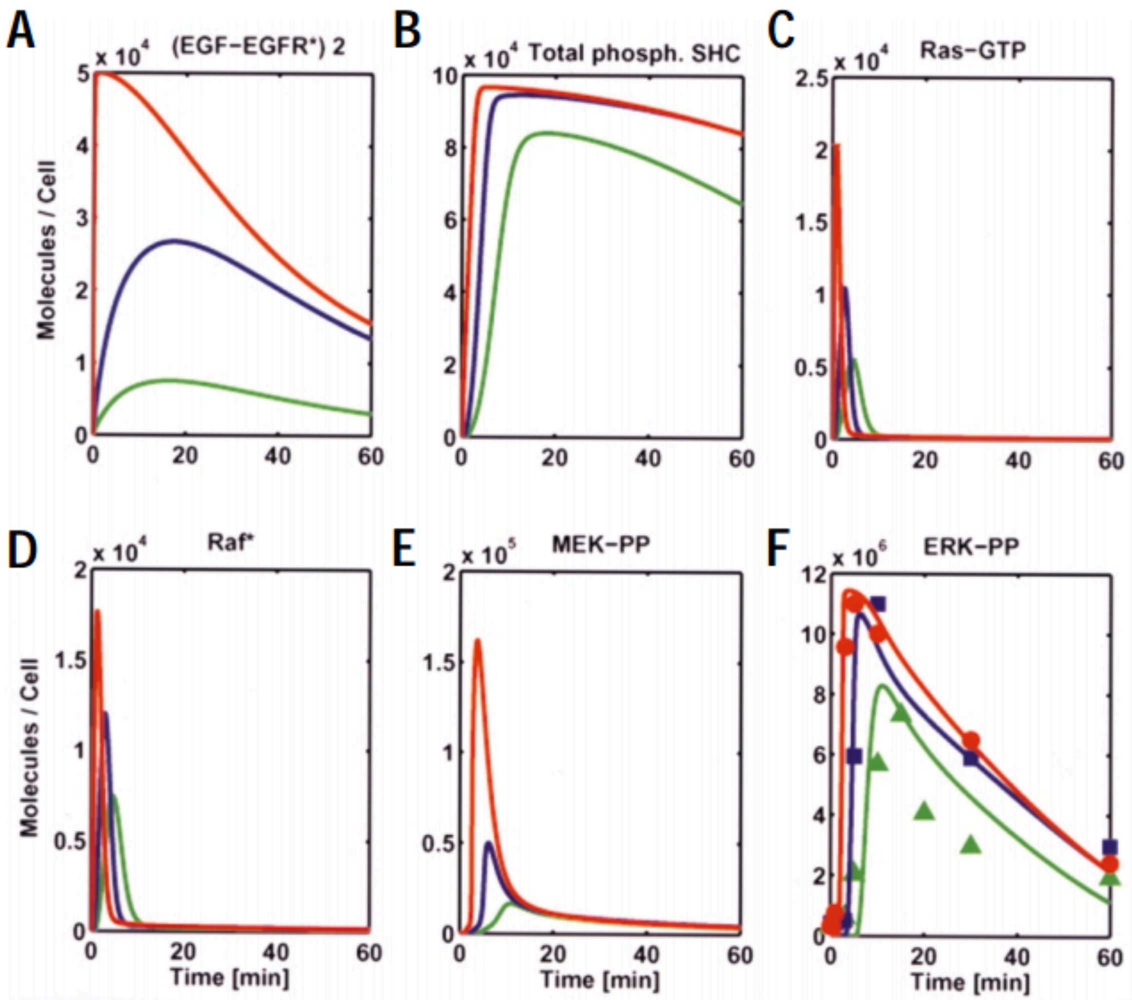






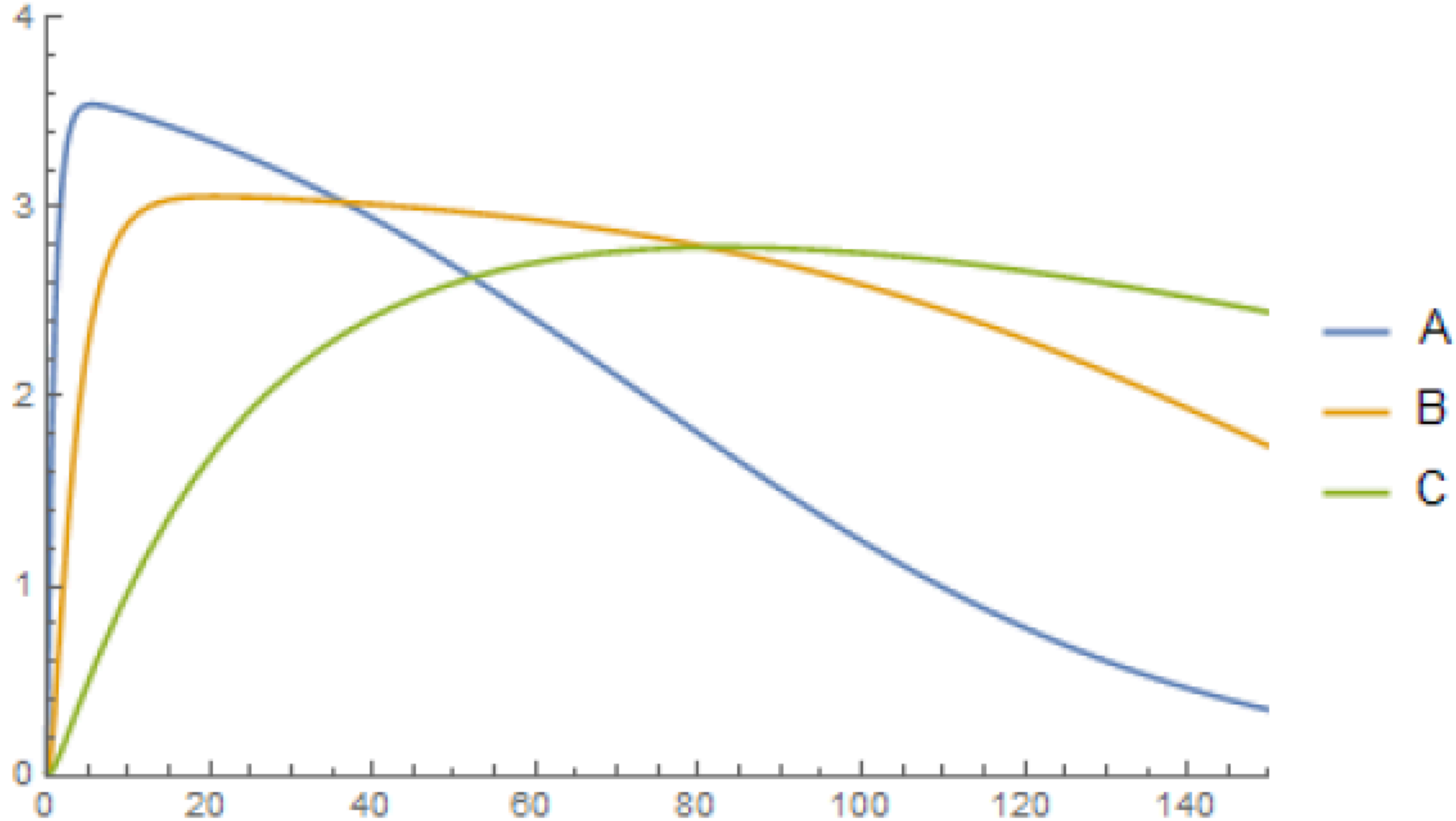


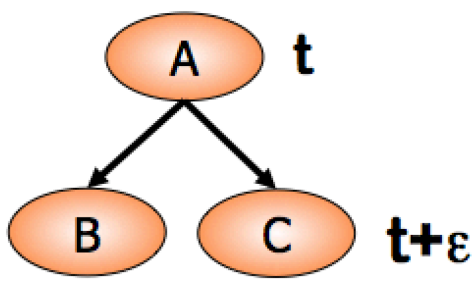
We are generally not in SS

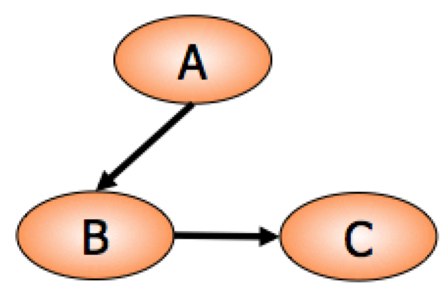


Schoeberl et al 2002

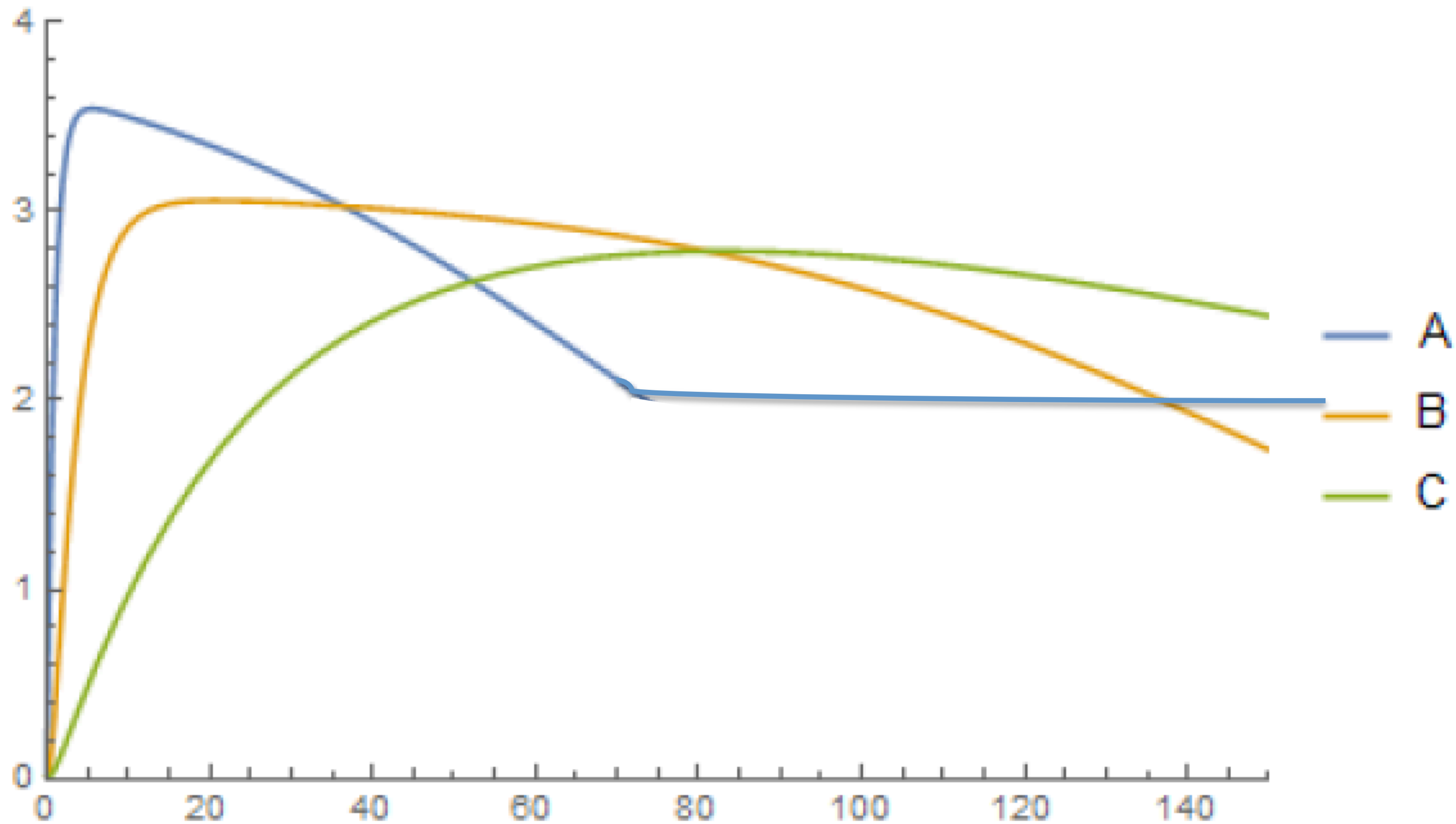
Avoid dependence on A's history? (How?)

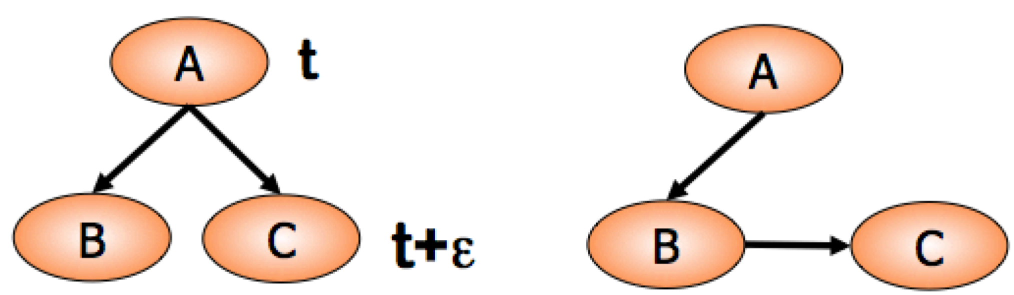






Hold A constant!

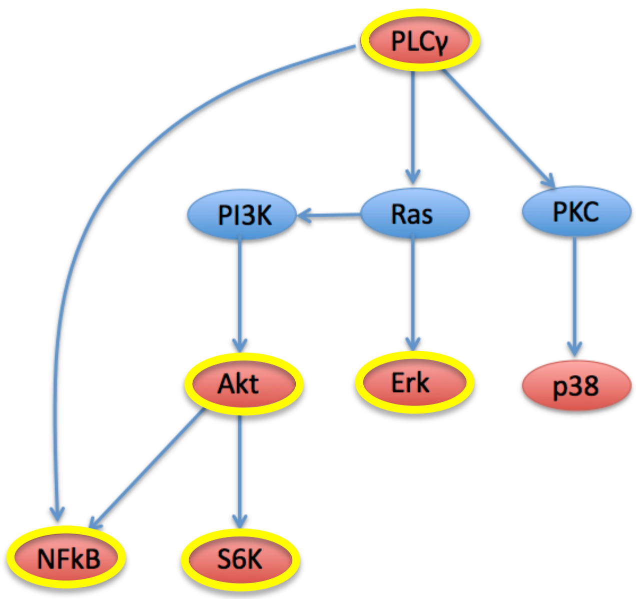




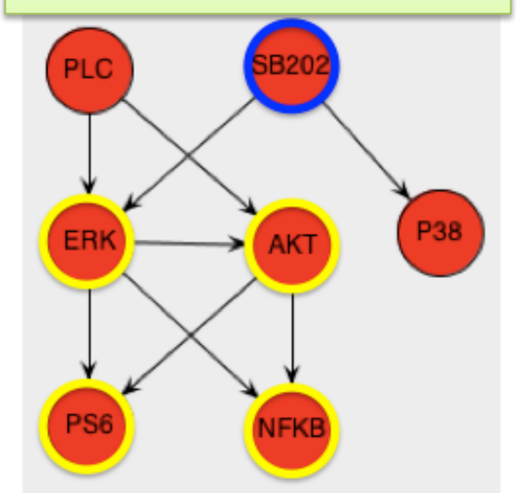
Algorithm for reducing noncausal edges

- Avoid dependence on history by learning from multi-inhibited conditions
- Some formalized results (see past talk)
- Continuing work based on feedback (work in progress)

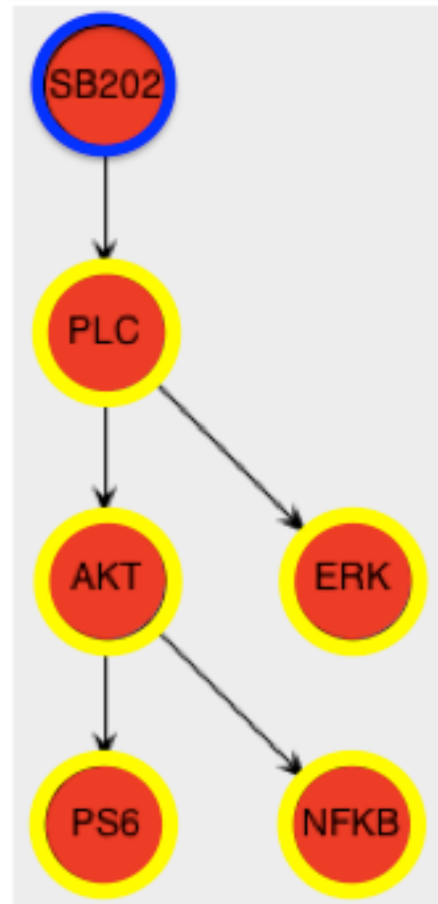
Reconstruction in T Cells



Standard BN

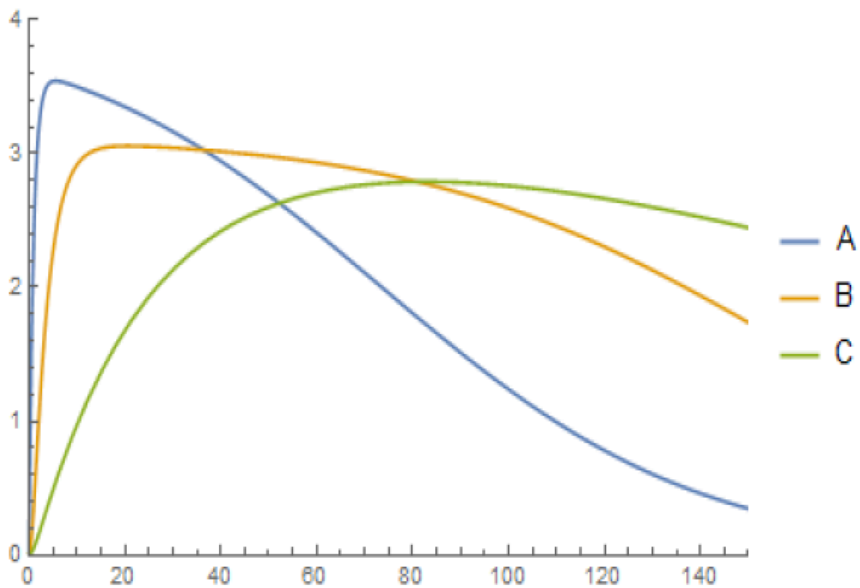


Combined inhibitions



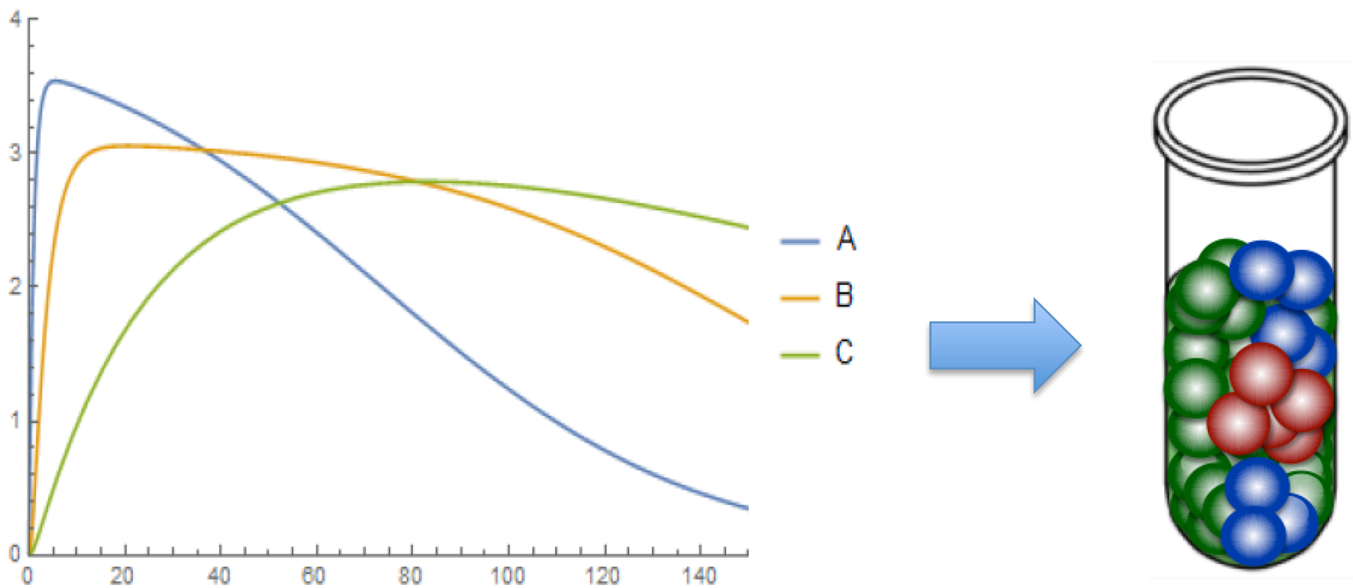
Other approaches?

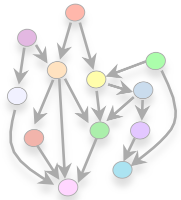
Other approaches?



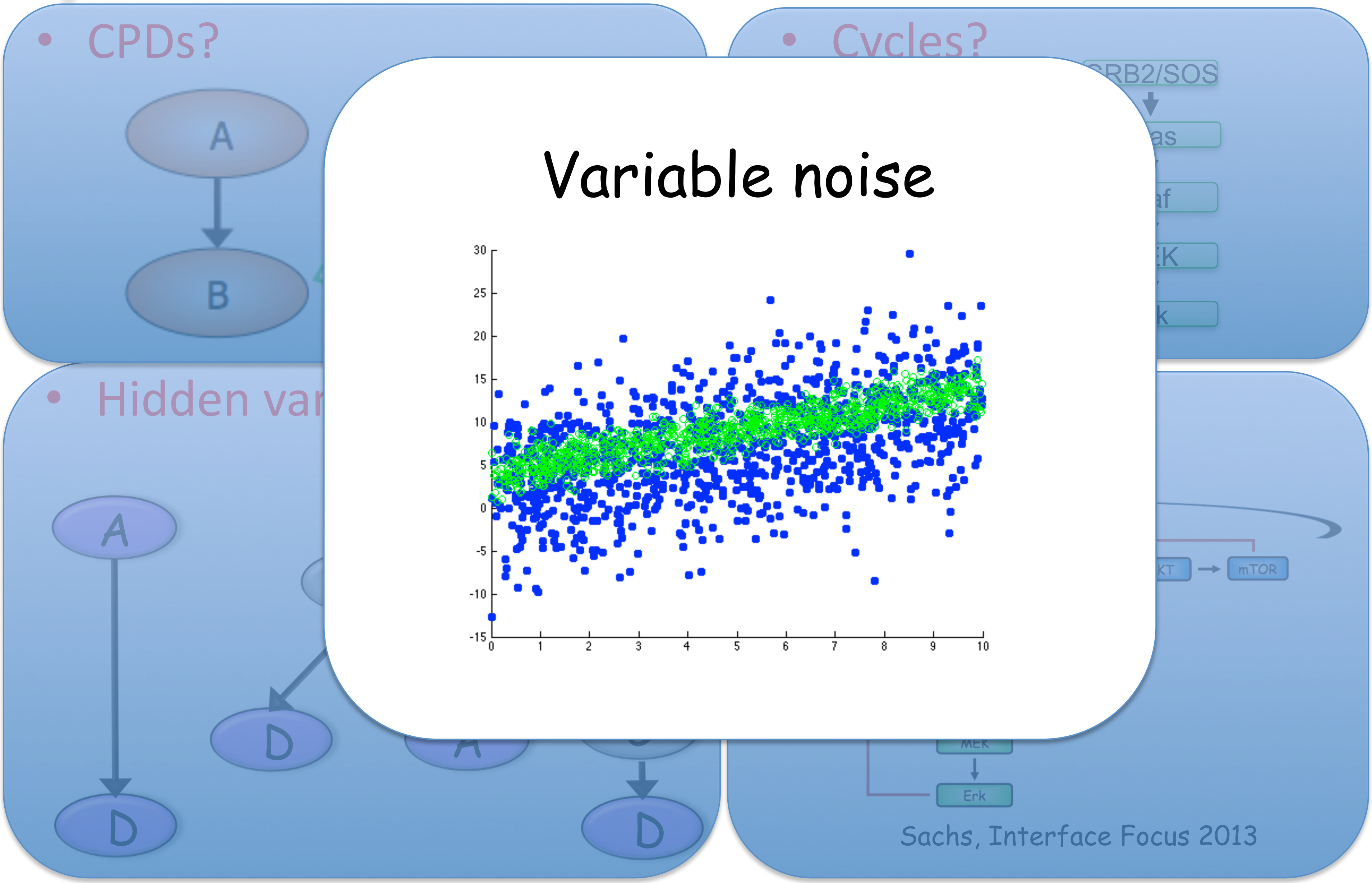
Other approaches?

How much fits into one snapshot?





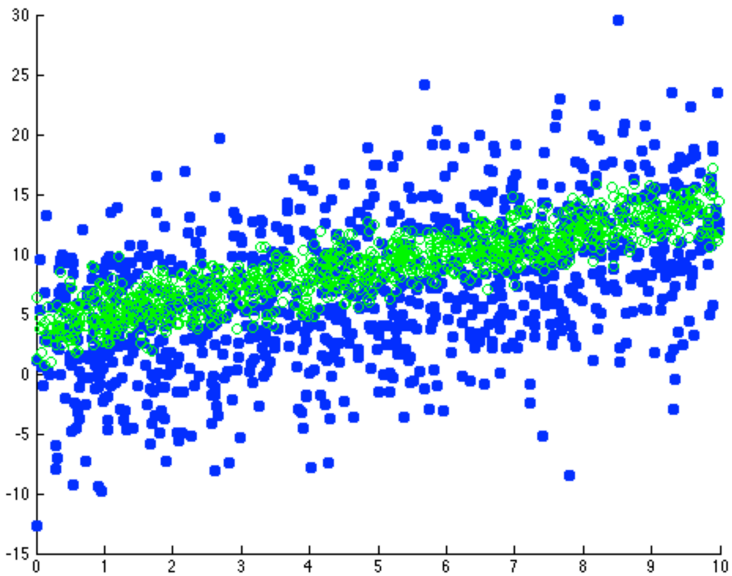
Bonus challenge



Bonus challenge: Variable noise

B has higher measurement noise

A



B
C

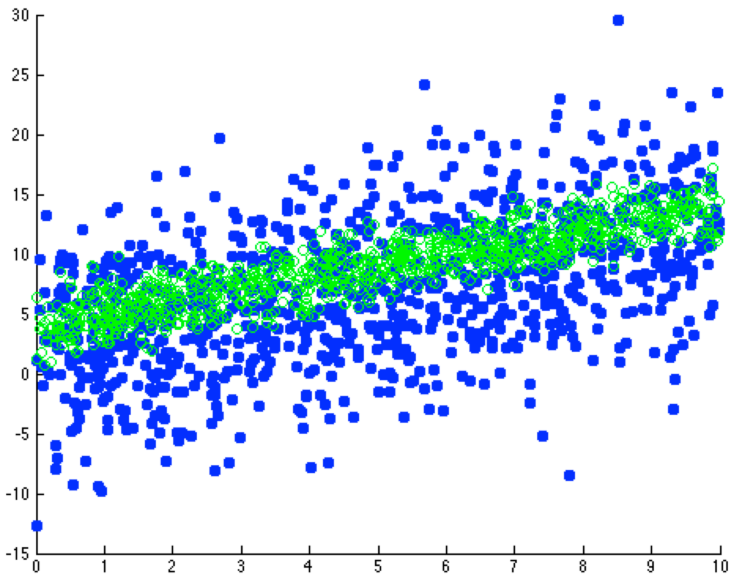
A → B → C

Ground truth

Bonus challenge: Variable noise

B has higher measurement noise

A



B

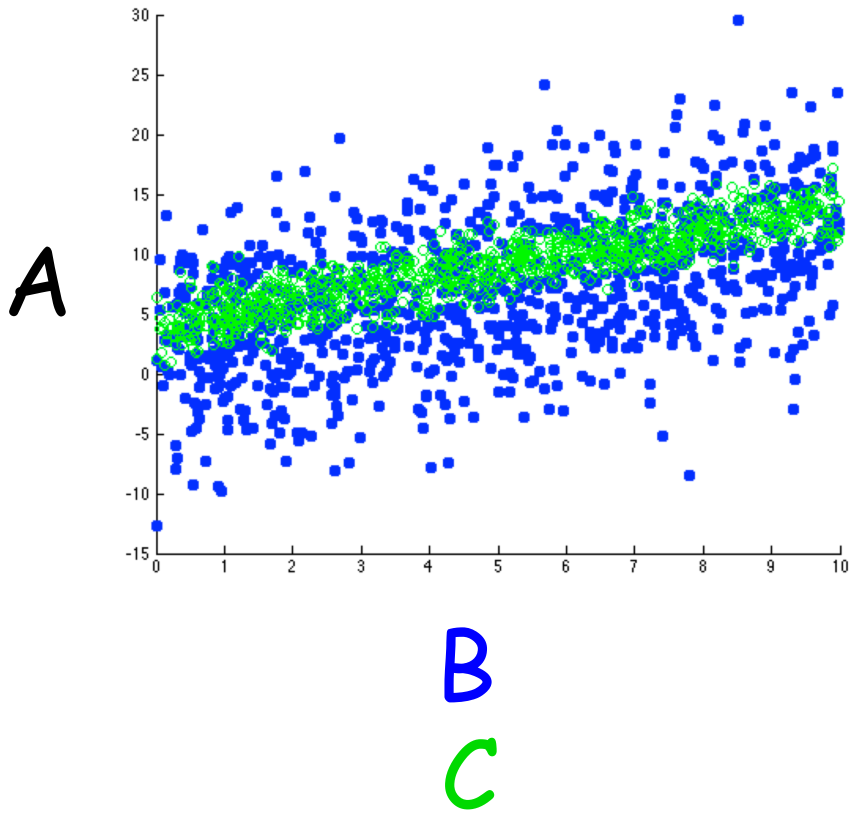
C

A → B → C

Ground truth

Bonus challenge: Variable noise

B has higher measurement noise



A → B → C

Ground truth

A → C → B

Learned model

Acknowledgements

Team Stanford

Garry Nolan
Tiffany Chen
Tim Lee
Peng Qiu
Sean Bendall
Erin Simonds

Wendy Fantl
Tyler Burns
Astraea Jager
Andrew Gentles
Sylvia Plevritis
Mingyu Chung

Team Berkeley, UMN, Merrimack, Harvard

Mohammed AlQuraishi
Birgit Schoeberl
Jonathan Fitzgerald
Solomon Itani
Claire Tomlin
Zohar Sachs

Funding

LLS post doctoral fellowship

The Leukemia & Lymphoma Society®
Fighting Blood Cancers

Thank you!

karensachs@stanford.edu