Linking Proteomic and Transcriptional Data through the Interactome and Epigenome Reveals a Map of Oncogene-induced Signaling

Anthony Gitter

Cancer Bioinformatics (BMI 826/CS 838)

April 16, 2015

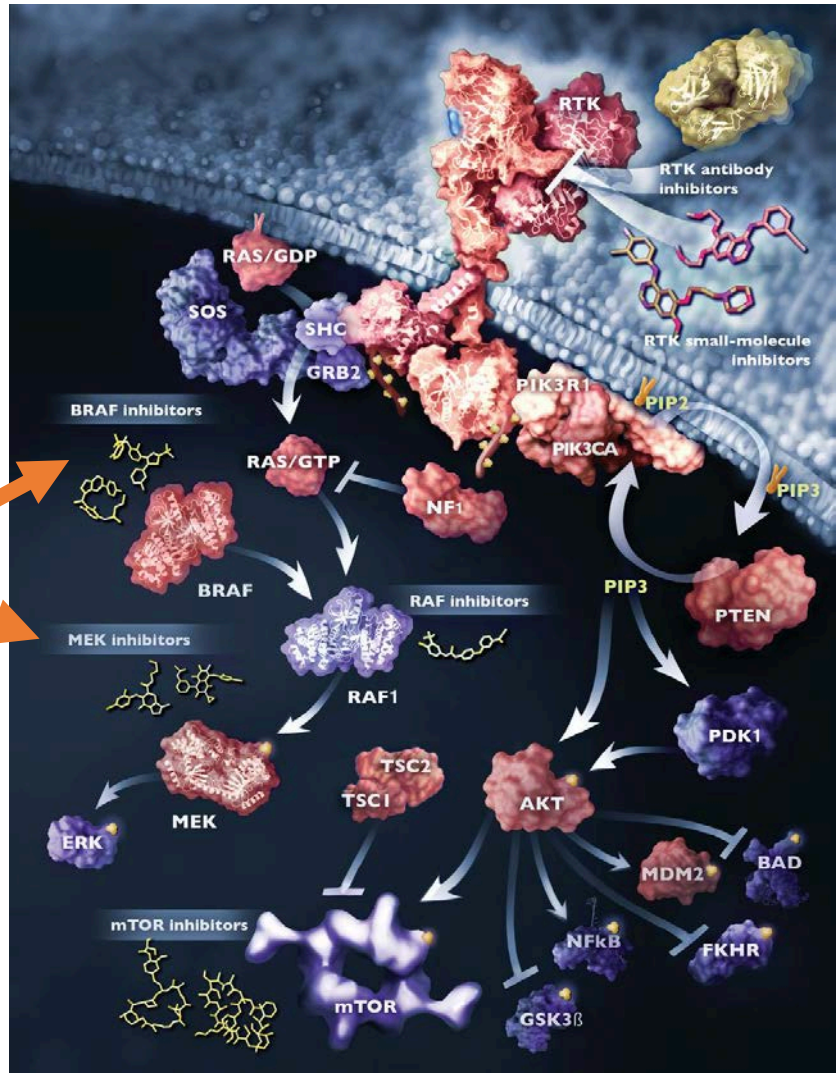
All figures and quotes from Huang2013 unless noted otherwise

Prize-Collecting Steiner Tree (PCST)

- Pathways needed to determine how heterogeneous drivers lead to cancer-related phenotypes
- PCST focuses on learning pathway structure

Pathway structure important for therapeutics

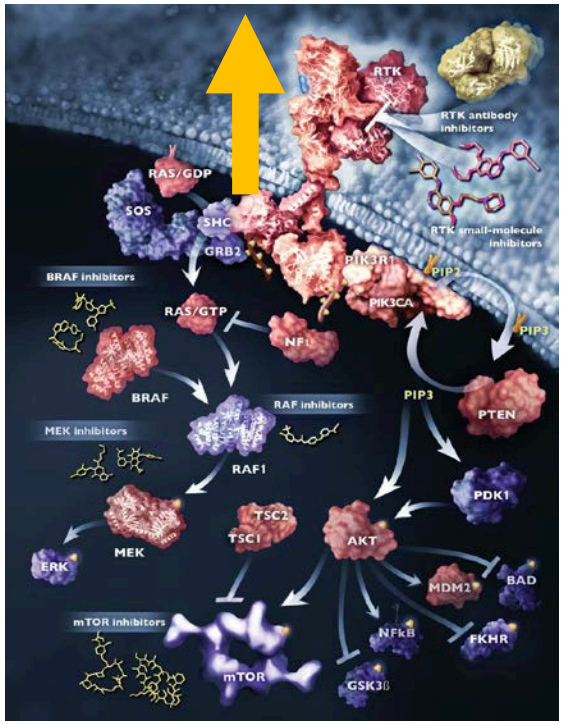
Kinase inhibitors target specific pathway members



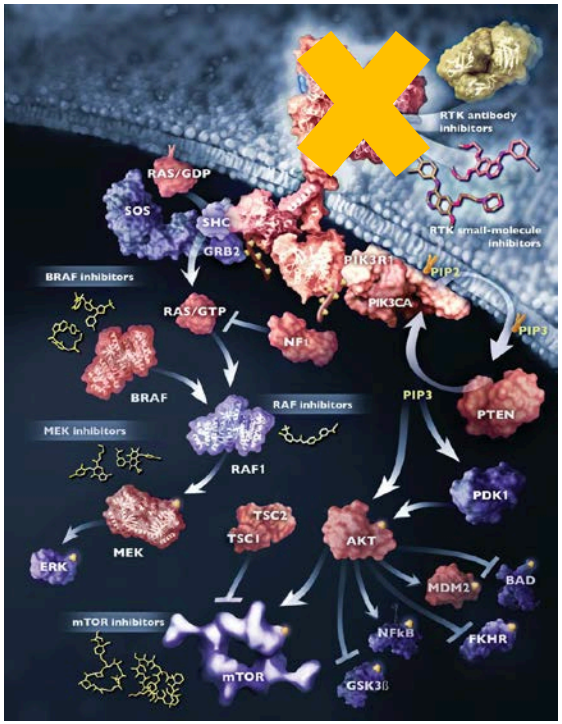
Vogelstein2013

Glioblastoma model

U87H – **H**igh EGFRvIII



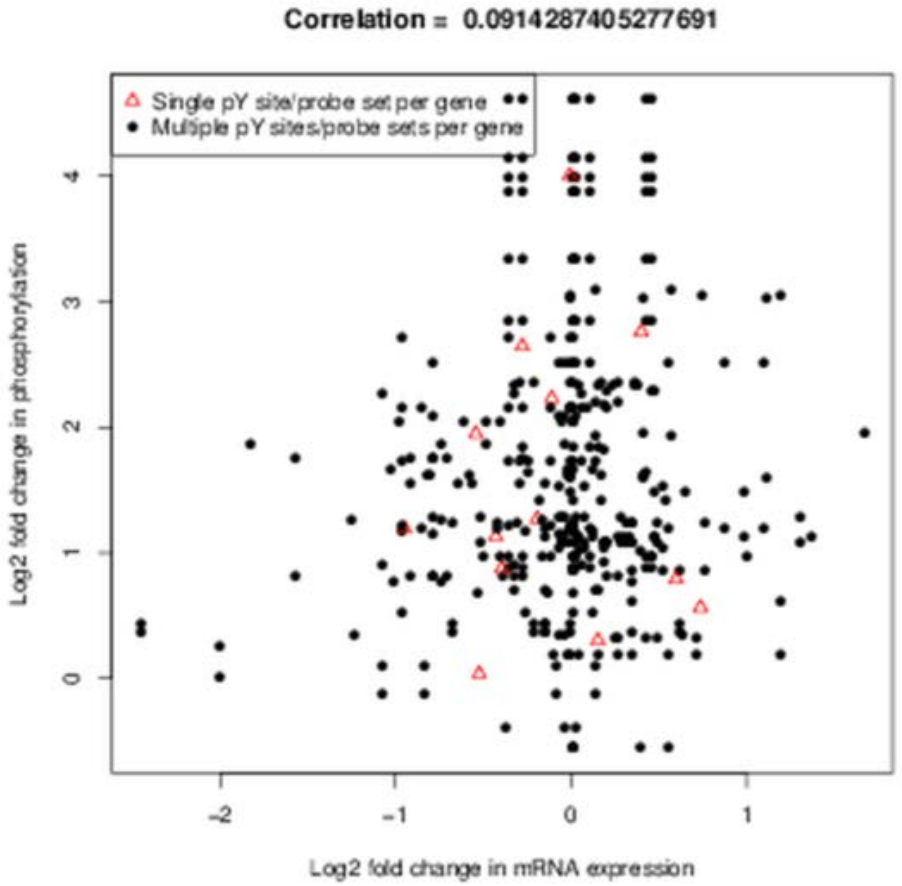
U87DK – **D**ead **K**inase



- Phosphorylation changes (88 proteins)
- DNase hypersensitivity changes (~13000 regions)
- Gene expression changes (1623 genes)

Pathways for integrating data

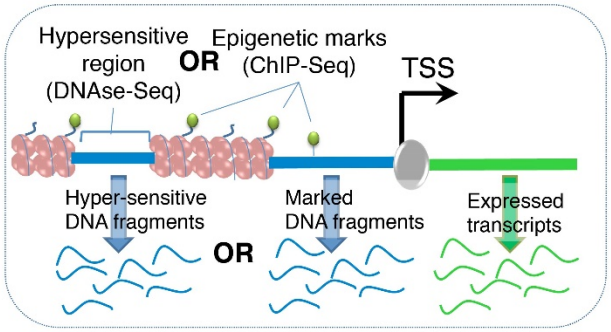
- Low correlation between phosphorylation and transcription
- Provide complementary information



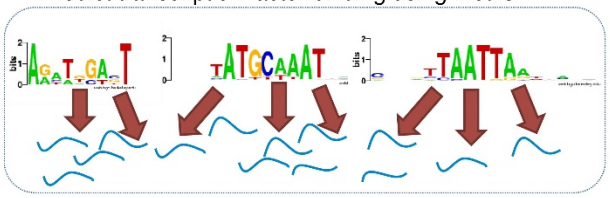
Two stages: define prizes, solve PCST

Garnet

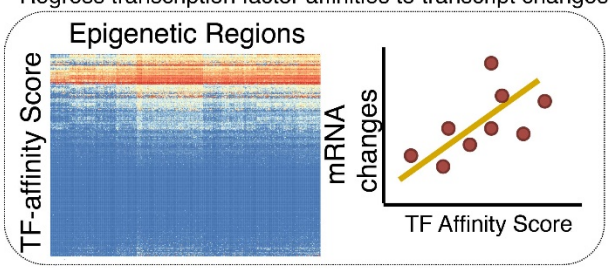
Map epigenetic regions to expressed transcripts



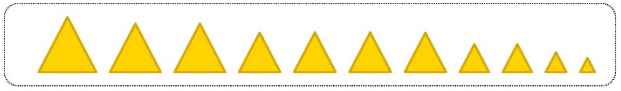
Predict transcription factor binding using motifs



Regress transcription factor affinities to transcript changes

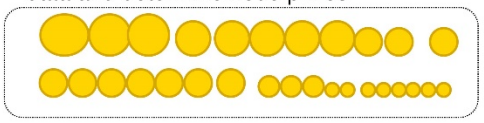


Select transcription factors as terminal nodes and assign prizes

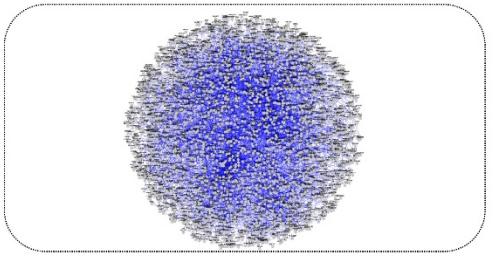


Forest

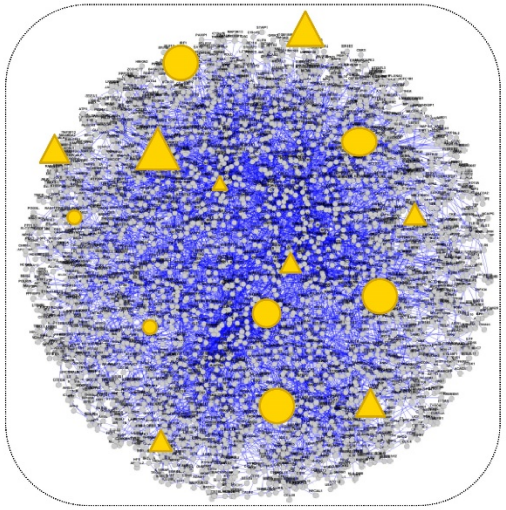
Define terminal node set from experimental data and determine node prizes



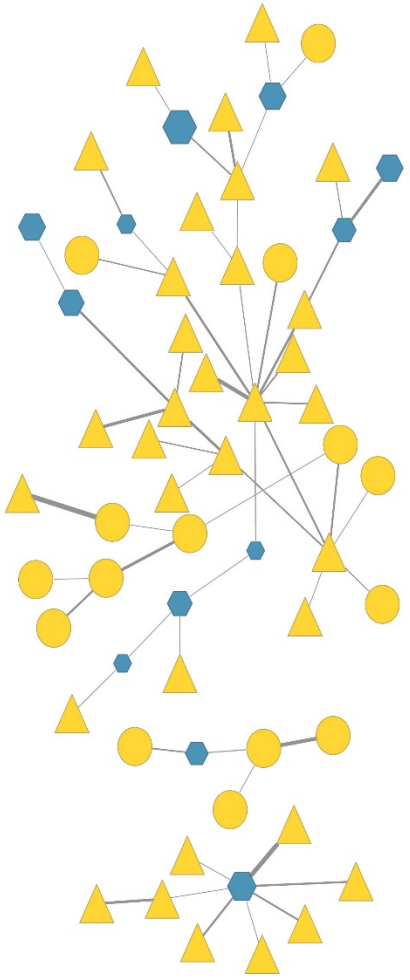
Collect weighted interactome from literature or construct your own



Weight nodes in graph by experimental data



Result



[OmicsIntegrator]

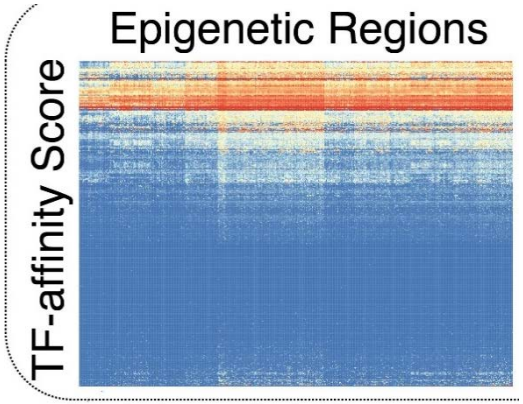
Stage 1: phosphorylation prizes

- Prize based on phosphorylation fold change
- Proteins with prizes called terminals
- p: protein
- $prize_{ph}$: phosphorylation prize

$$prize_{ph}(p) = \left| \log_2 \frac{phospho_{U87H}\,(p)}{phospho_{U87DK}\,(p)} \right|$$

Stage 1: TF prizes

- Find differentially expressed genes
- Find differential DNase peaks
- Create gene X TF score matrix



$$x_{g,\tau} = \sum_{i=1}^{|S_{g,H}|} x_{i,g,\tau,H} - \sum_{i=1}^{|S_{g,DK}|} x_{i,g,\tau,DK}$$

Peaks mapped to gene *g* in U87H cells

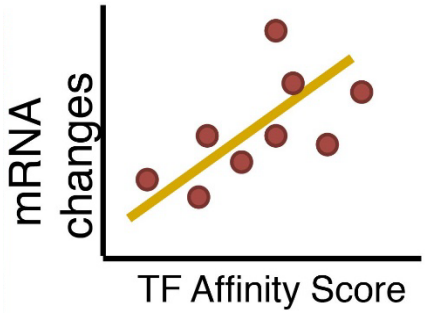
Peaks mapped to gene *g* in U87DK cells

DNA binding motif affinity for TF τ at score for peak *i* in condition *c* proximal to gene *g*

Stage 1: TF prizes continued

- Test each TF independently for association with differentially expressed gene *g*
- Are coefficients in linear regression significantly different from 0?

$$y_g = \alpha_\tau x_{g,\tau} + \varepsilon_g$$

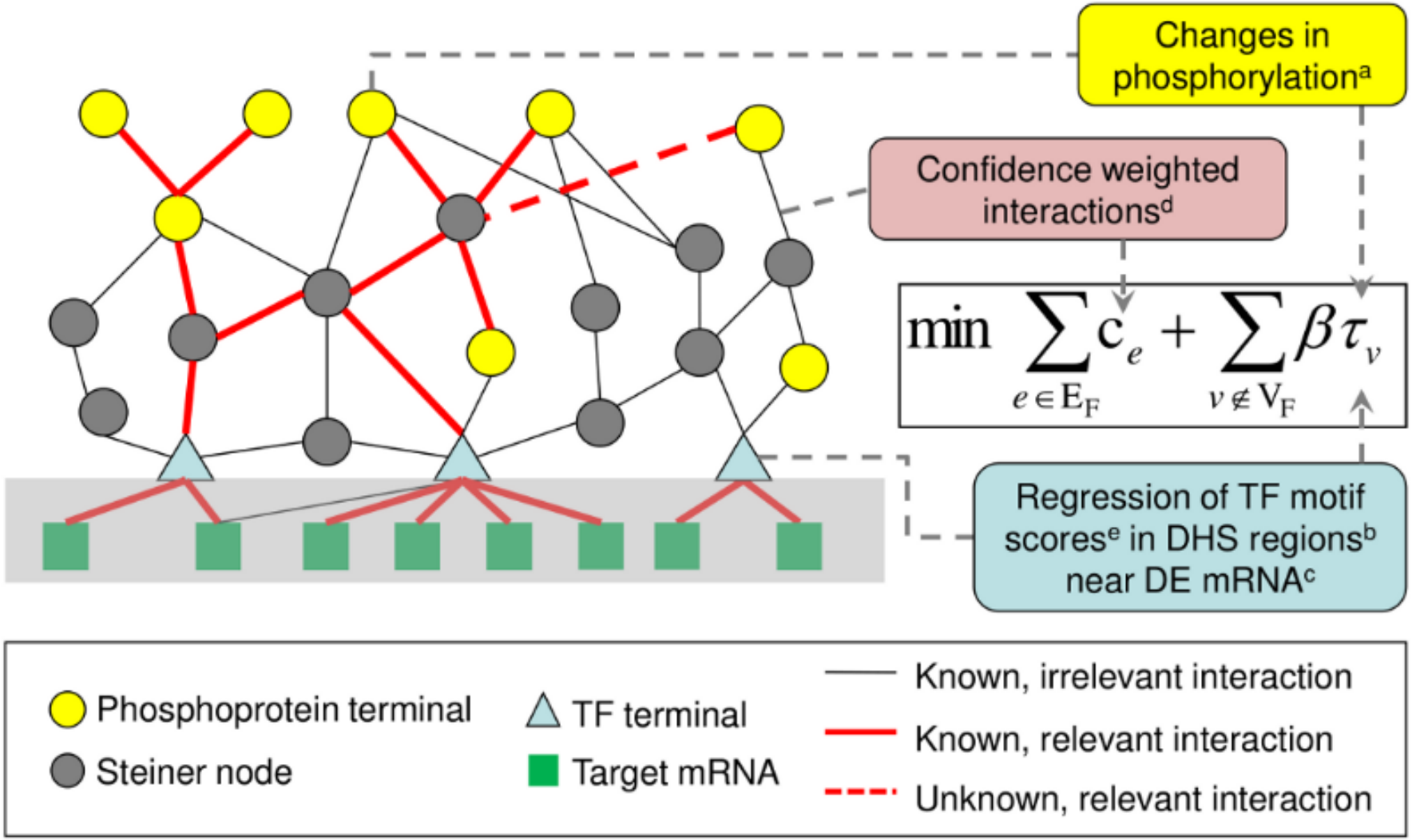


Gene *g* $\log_2$ expression fold change

Binding affinity for TF τ near gene *g*

Regression coefficient, use t-test statistic for prize

Stage 2: identify subnetwork



Solving PCST

$$o(F) = \beta \sum_{v \notin V_F} p(v) + \sum_{e \in E_F} c(e)$$

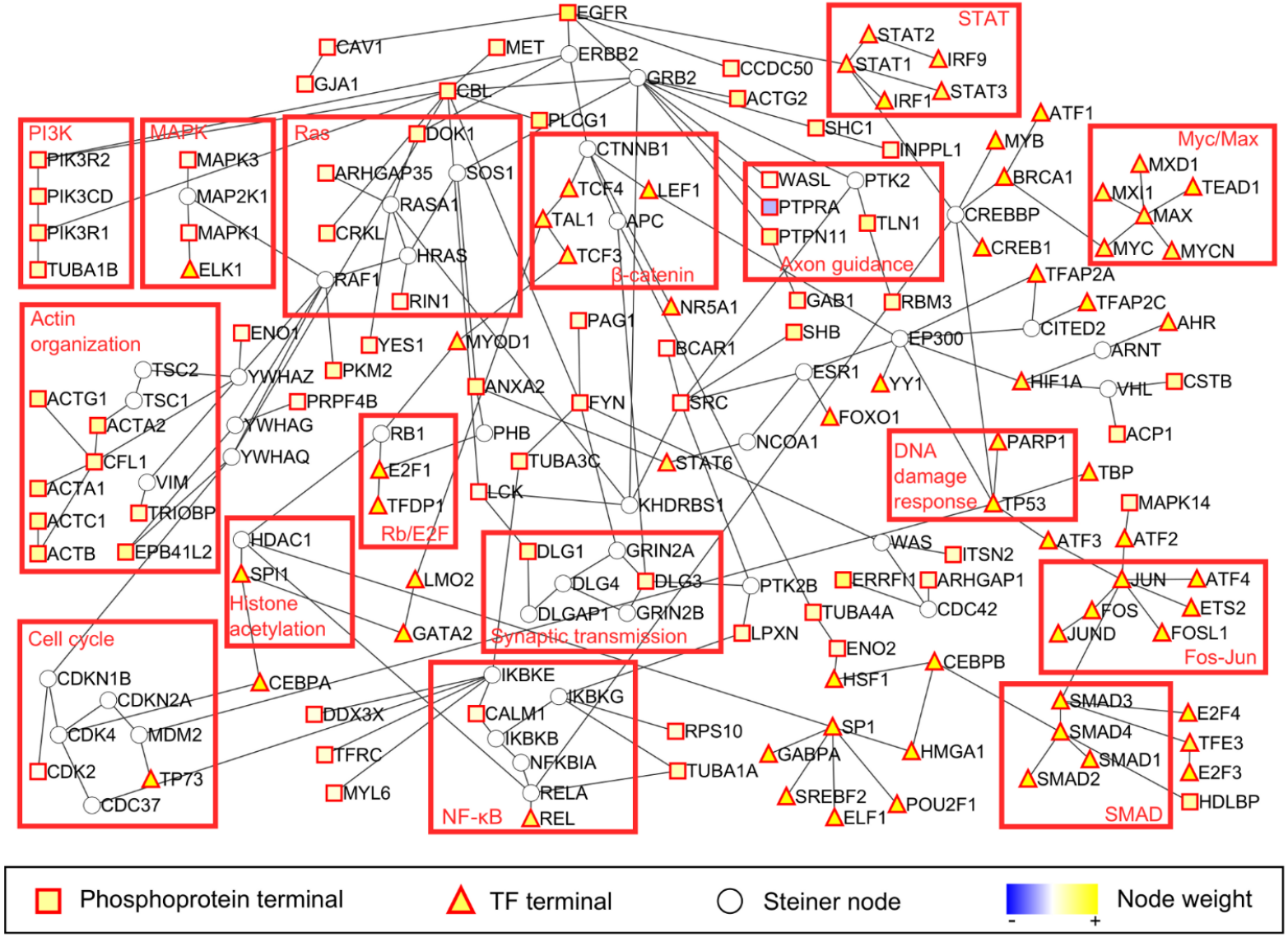
Prize vs. edge cost tradeoff

Cost of selected edges

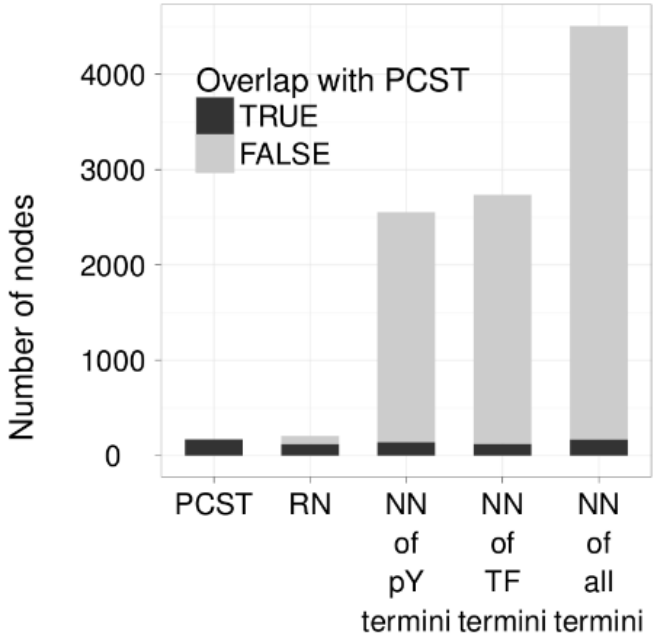
Penalty for omitted prize nodes

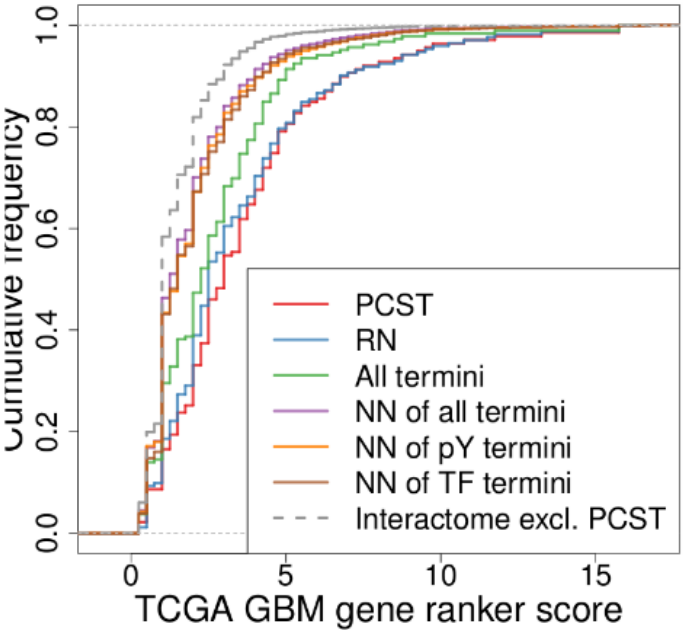
- Use off-the-shelf Steiner tree solver
- Solver creates an integer program

EGRFvIII signaling pathway



Comparing with other network approaches

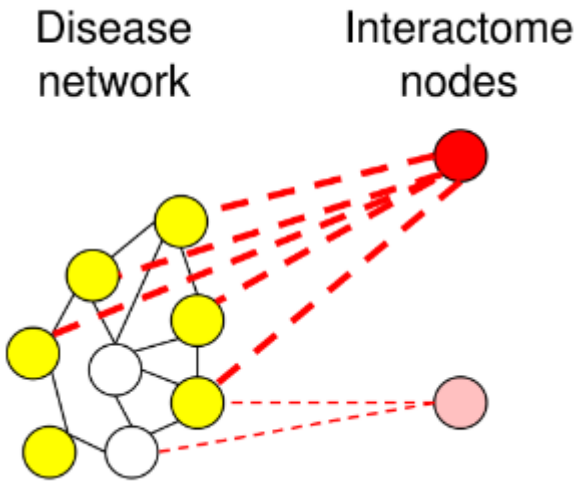




- Also compare to xenograft phosphorylation

Using pathway structure for validation

- Which proteins to test?
- What are appropriate negative controls?



$$score(v_i) = \sum_{e_{ij} \in E, v_j \in V_F} \left(1 - c(e_{ij})\right)$$

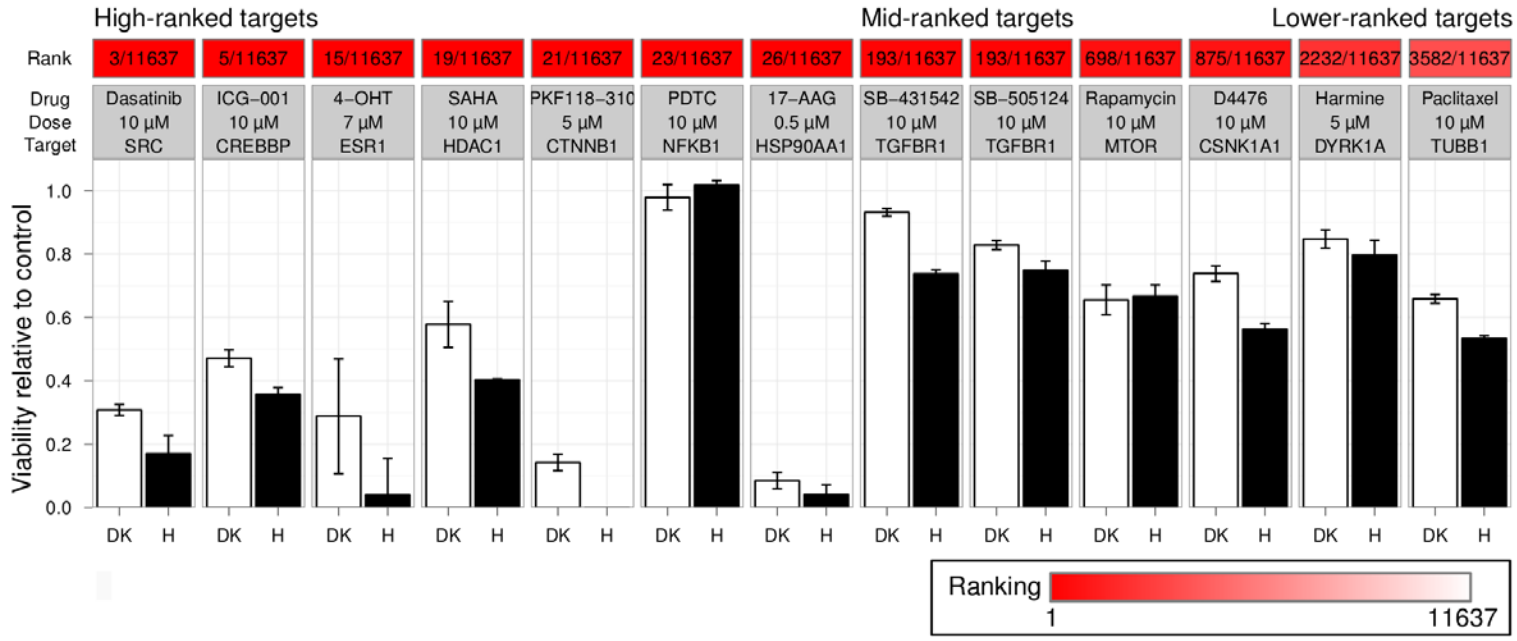
Three tiers of proteins to test

Experiment	Small molecule inhibitor	Antibody	Target	Target rank	Target type
Viability	Dasatinib		SRC	3	High-ranked target
			FYN	12	
ChIP-Seq		sc-585x	EP300	4	High-ranked target
Viability	ICG-001		CREBBP	5	High-ranked target
Viability	4-hydroxytamoxifen (4-OHT)		ESR1	15	High-ranked target
Viability	suberoylanilide hydroxamic acid (SAHA)		HDAC1	19	High-ranked target
Viability	PKF118–310		CTNNB1	21	High-ranked target
Viability	ammonium pyrrolidinedithiocarbamate (PDTC)		NFKB1	23	High-ranked target
Viability	17-N-Allylamino-17-demethoxygeldanamycin (17-AAG)		HSP90AA1	26	High-ranked target
Viability	SB-505124		TGFBR1	193	Mid-ranked target
Viability	SB-431542		TGFBR1	193	Mid-ranked target
			ACVR1B	1695	
Viability	Rapamycin		MTOR	698	Lower-ranked target
Viability	D4476		CSNK1A1	875	Lower-ranked target
Viability	Harmine		DYRK1A	2232	Lower-ranked target
			MAOA	8508.5	
Viability	Paclitaxel		TUBB1	3582	Lower-ranked target

For cell viability assays, the inhibitors used are listed. Note that some inhibitors have multiple targets. For ChIP-Seq, the antibody used is listed.
doi:10.1371/journal.pcbi.1002887.t001

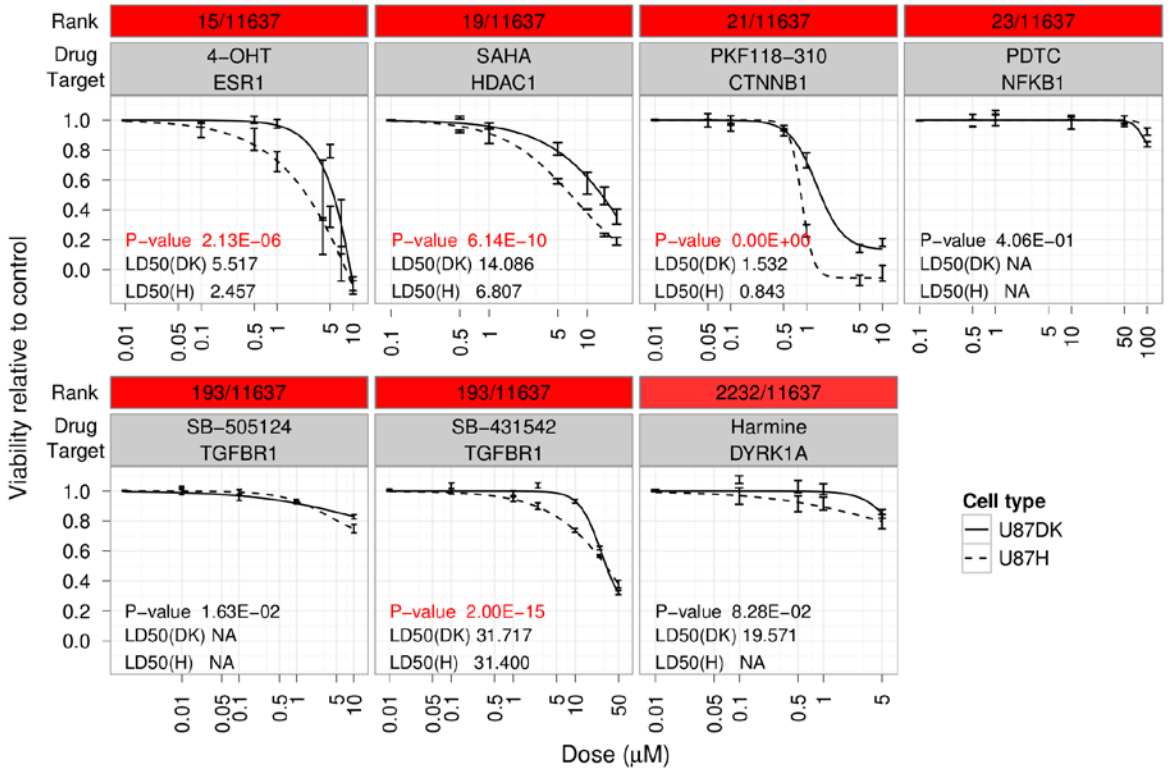
Inhibitor viability screening

- U87H cells very sensitive to inhibiting high-ranked targets



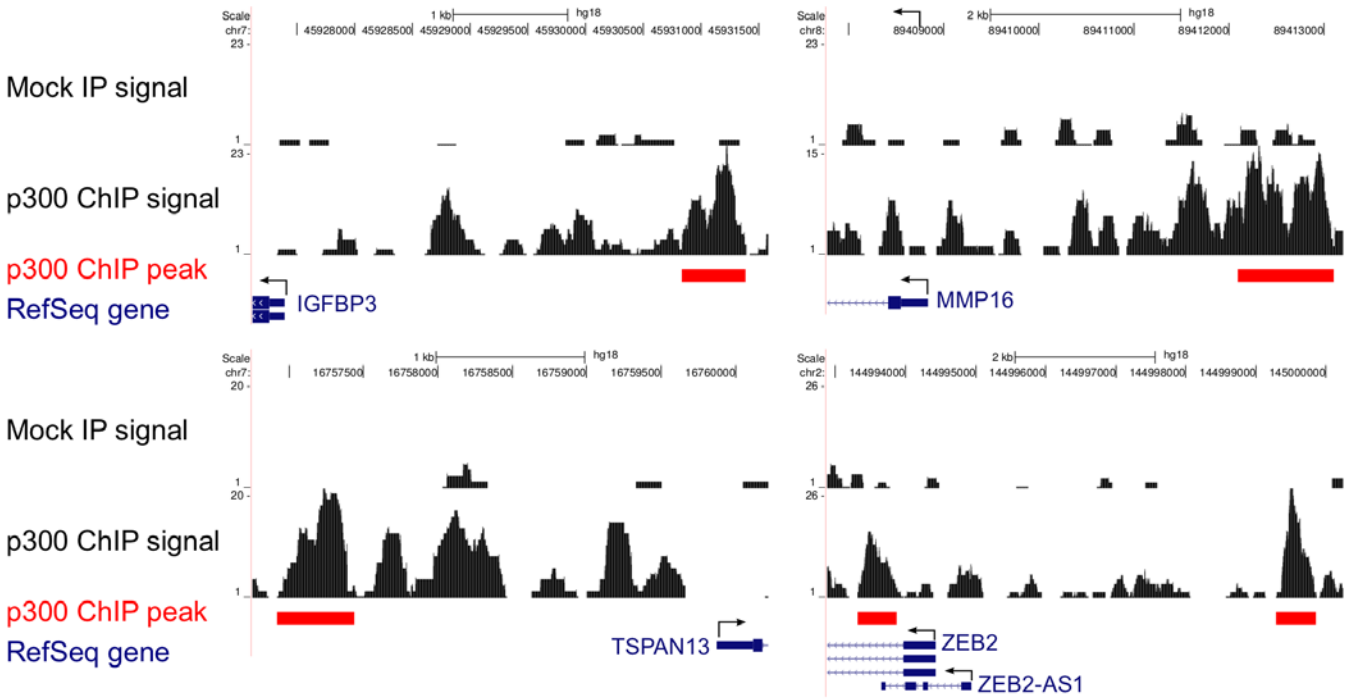
Inhibitor viability screening

- 3 of 4 top targets more toxic in U87H
- 1 of 3 lower targets more toxic, requires high dose



ChIP-seq to explore EP300 role

- EP300 is a Steiner node, top-ranked TF
- Find its targets include epithelial-mesenchymal transformation markers

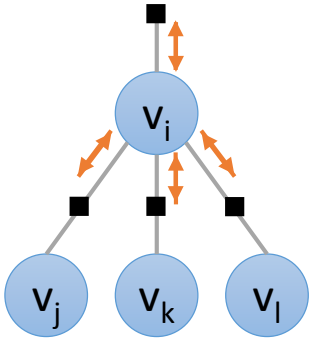


PCST versus other pathway approaches

- MEMo / Multi-Dendrix (mutual exclusivity)
- RPPA regression
- GSEA / PARADIGM (pathway enrichment / activity)
- HotNet / NBS (network diffusion)
- ActiveDriver (phosphorylation impact)

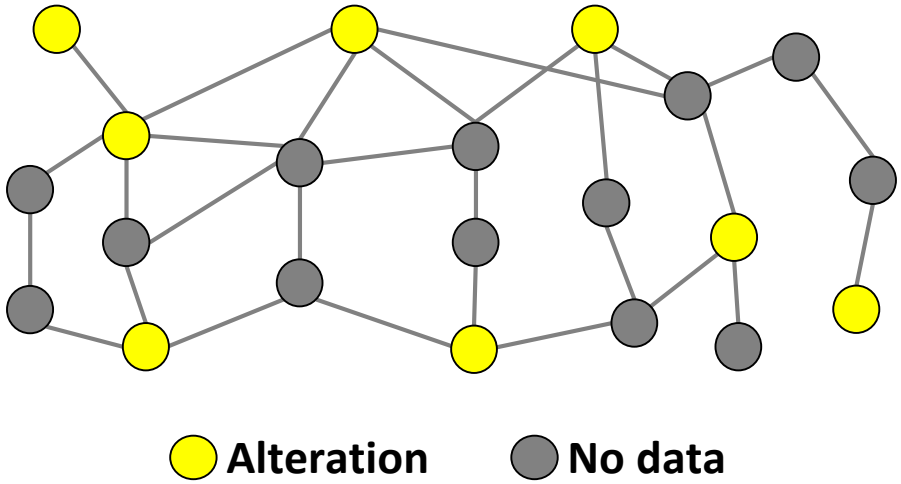
Multi-sample Prize-Collecting Steiner Forest (Multi-PCSF)

- PCST => PCSF: allow multiple disjoint trees
- PCSF => Multi-PCSF: jointly optimize pathways for many related samples (e.g. tumors)
- Approximate optimization with belief propagation instead of integer program



PCST formulation

Protein-protein interactions



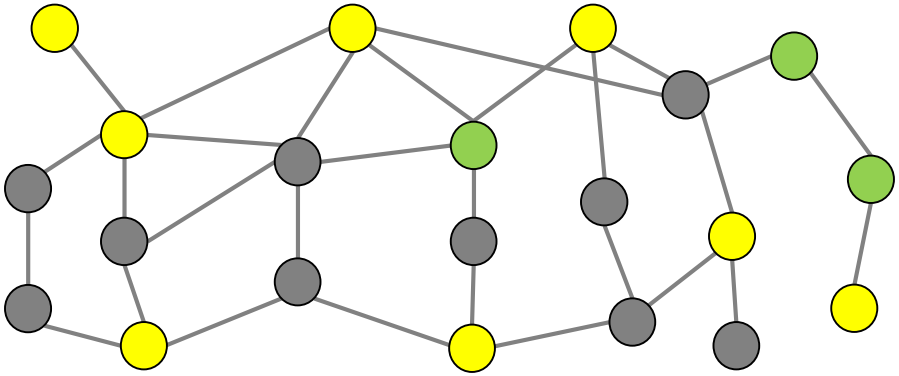
$$o(F) = \beta \sum_{v \notin V_F} p(v) + \sum_{e \in E_F} c(e)$$

Prize vs. edge cost tradeoff

Cost of selected edges

Penalty for omitted prize nodes

Multi-PCSF formulation



Alteration No data Important in related tumors

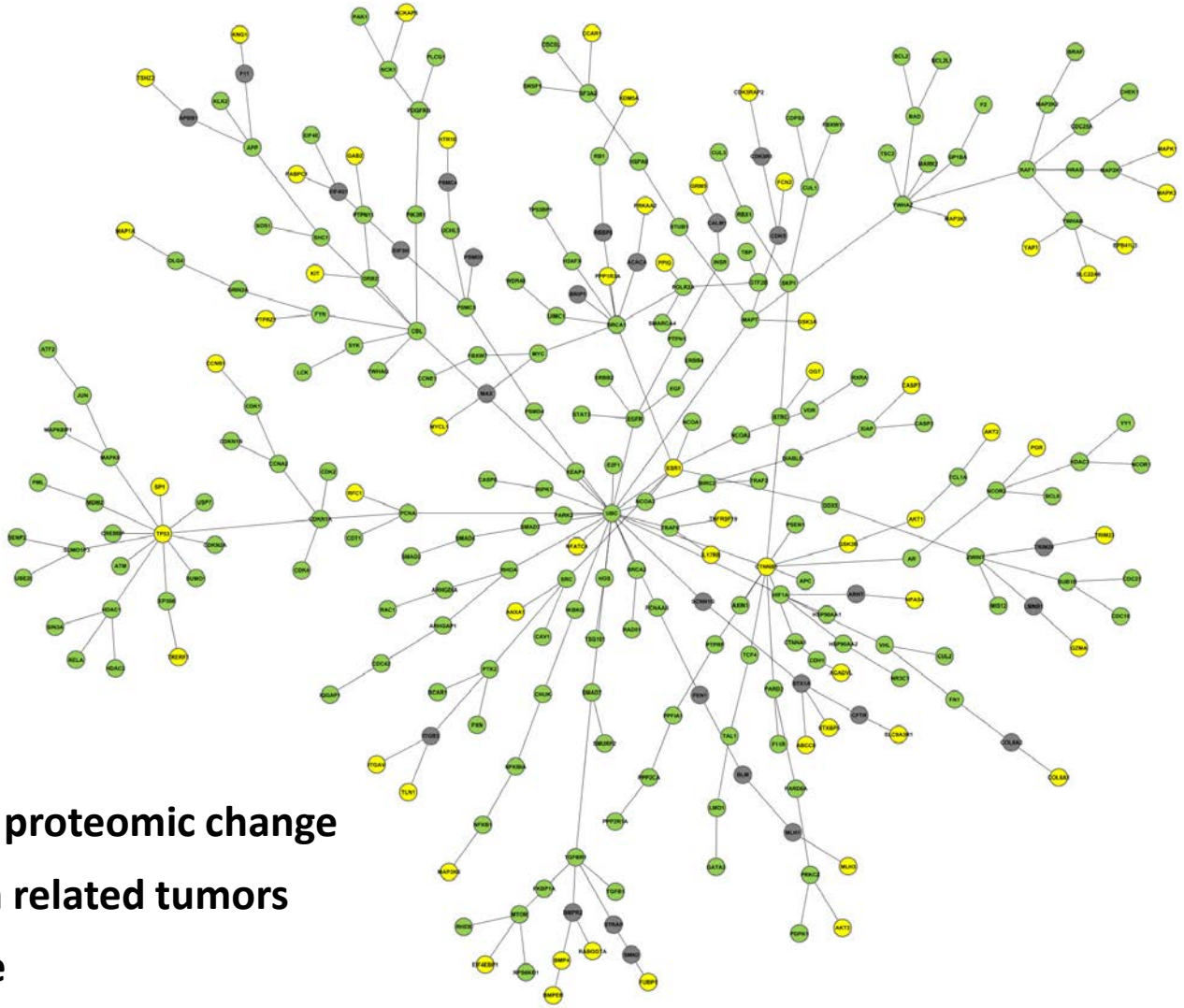
$$o(\mathcal{F}) = \sum_{i=1}^{N} o\left(F^i\right) + \lambda \sum_{i=1}^{N} \sum_{v \notin V_{F^i}} \phi(\alpha, v, p^i(v), \mathcal{F} \backslash F^i)$$

Original Steiner tree objective

Artificial prizes to share information

Explaining individuals' data vs. similarity tradeoff

Breast cancer tumor TCGA-AN-A0AR

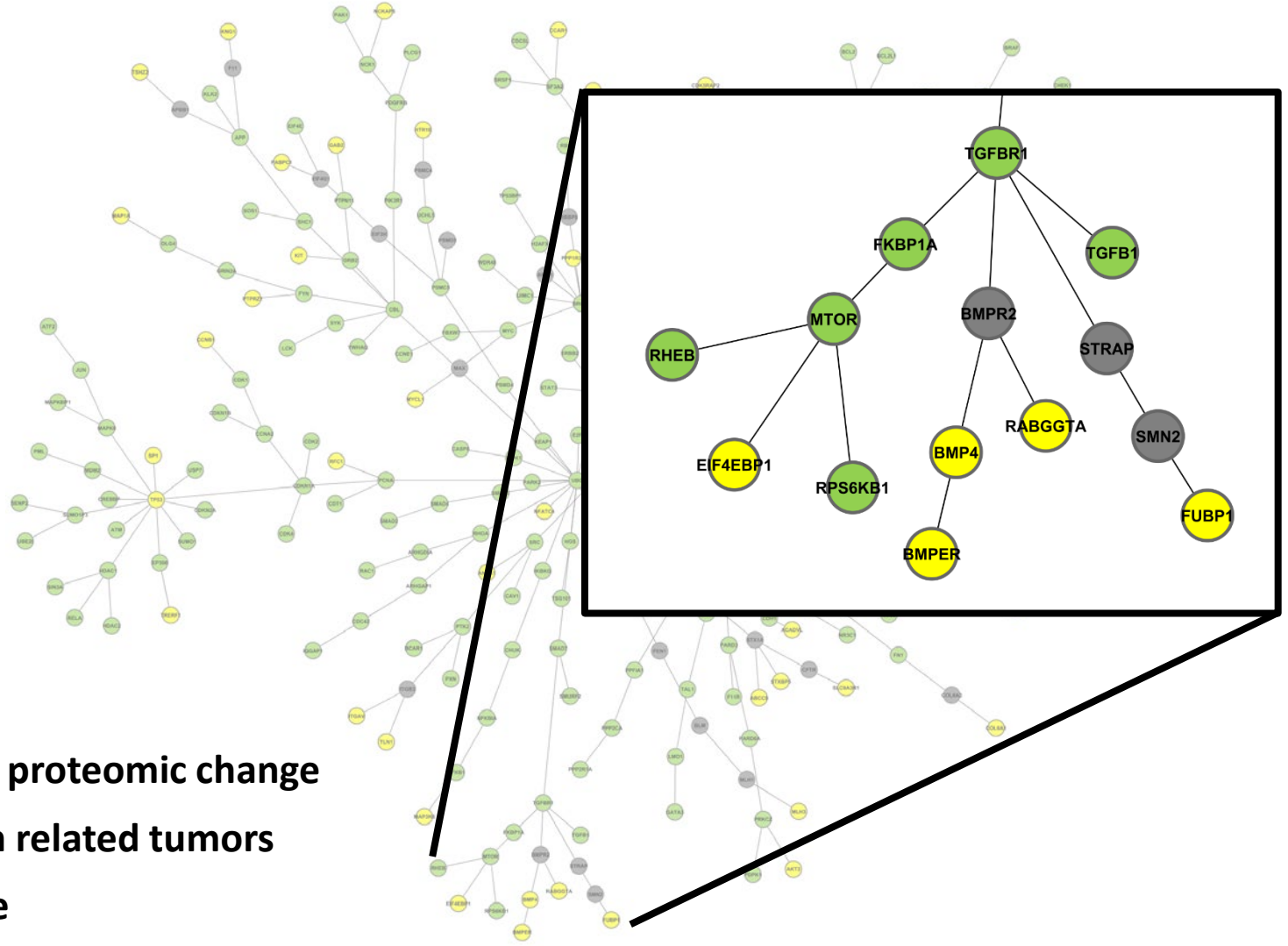


Mutation or proteomic change

Important in related tumors

Steiner node

Unique pathways with common core



- **Mutation or proteomic change**
- **Important in related tumors**
- **Steiner node**

Appendix

Alternative pathway identification algorithms

- Steiner tree/forest (related to PCST)
  - Prize-collecting Steiner forest (PCSF)
  - Belief propagation approximation (msgsteiner)
- k-shortest paths
  - Ruths2007
  - Shih2012
- Integer programs
  - Signaling-regulatory Pathway INferencE (SPINE)
  - Chasman2014

Alternative pathway identification algorithms continued

- Path-based objectives
  - Physical Network Models (PNM)
  - Maximum Edge Orientation (MEO)
  - Signaling and Dynamic Regulatory Events Miner (SDREM)
- Maximum flow
  - ResponseNet
- Hybrid approaches
  - PathLinker: random walk + shortest paths
  - ANAT: shortest path + Steiner tree

Recent developments in pathway discovery

- Multi-task learning: jointly model several related biological conditions
  - ResponseNet extension: SAMNet
  - Steiner forest extension: Multi-PCSF
  - SDREM extension: MT-SDREM
- Temporal data
  - ResponseNet extension: TimeXNet
  - Pathway synthesis