# *ALGEBRAIC AND TOPOLOGICAL INDICES OF MOLECULAR PATHWAY NETWORKS IN HUMAN CANCERS*

*JACK A. TUSZYNSKI*

*Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada*
*Department of Oncology, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada*

# collaborators

- **DYLAN BREITKREUTZ**

  **Department of Oncology, University of Alberta,**
- **EDWARD A. RIETMAN, GIANNOULA KLEMENT**

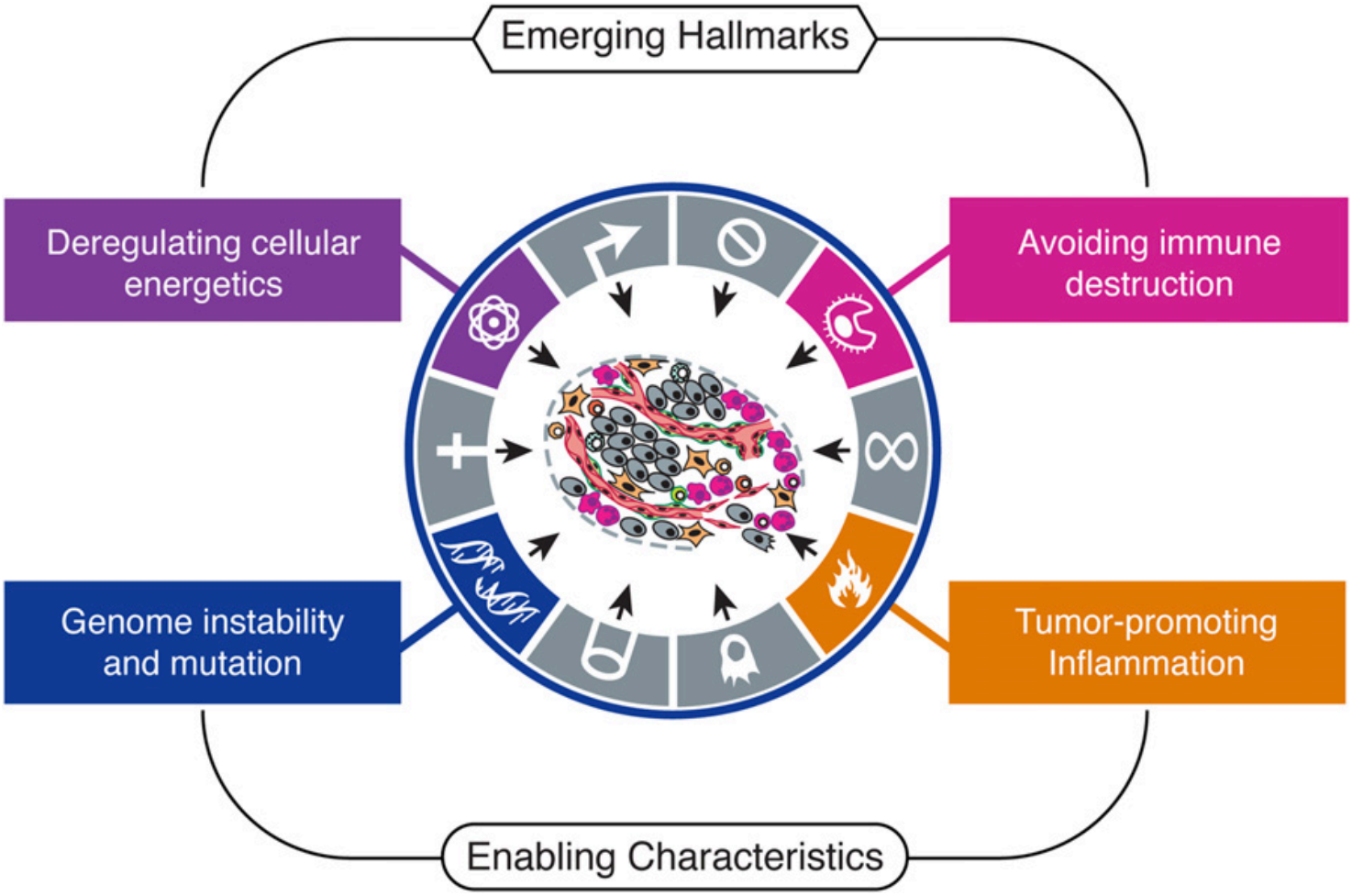
  **Center of Cancer Systems Biology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, MA, USA**
- **PETER HINOW**

  **Department of Mathematical Sciences, University of Wisconsin – Milwaukee, Milwaukee, WI, USA**
- **NICOLAS CARELS, TATIANA TILLI**
- **Fundacao Oswaldo Cruz, Rio de Janeiro, Brazil**

# References


- S. Benzekry, J. A Tuszynski, E. A Rietman and G. L Klement, Design Principles for Cancer Therapy guided by changes in complexity of Protein-Protein Interaction Networks, **Biology Direct** 10:32 (2015), DOI 10.1186/s13062-015-0058-5.
- P. Hinow, E.A. Rietman, S.I. Omar and J.A. Tuszynski, Algebraic and Topological Indices of Molecular Pathway Networks in Human Cancers, **Mathematical Biosciences and Engineering** vol. 12, No. 6, December pp. 1289-1302 (2015)
- D.Breitkreutz, L. Hlatky, E. A. Rietman, and J. A. Tuszynski, Molecular Signaling Network Complexity Is Correlated with Cancer Patient Survivability, **Proceedings of the National Academy of Sciences of the USA,** Jun 5;109(23):9209-12. Epub 2012 (2012),
- T. Tilli, N. Carels and J.A. Tuszynski, Optimization of combination chemotherapy based on the calculation of network entropy for protein-protein interactions in breast cancer cell lines, **European Physical Journal Nonlinear Biomedical Physics 3**:6
- N. Carels, T. Tilli and J.A. Tuszynski, A computational strategy to select optimized protein targets for drug development toward the control of cancer diseases, **PLoS One**, January 27, 2015, DOI: 10.1371/journal.pone.0115054,

# The Cancer Problem




Hanahan and Weinberg 2011, Cell 144:646-674.

# Cancer Cell Network

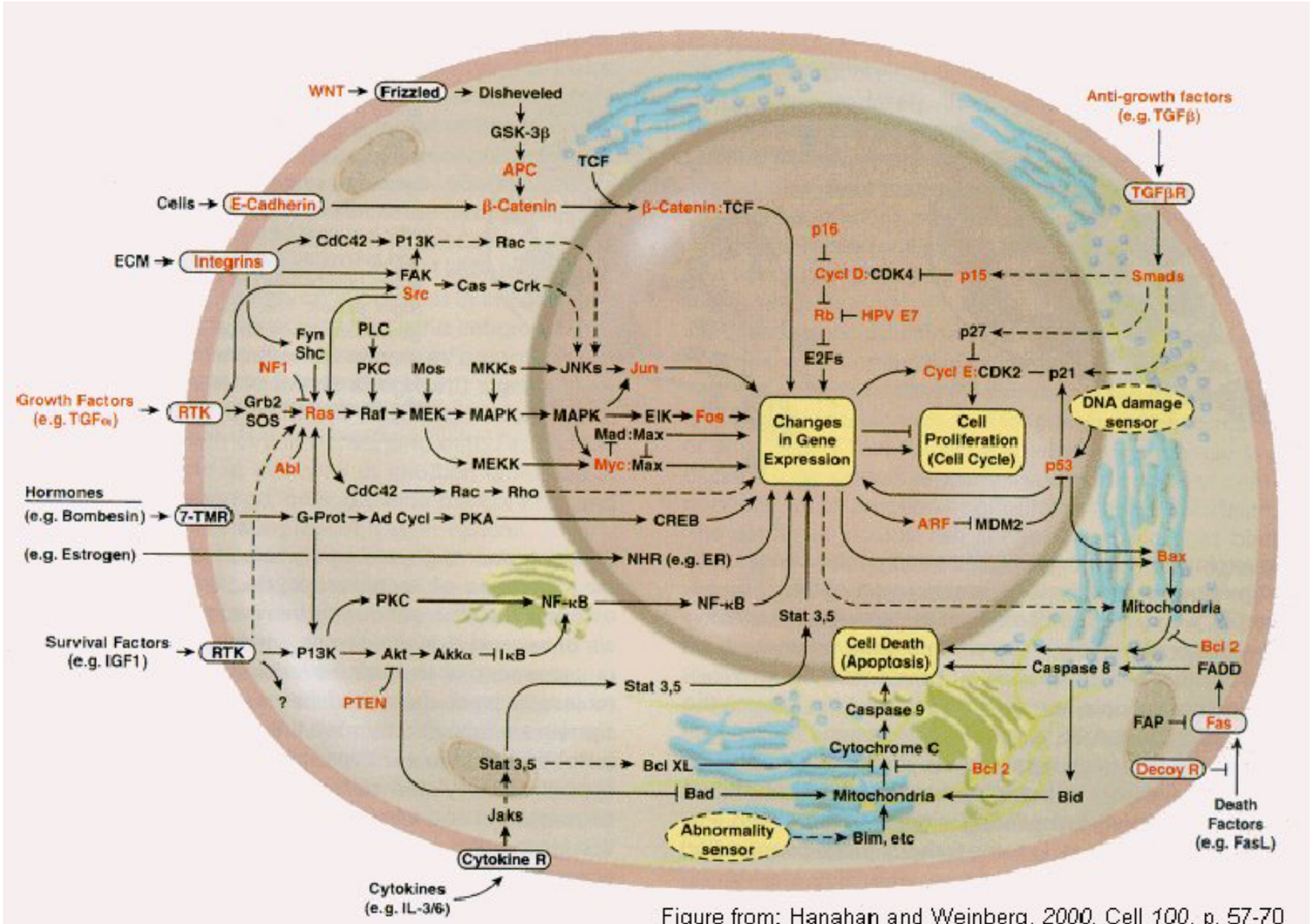
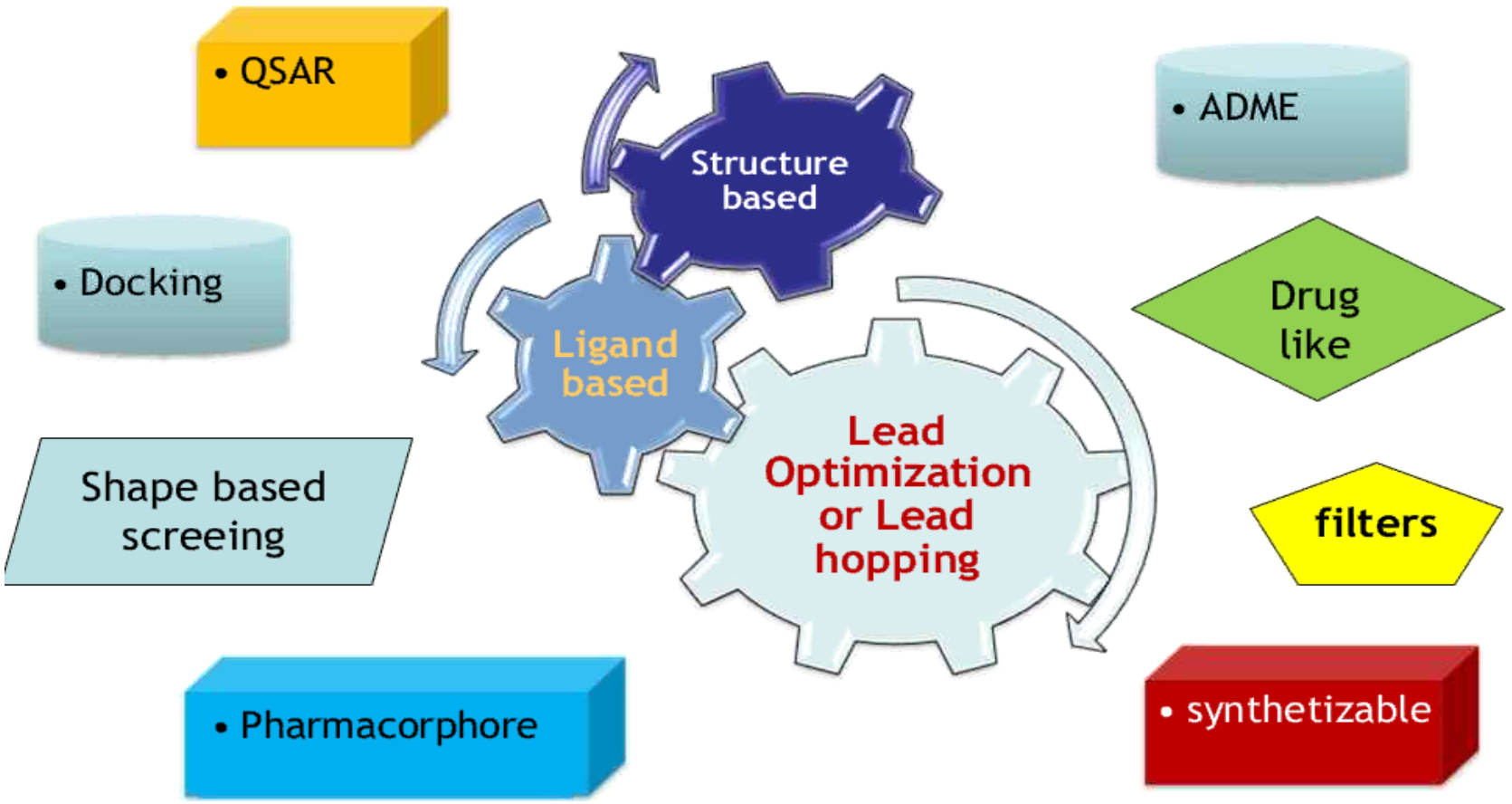



Figure from: Hanahan and Weinberg, *2000*. Cell *100*, p. 57-70

# We can find inhibitors for almost all targets




But how do we select targets?

# Introduction

- Overall Goal: To look at cancer and chemotherapy in a different way and to ultimately improve treatment.
- Investigate a quantitative measure of the robustness of cancer signaling pathways.
- Cell biology normally focuses on individual components and processes.
- Network biology focuses on the interaction of all components of a biological system.
- Chemotherapy traditionally focuses on single targets.
- Cellular processes are more complex than this.
- We want to find a way to evaluate the effectiveness of chemotherapeutics on a network level.

# Information available

- Data bases of drugs approved and investigational, their mode of action,  targets, applications in cancer
- Biochemical pathways (identify where drugs inhibit them)
- List of “druggable targets”
- Methods of modeling networks and pathways

# Key Databases for these Studies

Database of pathway networks for cancer (and other stuff)
KEGG (Kyoto Encyclopedia of Genes and Genomes)
http://www.genome.jp/kegg/

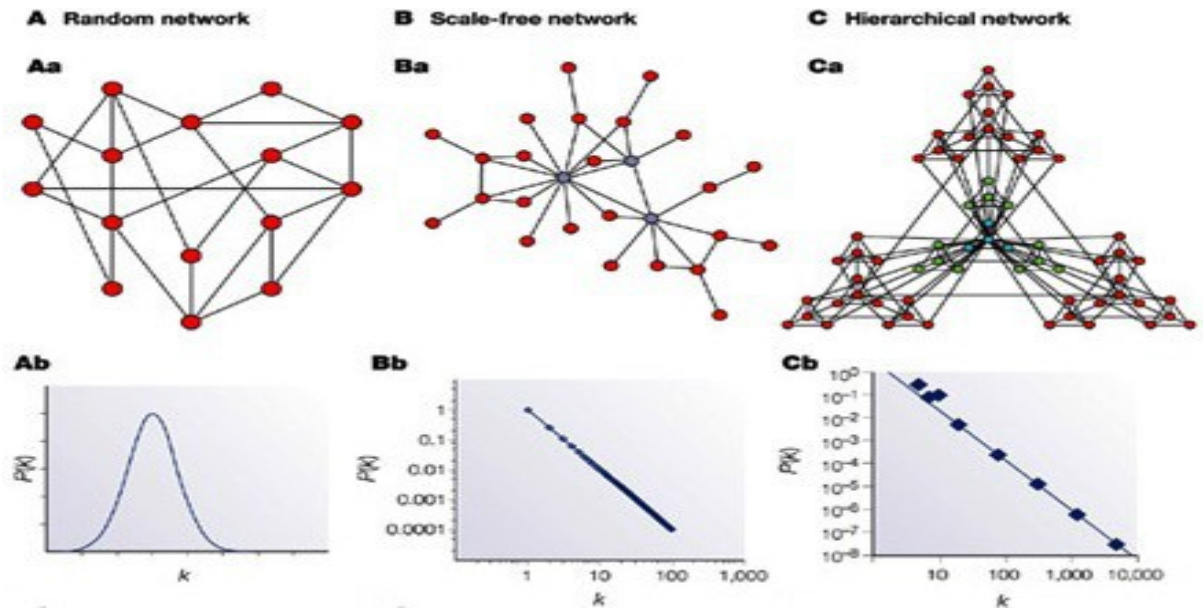
Database of statistical information on survival.
SEER (Surveillance Epidemiology and End Results)
http://seer.cancer.gov/

Database of protein-protein interaction networks
BioGrid
http://www.biogrid.org

Database of cancer data
TCGA (The Cancer Genome Atlas)
http://cancergenome.nih.gov/

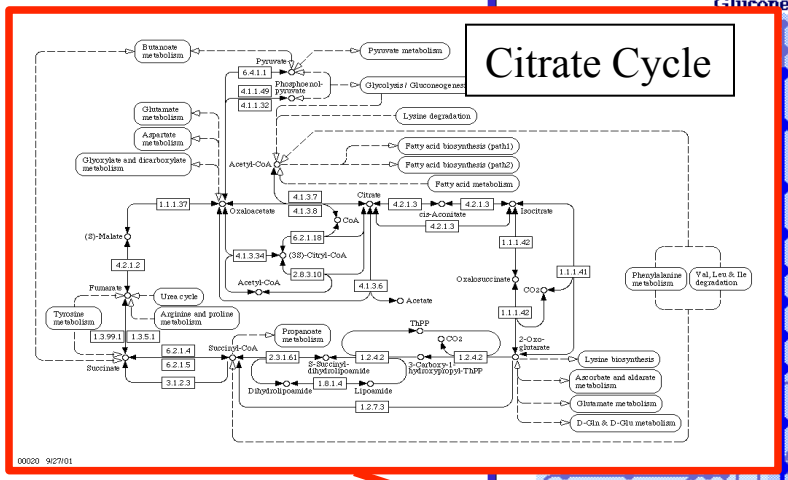
Database of drugs, both approved and investigational
Drugbank.ca

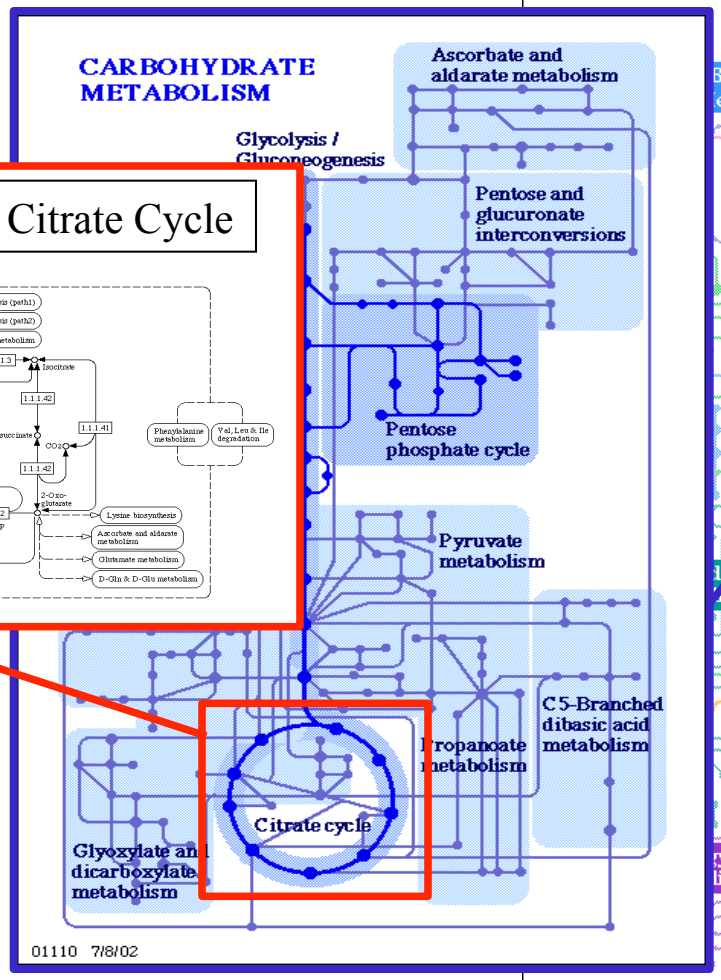
# Organization vs. Randomness

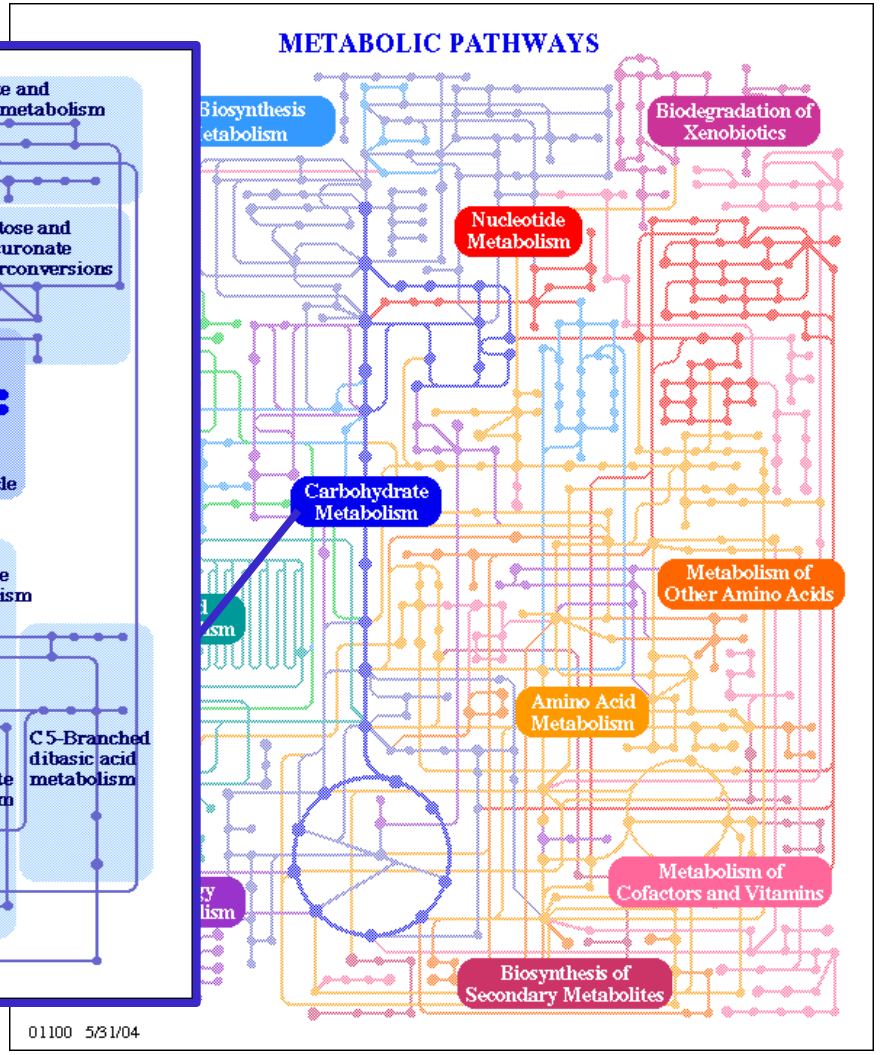



Networks in nature and technology differ from “random networks” in many aspects.

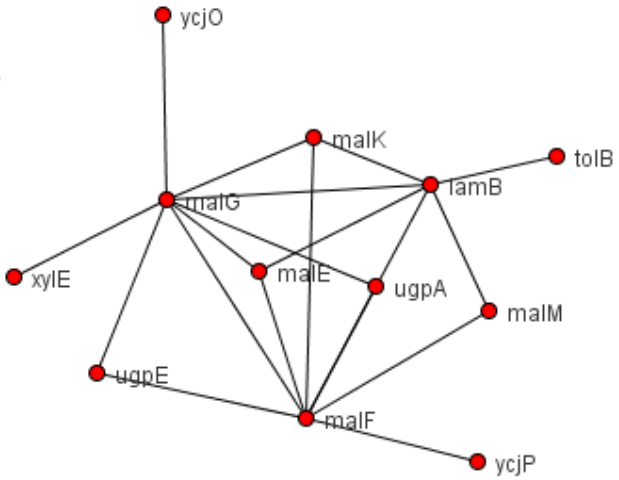
# Biological pathways are complex networks



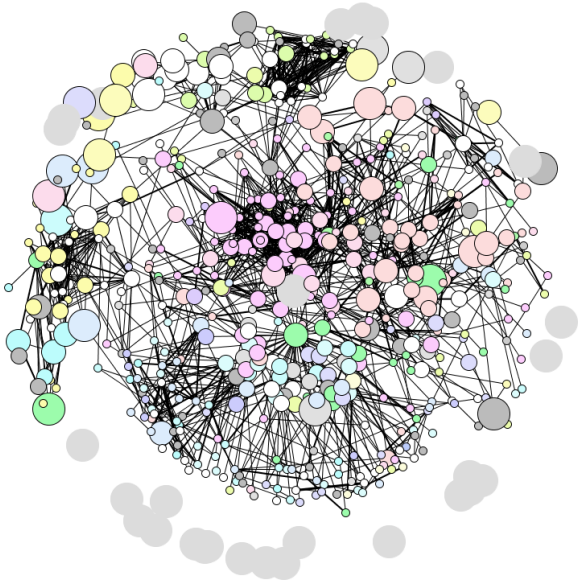
# Networks may be analyzed using graph theory



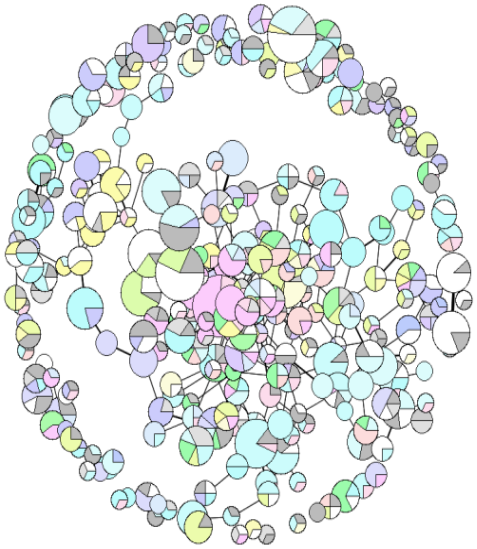

Nodes represent e.g. proteins, genes or substrates

Links between nodes represent interactions e.g. physical, genetic, biochemical

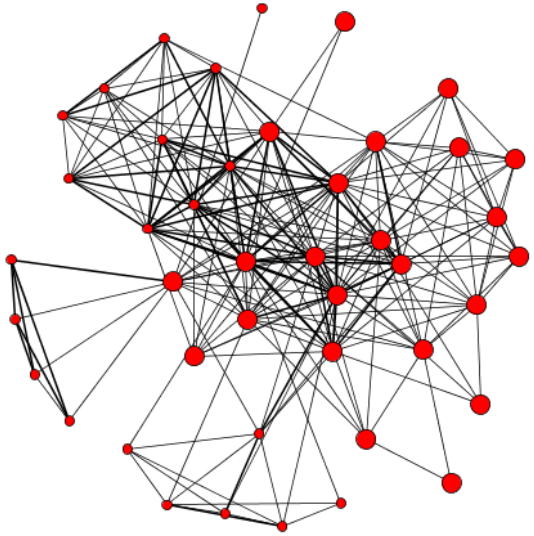
Analyses can be performed at different levels



Global



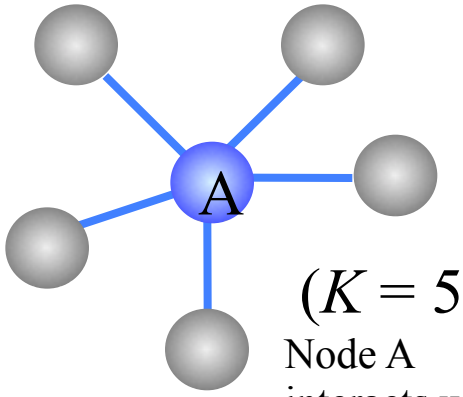
Complexes / cliques / pathways



Local

# Topological properties of networks

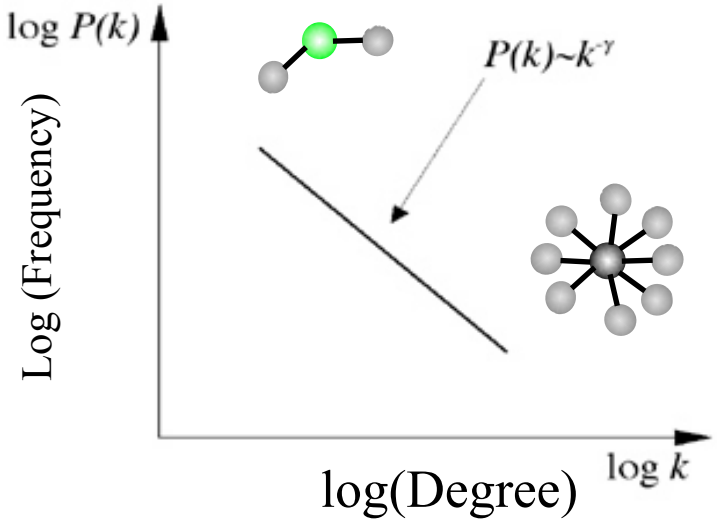
Global analyses of these properties over an entire network provide insights into its organization




($K = 5$)

Node A interacts with five other nodes

Many networks display small world / *scale free* behaviour (many nodes with few connections; few nodes with many connections)



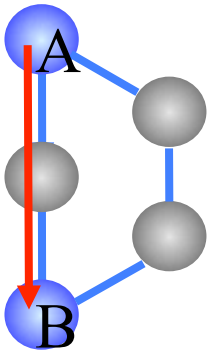

One of the more commonly used is

***Node Degree or rank order***

Scale free networks are thought to be more resistant to disruption

# Topological properties of networks

## Shortest path length

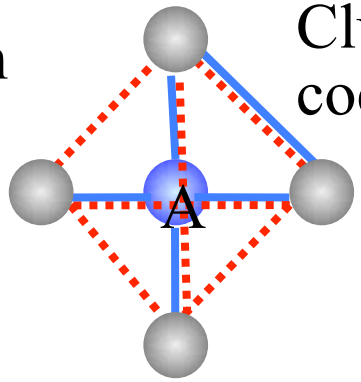



$(L = 2)$

The shortest path between A and B is via 2 links

Mean path length offers a measure of a networks overall navigability

## Cluster coefficient



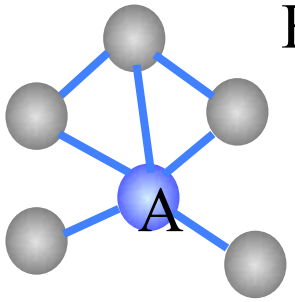

$(C = 1/6)$

Of the six possible connections between the neighbours of A, only one is actually made:

n!/2! (n-2)!=6; r=1

Average cluster coefficient characterizes the overall tendency of the network to form clusters

## Betweenness

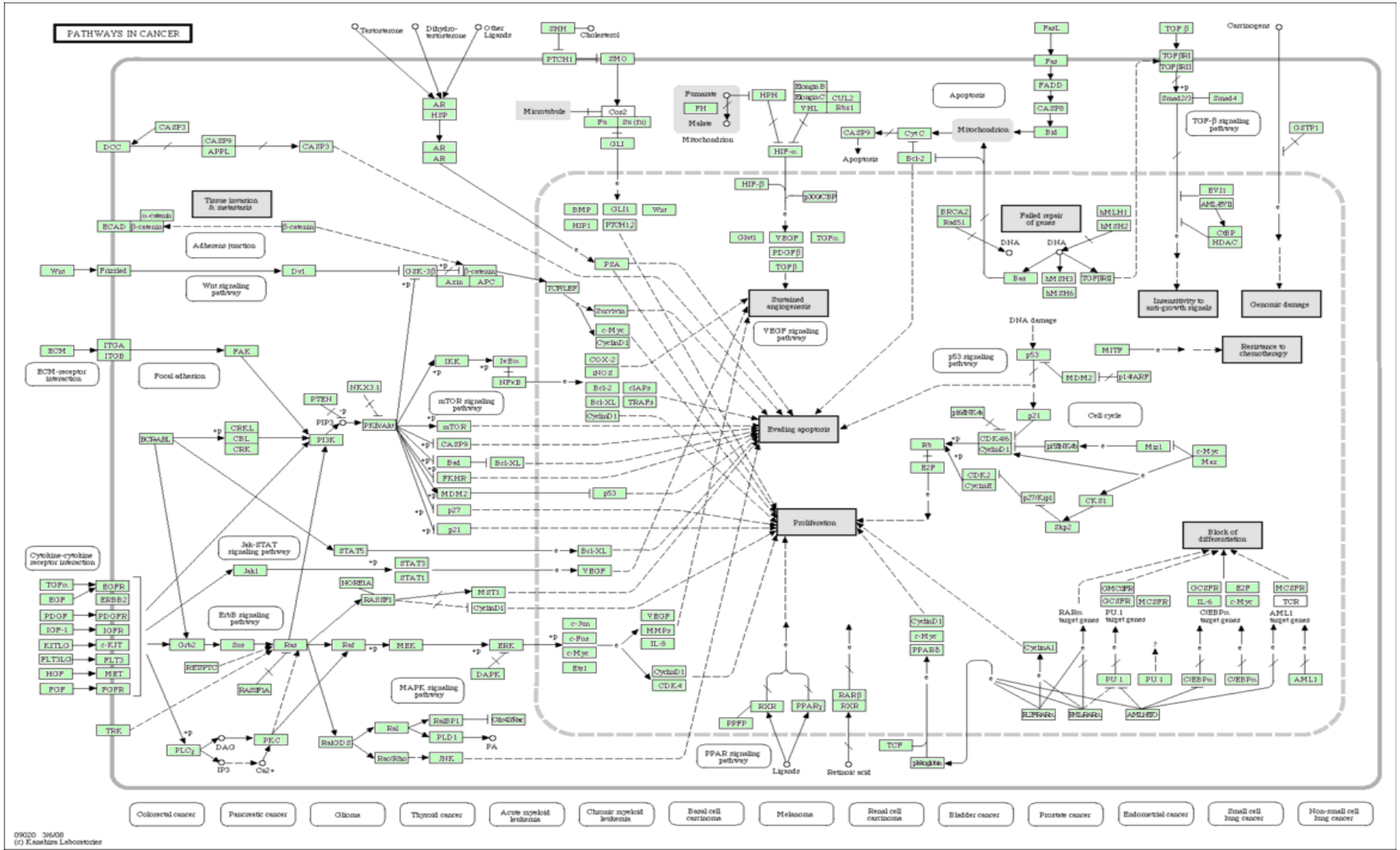



*(B = 13/15)*

13 out 15 shortest paths in the network go through node A

Nodes with high betweenness are 'central' to the network

# General cancer pathway



# Molecular Pathway Networks

A *network* is an undirected graph $G = (V, E)$ with vertex and edge sets $V$ and $E$, respectively. The vertices are proteins and two vertices are connected by an edge if there is a known interaction of the two partners, either by direct binding or by enzymatic catalysis.

We obtain the cancer pathways from KEGG

http://www.genome.jp/kegg/

with the open source software packages `KEGGgraph` and `cytoscape`.



`http://bioconductor.org` `http://cytoscape.org`

# Chemotherapeutic agents' interactions with targets

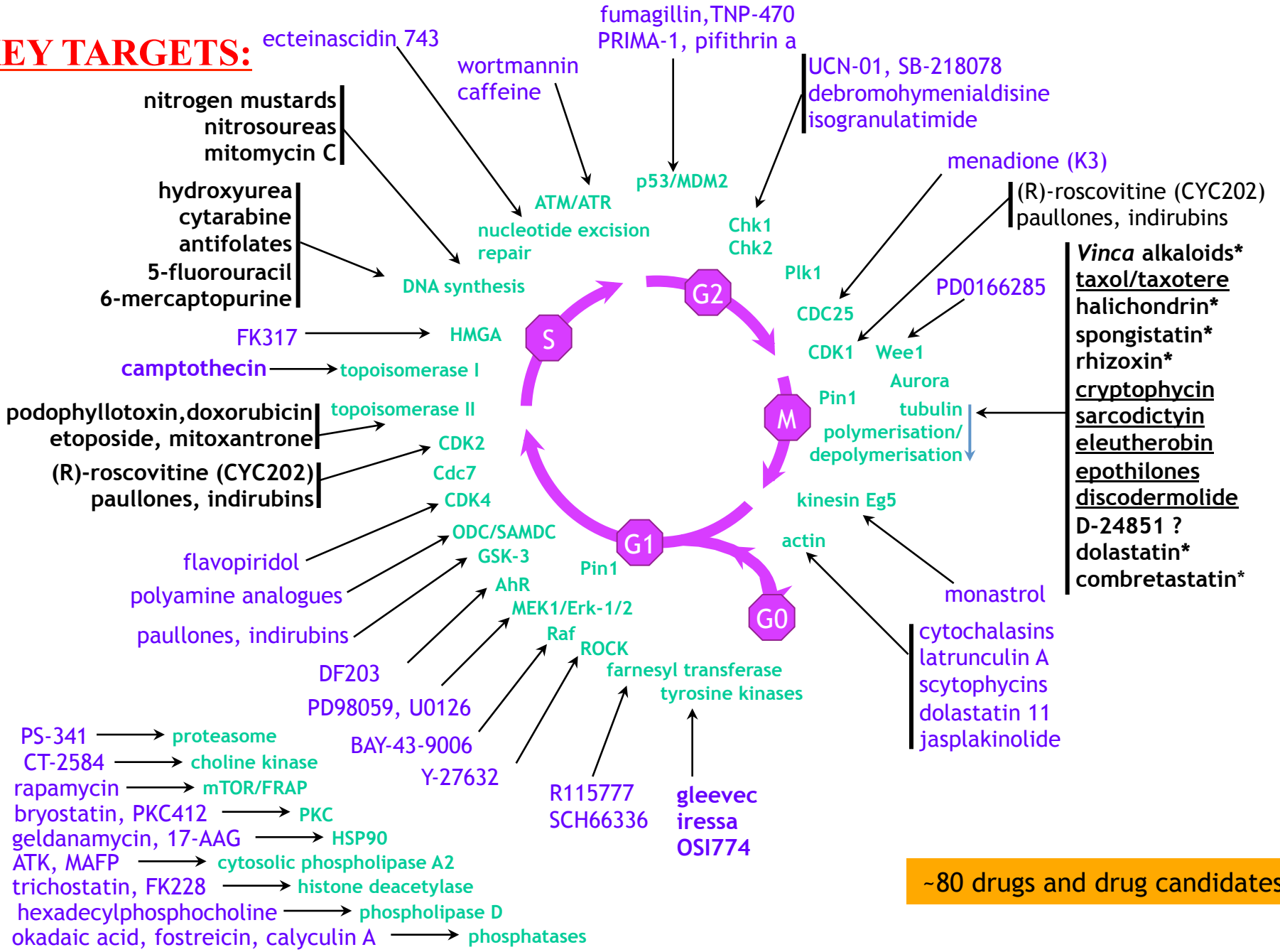
*Target : a molecule whose interaction with an anticancer agent will induce a cytotoxic effect*

Targets are key bio- molecules involved or required for cell mitosis and/or survival

Conventional chemotherapy acts on dividing cells only, but does not distinguish normal and abnormal dividing cells

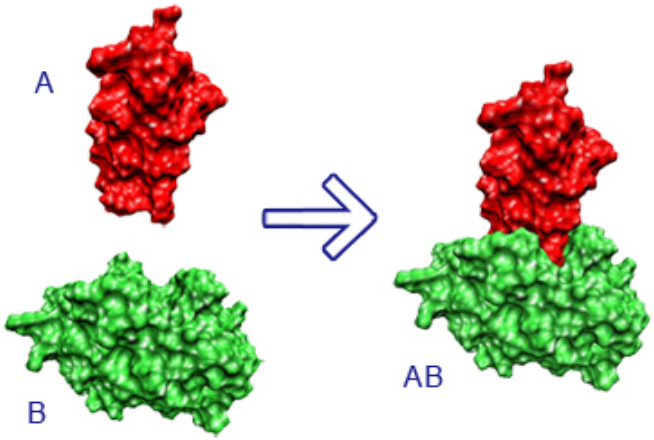
Targeted agents are designed to act on targets which are specific for tumor cells

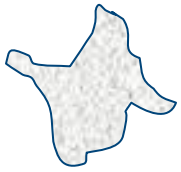
# KEY TARGETS:



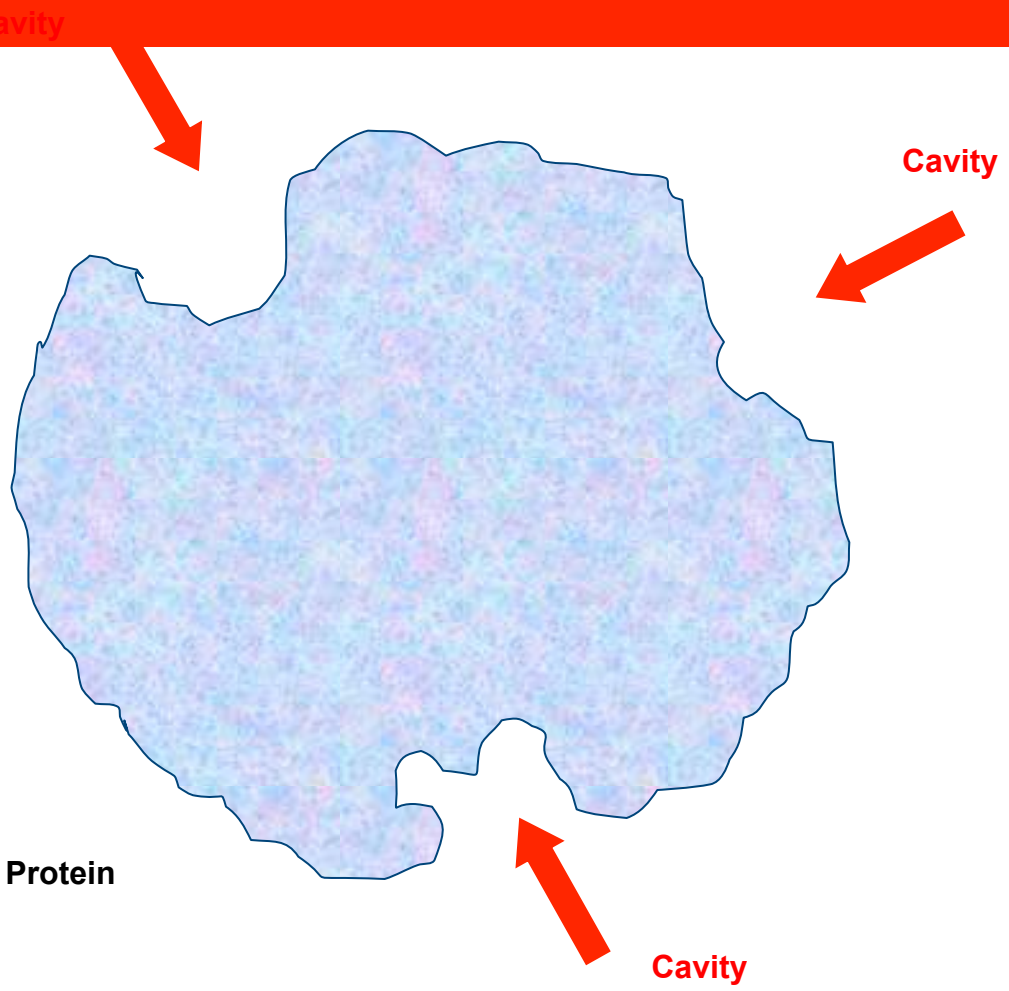

*Source: Cell cycle laboratory, L. Meijer, Roscoff, France*

# Drug Binding: Inhibition of Protein-Protein Interactions



Drug / Ligand



# Robustness of Biological Networks

- We want to know how resilient these pathways are to chemotherapy.
- How does the inhibition of an interaction effect function of the entire network?
- Robustness is the measure of how well networks function under random perturbation.
- Network robustness can be quantified as entropy.

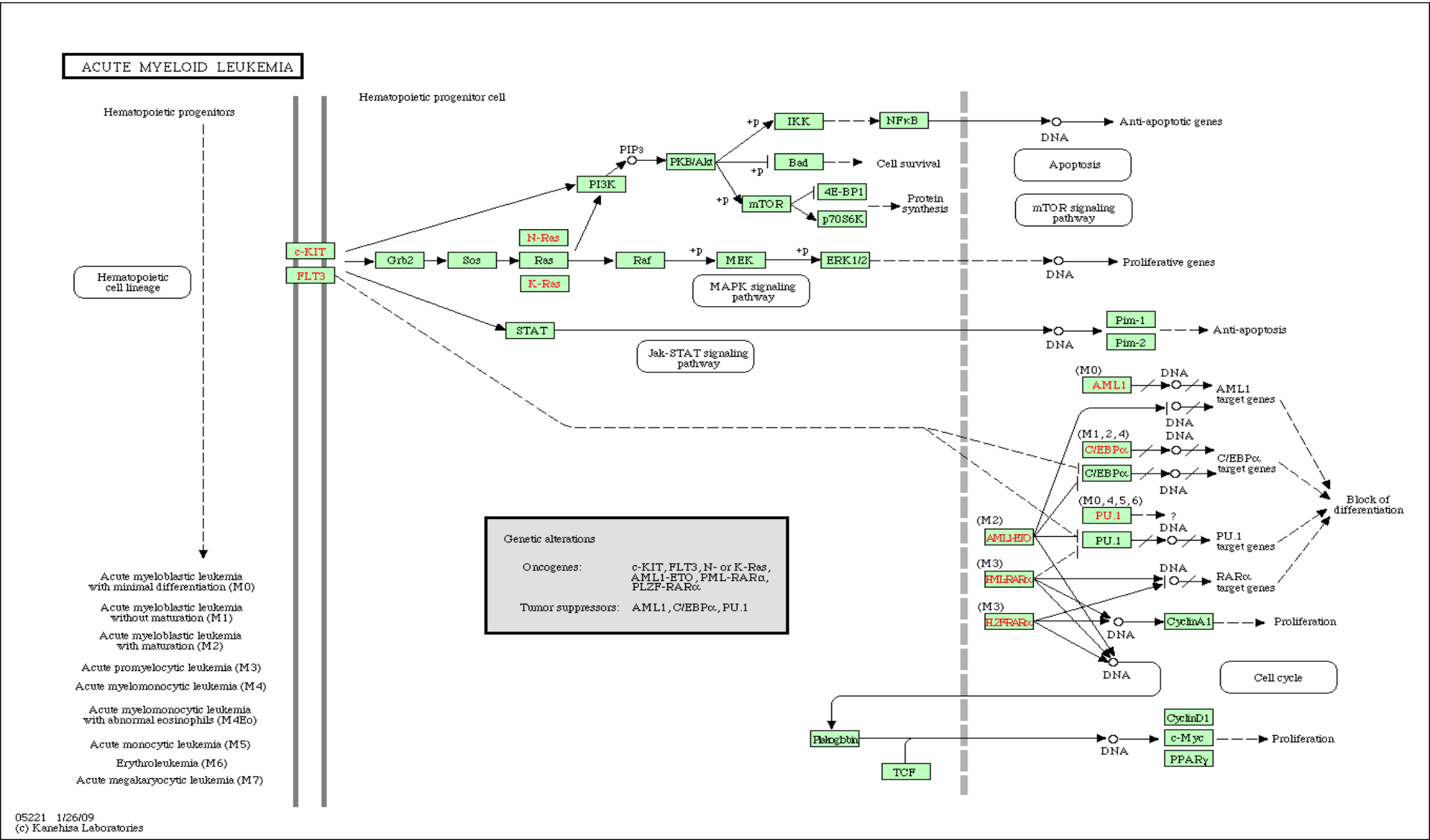
# Graph Theory and Entropy

- A graph is a collection of nodes and edges.
- In this case nodes represent proteins and genes while edges represent interactions between them.
- The degree of a node is a count of how many edges lead to or from it.
- Pathways were converted into graphs using R and KEGGgraph.
- The entropy of these graphs is then given by $H = -\Sigma\, p(k)*\ln(p(k))$
- where p(k) is the probability that a node has degree k.

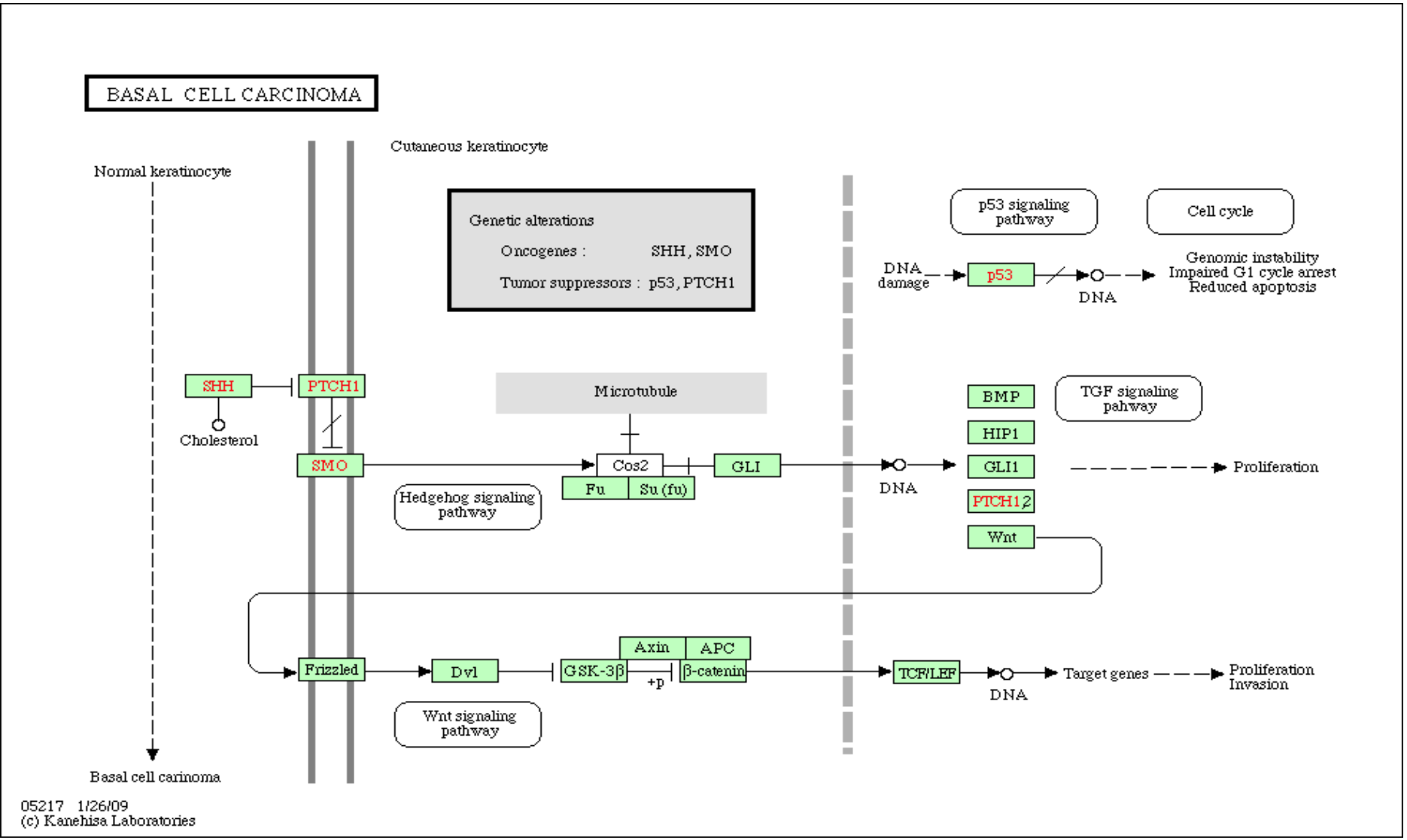
# Analysis of Pathways

- The next step was to calculate the entropy of each pathway.
- This was done using both R and Excel.
- After that,  to draw useful information from these entropies.
- We hypothesized that there should be a correlation between entropy and lethality.
- The most lethal cancers should be the most robust.

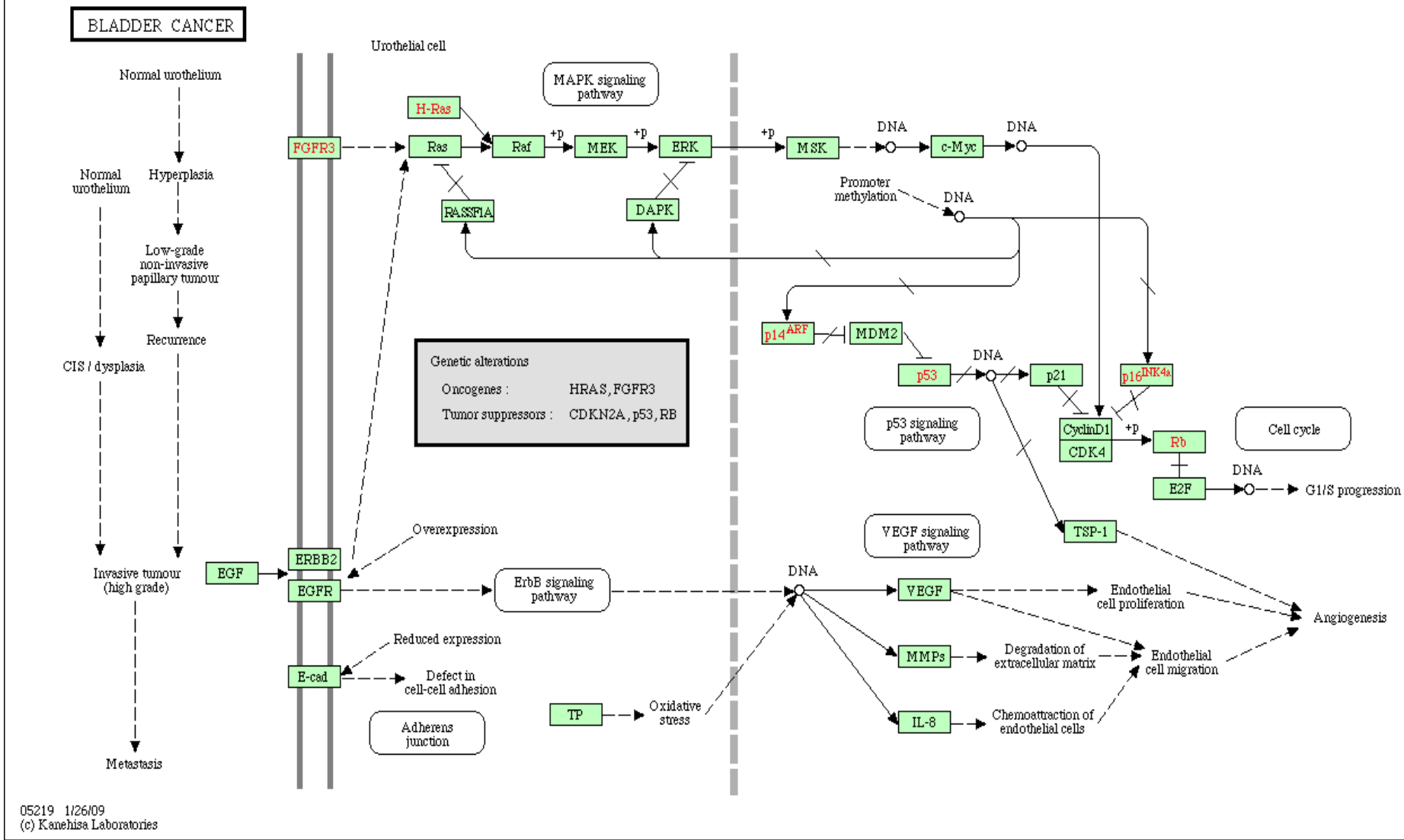
# Acute myeloid leukemia



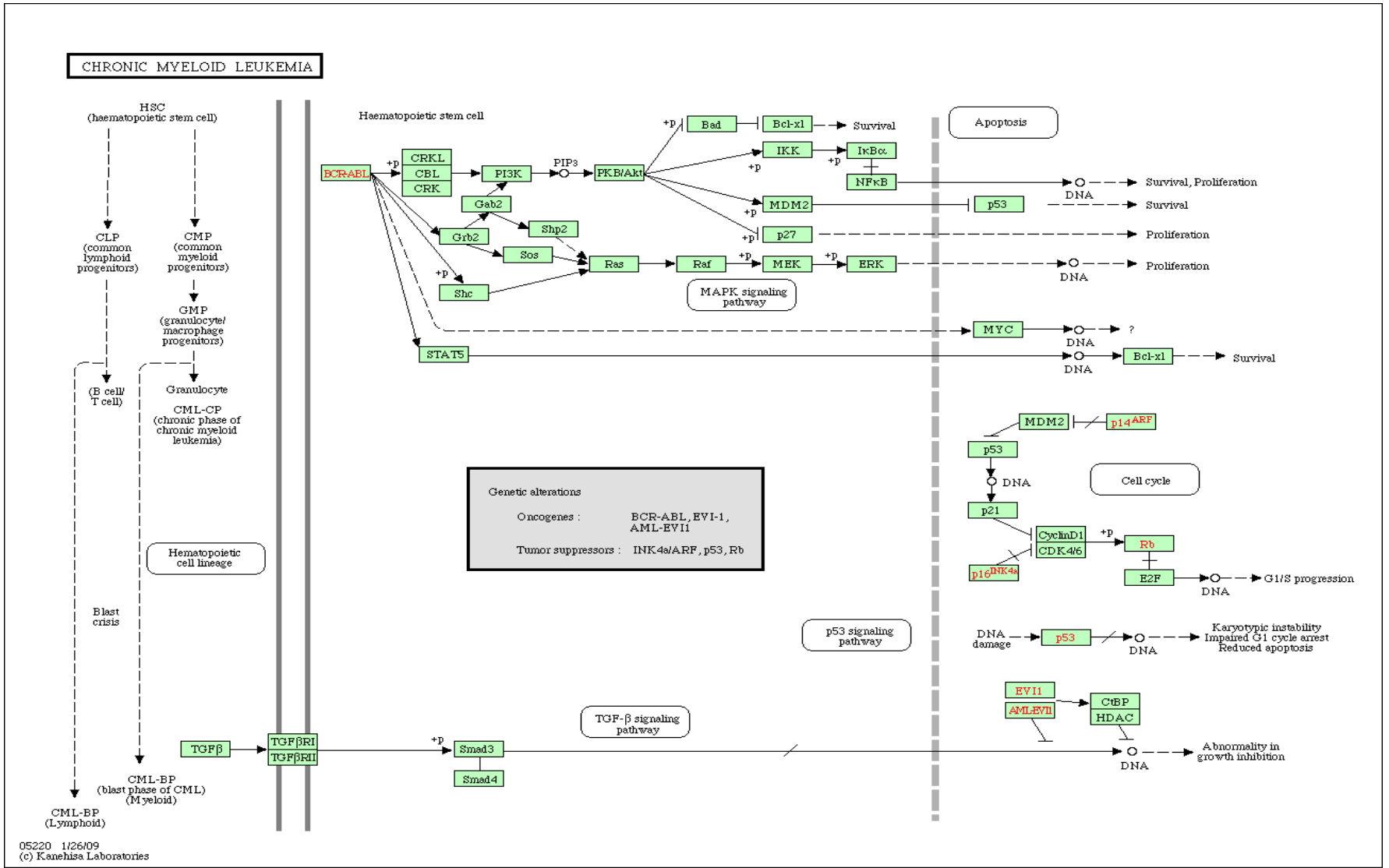
# Basal cell carcinoma



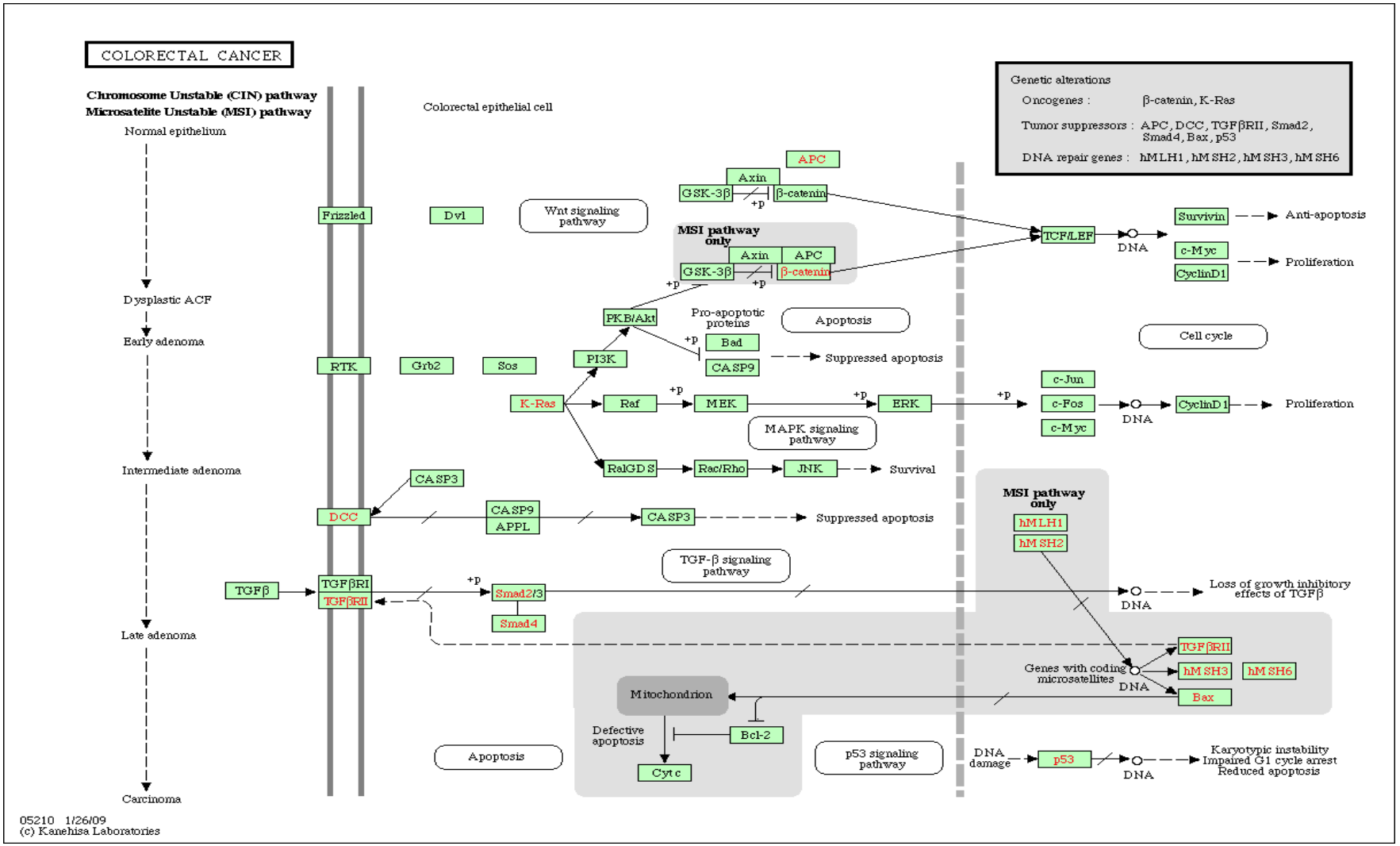
# Bladder cancer



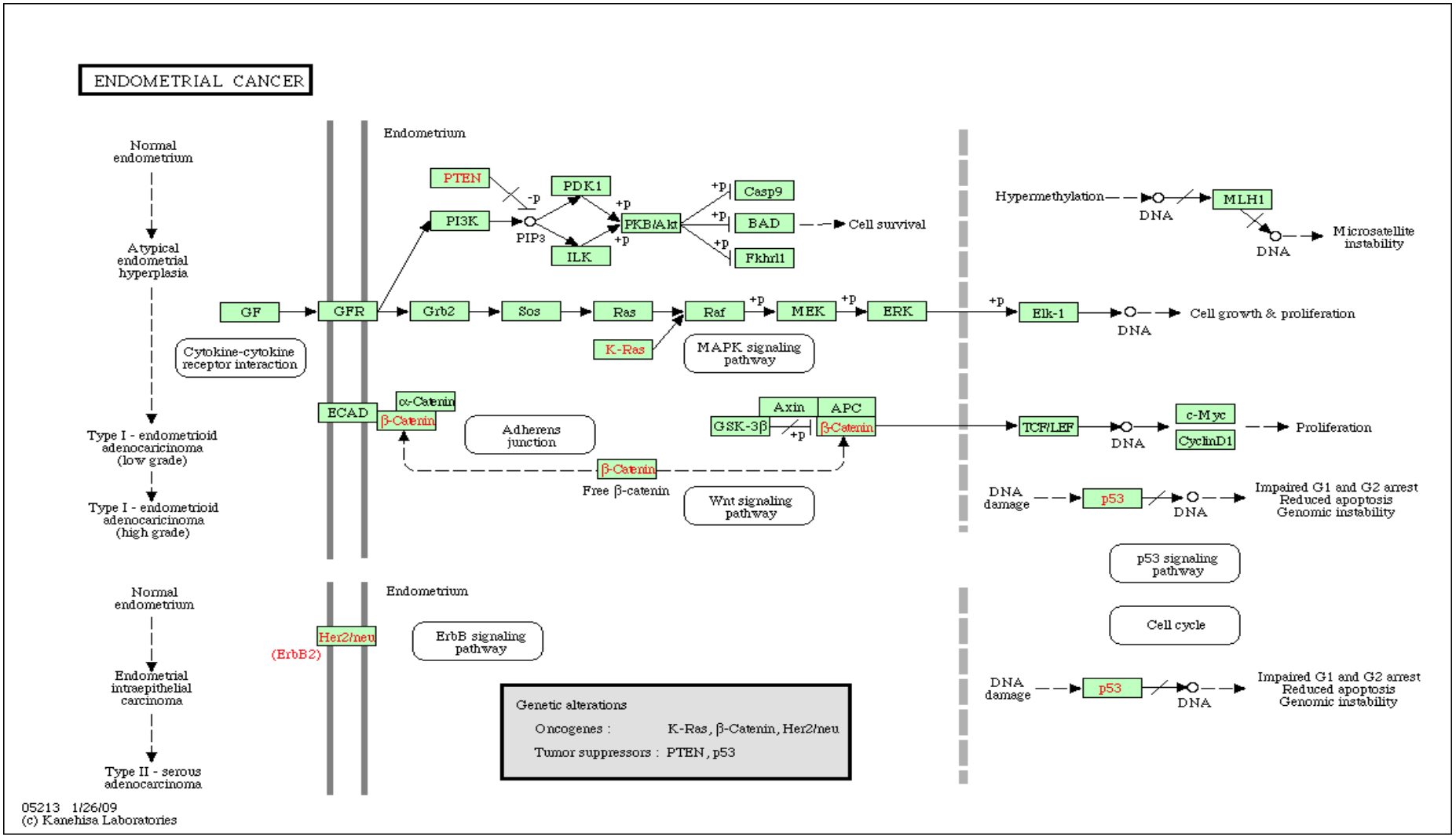
# Chronic myeloid leukemia



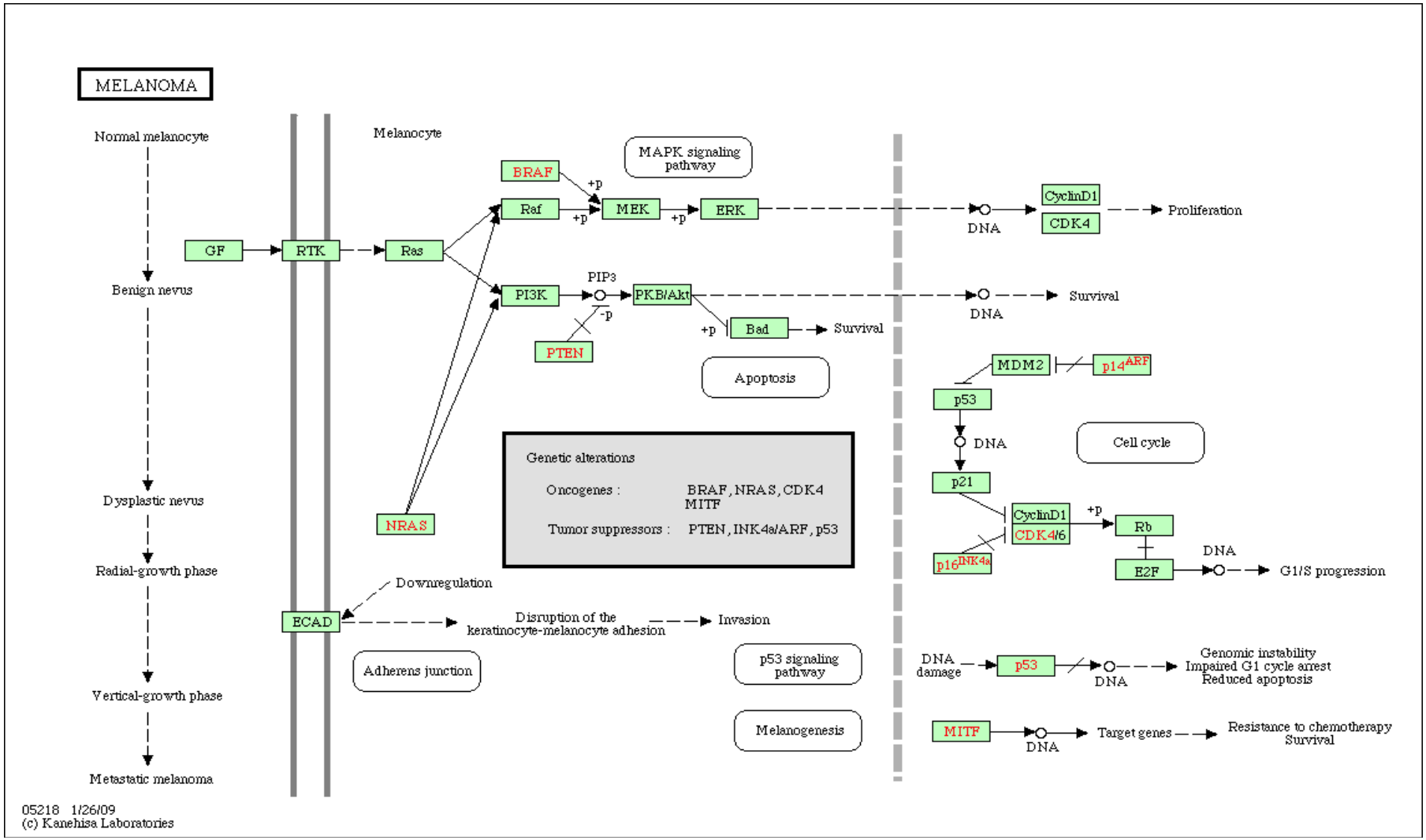
# Colorectal cancer



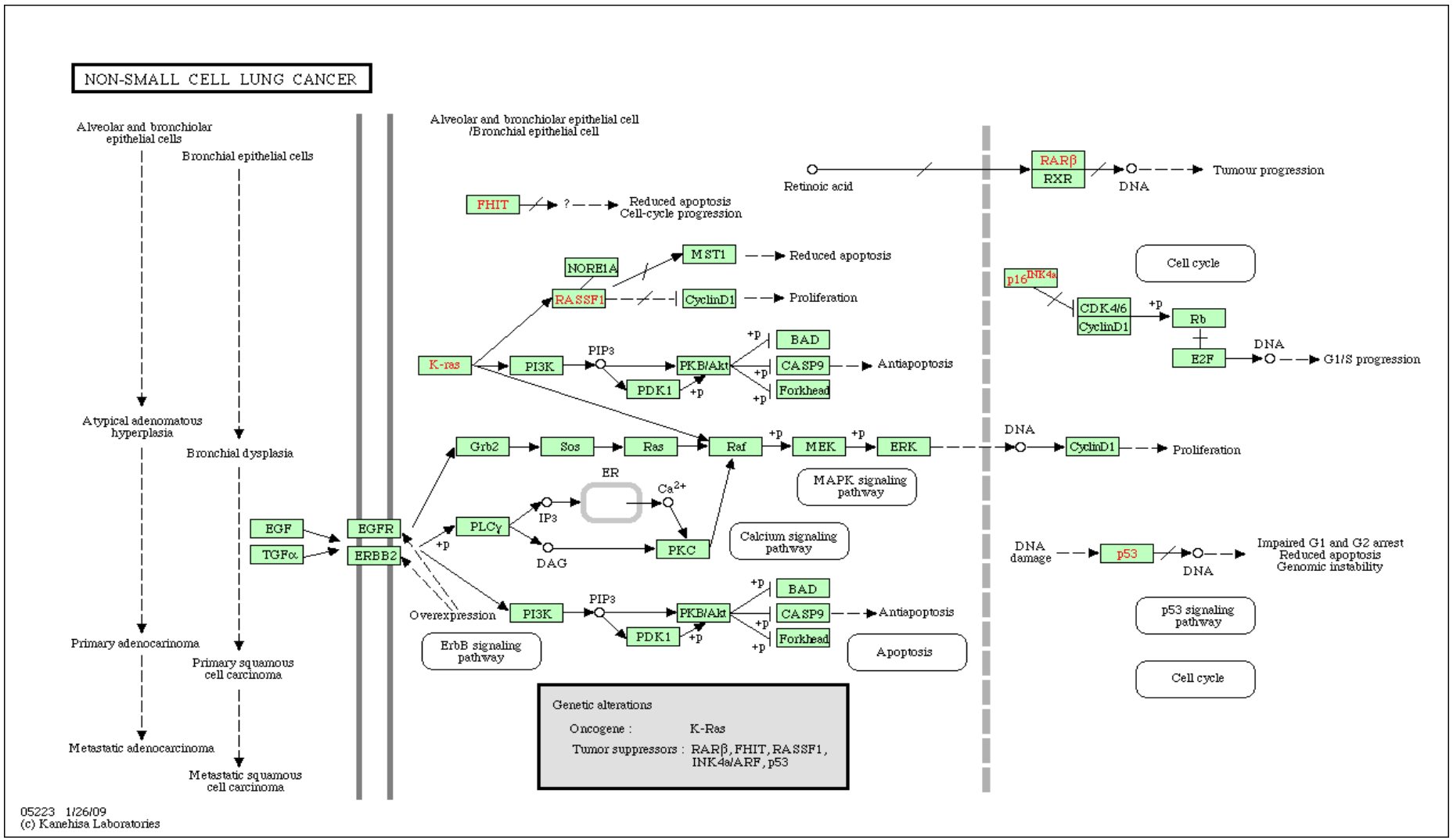
# Endometrial cancer



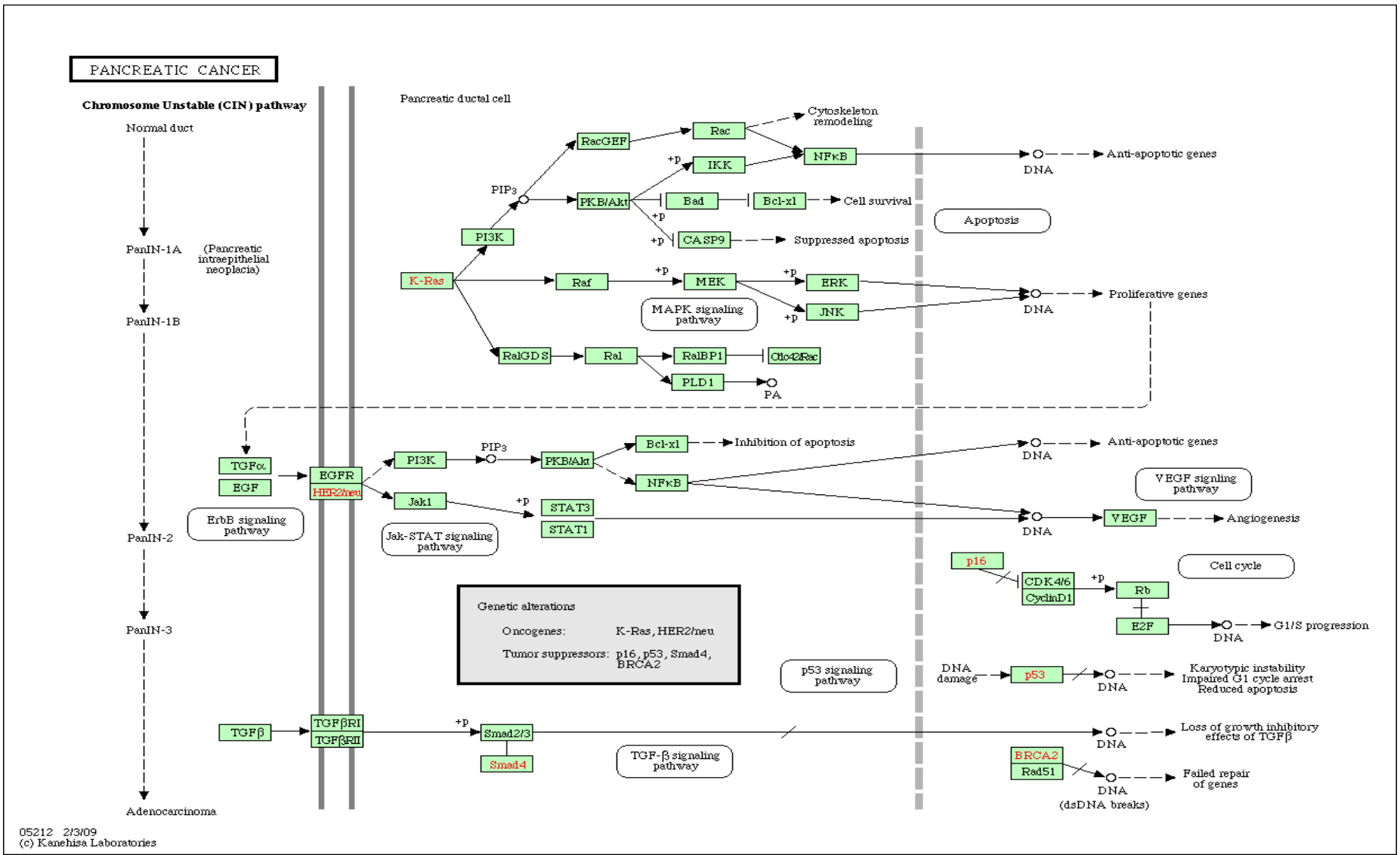
# melanoma


MELANOMA
Normal melanocyte
Benign nevus
Dysplastic nevus
Radial-growth phase
Vertical-growth phase
Metastatic melanoma
Melanocyte
GF
RTK
Ras
BRAF
Raf
MEK
ERK
MAPK signaling pathway
+p
+p
+p
DNA
CyclinD1
CDK4
Proliferation
PI3K
PIP3
PKB/Akt
-p
PTEN
+p
Bad
Survival
DNA
Survival
Apoptosis
NRAS
Genetic alterations
Oncogenes : BRAF, NRAS, CDK4 MITF
Tumor suppressors : PTEN, INK4a/ARF, p53
MDM2
p14ARF
p53
DNA
Cell cycle
p21
CyclinD1
CDK4/6
+p
Rb
p16INK4a
DNA
E2F
G1/S progression
Downregulation
ECAD
Disruption of the keratinocyte-melanocyte adhesion
Invasion
Adherens junction
p53 signaling pathway
Melanogenesis
DNA damage
p53
DNA
Genomic instability
Impaired G1 cycle arrest
Reduced apoptosis
MITF
DNA
Target genes
Resistance to chemotherapy
Survival
05218 1/26/09
(c) Kanehisa Laboratories

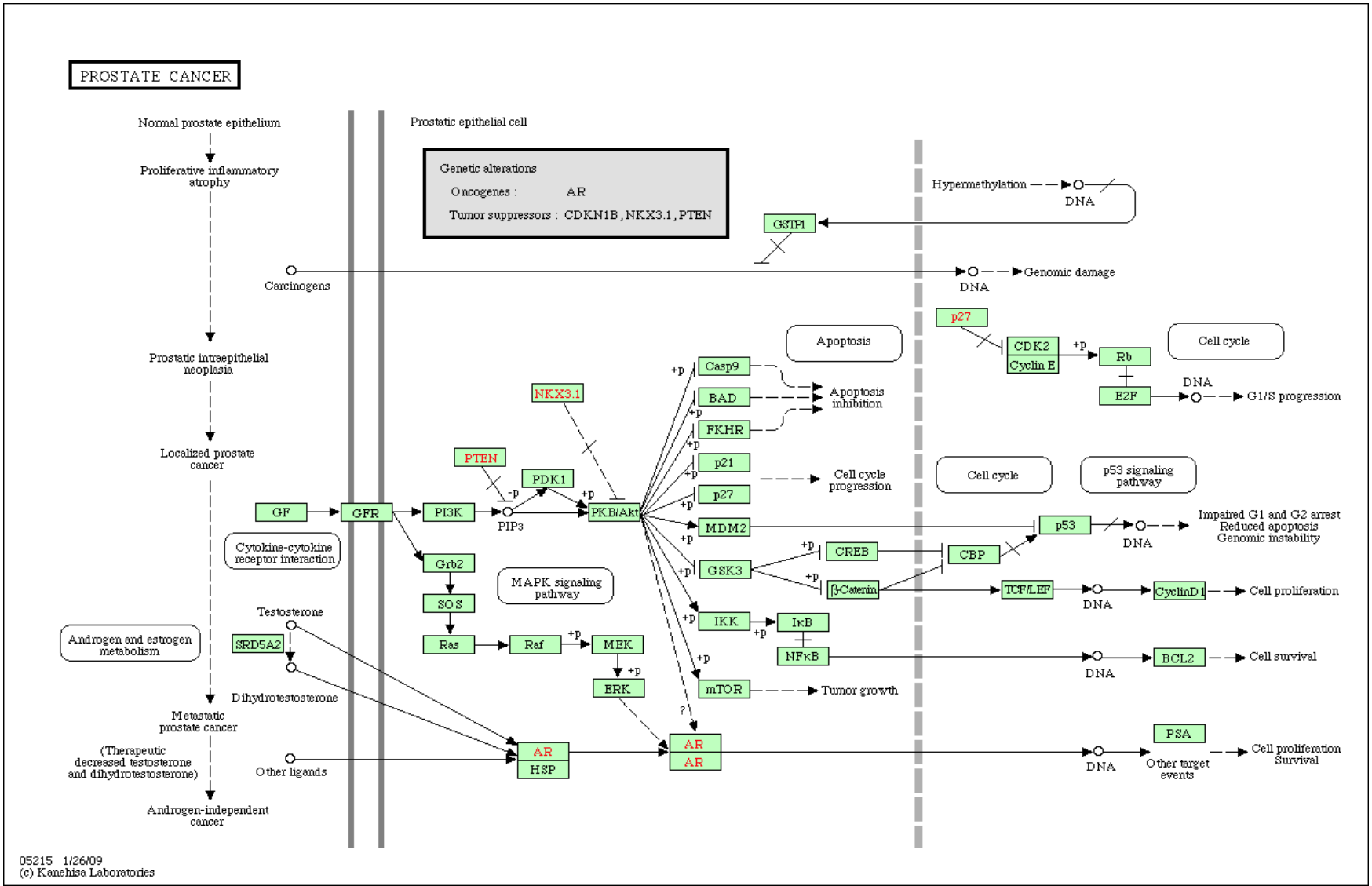
# Non-small cell lung cancer



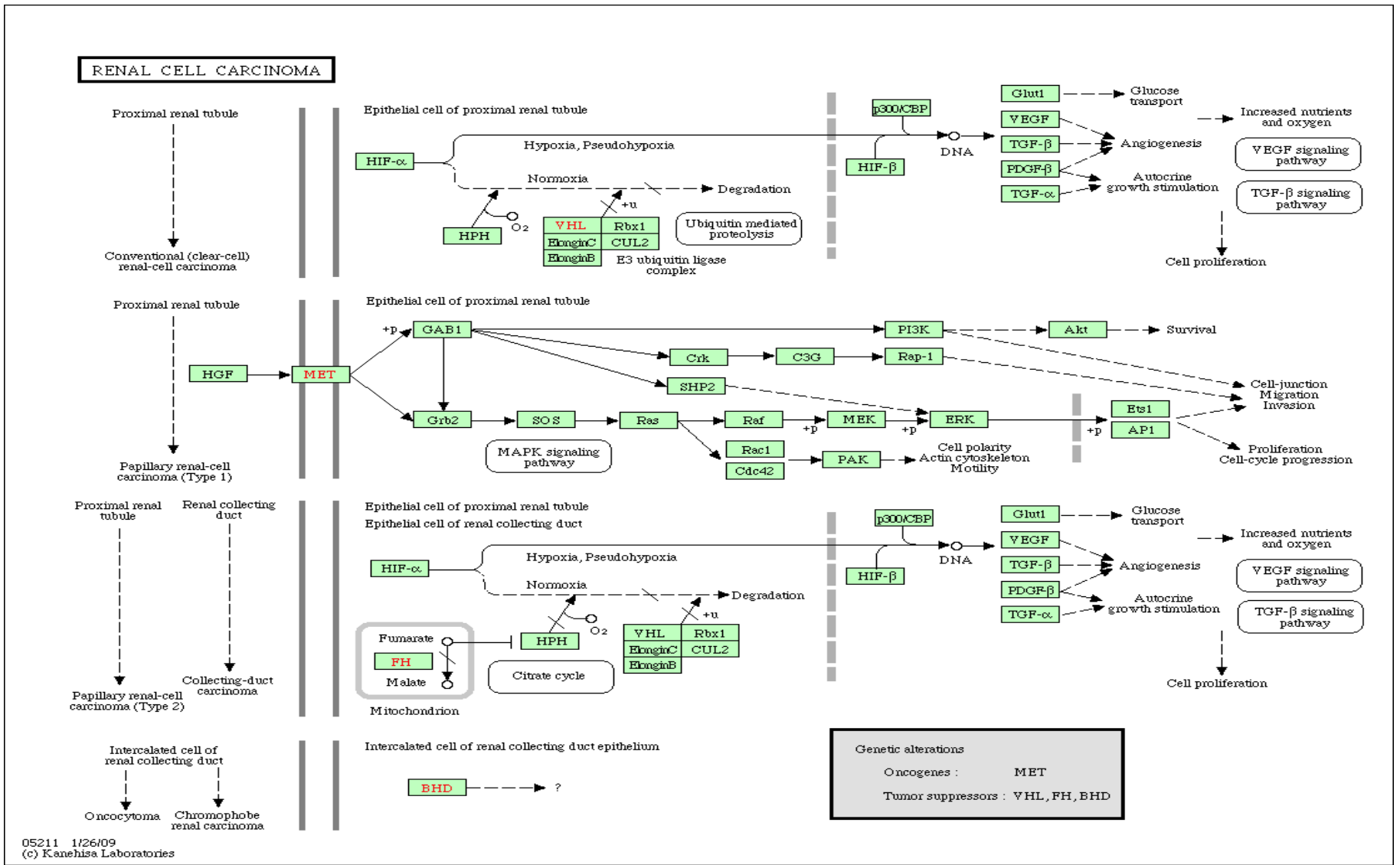
# Pancreatic cancer



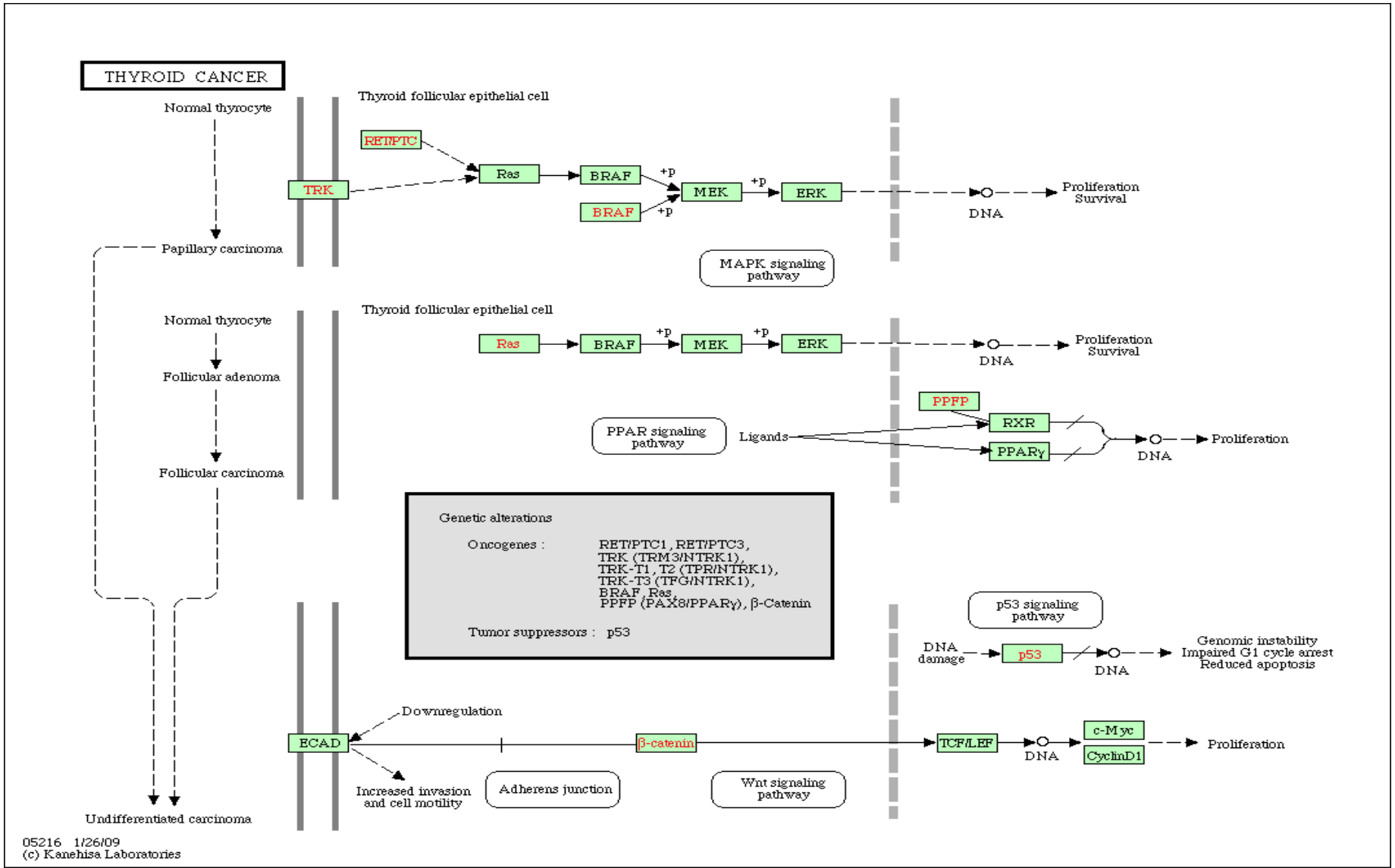
# Prostate cancer



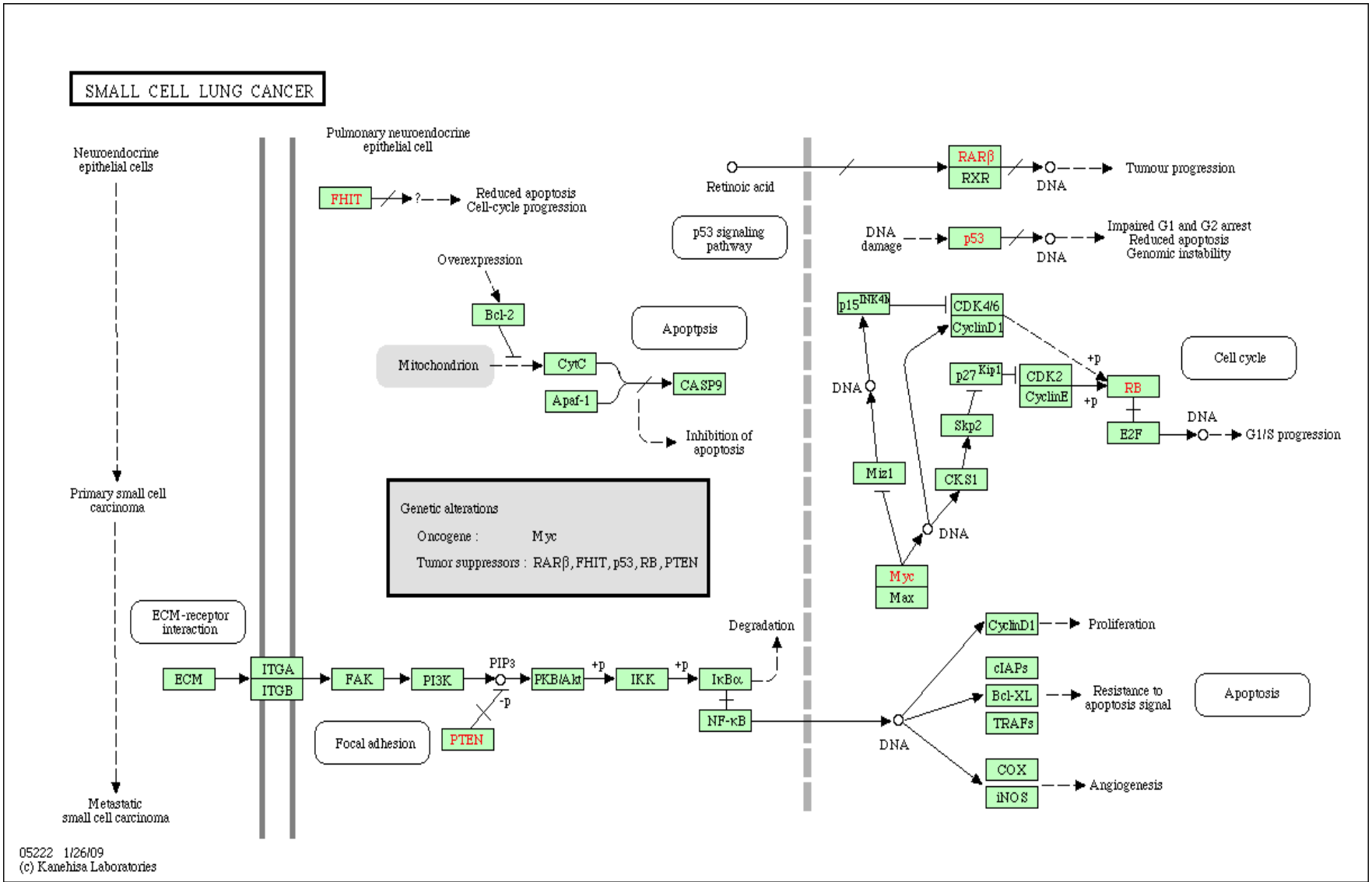
# Renal cell carcinoma



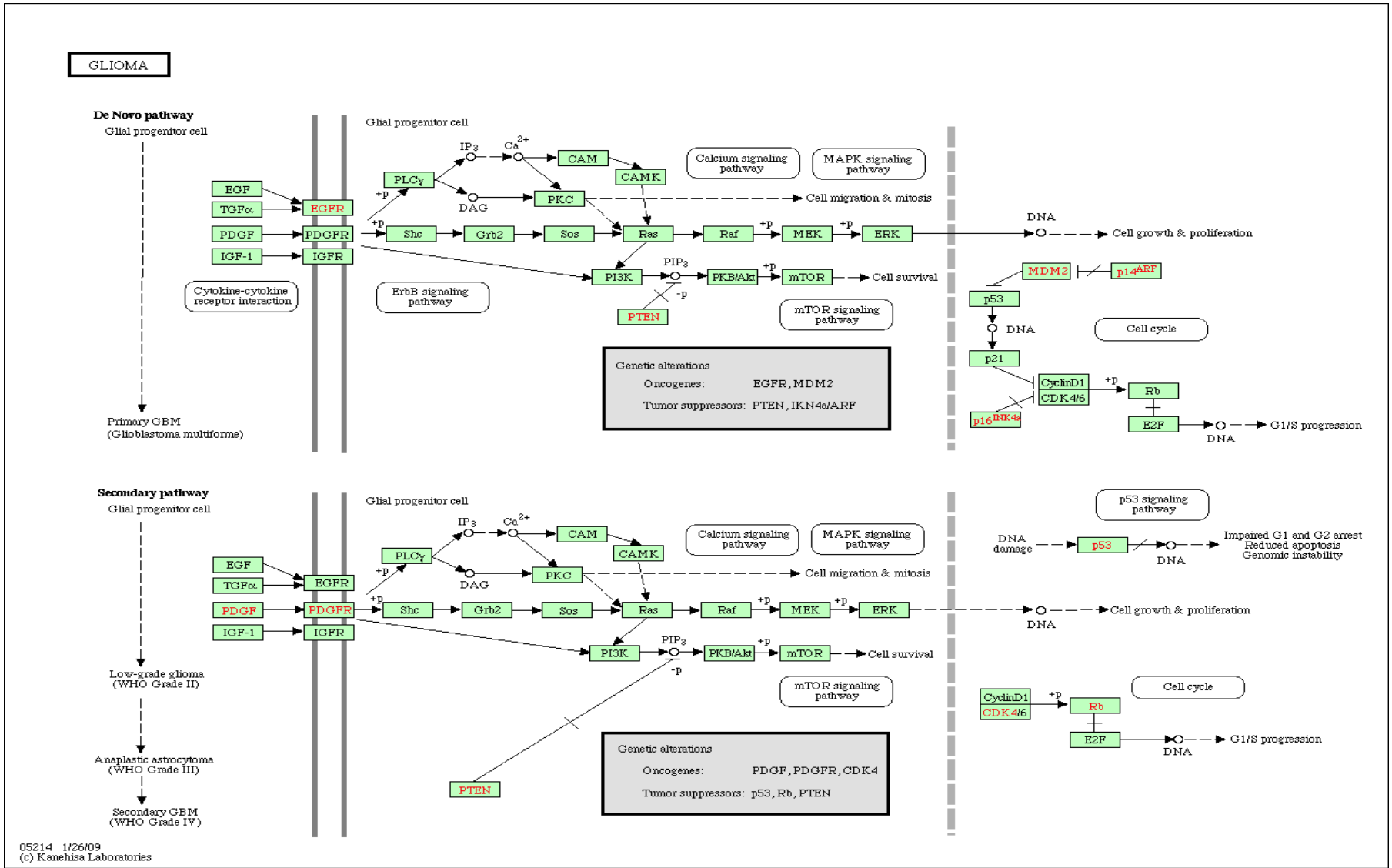
# Thyroid cancer



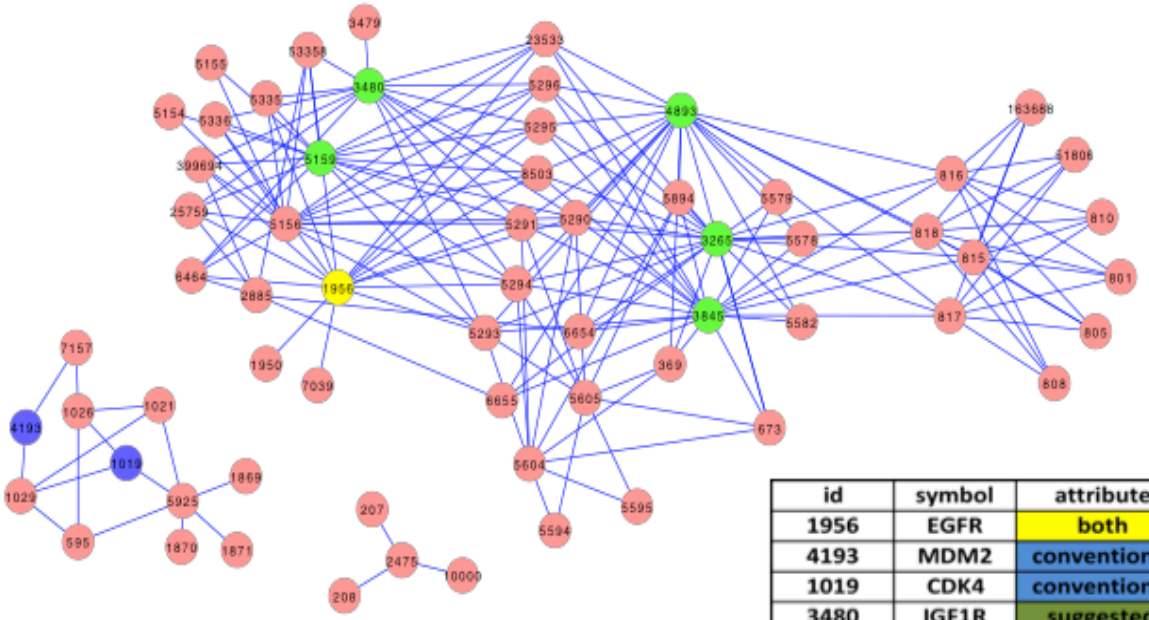
# Small cell lung cancer



# glioma


GLIOMA
De Novo pathway
Glial progenitor cell
Primary GBM
(Glioblastoma multiforme)
Glial progenitor cell
EGF
TGFα
PDGF
IGF-1
EGFR
PDGFR
IGFR
Cytokine-cytokine receptor interaction
PLCγ
IP3
Ca2+
CAM
CAMK
DAG
PKC
Calcium signaling pathway
MAPK signaling pathway
Cell migration & mitosis
Shc
Grb2
Sos
Ras
Raf
MEK
ERK
DNA
Cell growth & proliferation
PI3K
PIP3
PKB/Akt
mTOR
Cell survival
PTEN
ErbB signaling pathway
mTOR signaling pathway
MDM2
p14ARF
p53
DNA
Cell cycle
p21
CyclinD1
CDK4/6
Rb
p16INK4a
E2F
G1/S progression
Genetic alterations
Oncogenes: EGFR, MDM2
Tumor suppressors: PTEN, IKN4a/ARF
Secondary pathway
Glial progenitor cell
Low-grade glioma
(WHO Grade II)
Anaplastic astrocytoma
(WHO Grade III)
Secondary GBM
(WHO Grade IV)
p53 signaling pathway
DNA damage
Impaired G1 and G2 arrest
Reduced apoptosis
Genomic instability
Genetic alterations
Oncogenes: PDGF, PDGFR, CDK4
Tumor suppressors: p53, Rb, PTEN
05214 1/26/09
(c) Kanehisa Laboratories

# Glioma Carcinoma Protein-Protein Interaction Network




| id | symbol | attribute |
|---|---|---|
| 1956 | EGFR | both |
| 4193 | MDM2 | conventional |
| 1019 | CDK4 | conventional |
| 3480 | IGF1R | suggested |
| 3265 | HRAS | suggested |
| 4893 | NRAS | suggested |
| 3845 | KRAS | suggested |
| 5159 | PDGFRB | suggested |

# Calculating Betweenness-Centrality

Betweenness centrality, or just betweenness, is a network topological metric and a measure of the centrality of a node, $v_i$. Specifically, it is the sum of the fractions of shortest paths that pass through $v_i$. The relation is given by

$$c_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (1)$$

where $\sigma_{st}$ is the number of shortest paths between two nodes $(s,t)$ and $\sigma_{st}(v)$ is the number of those paths passing through $v_i$ (Newman, 2010). In other words, betweenness centrality is a measure of the extent that a node lays on the paths between other nodes. This is important because it may indicate the influence within the network that this node plays in controlling information transfer between other nodes.

# Calculating Degree Entropy

The second network topology metric we explored, for which we did find correlation with 5-year survival probability, was network entropy, specifically degree-entropy, which is simply stated as:

$$H = -\sum_{k=1}^{N-1} p(k) \log p(k) \qquad (2)$$

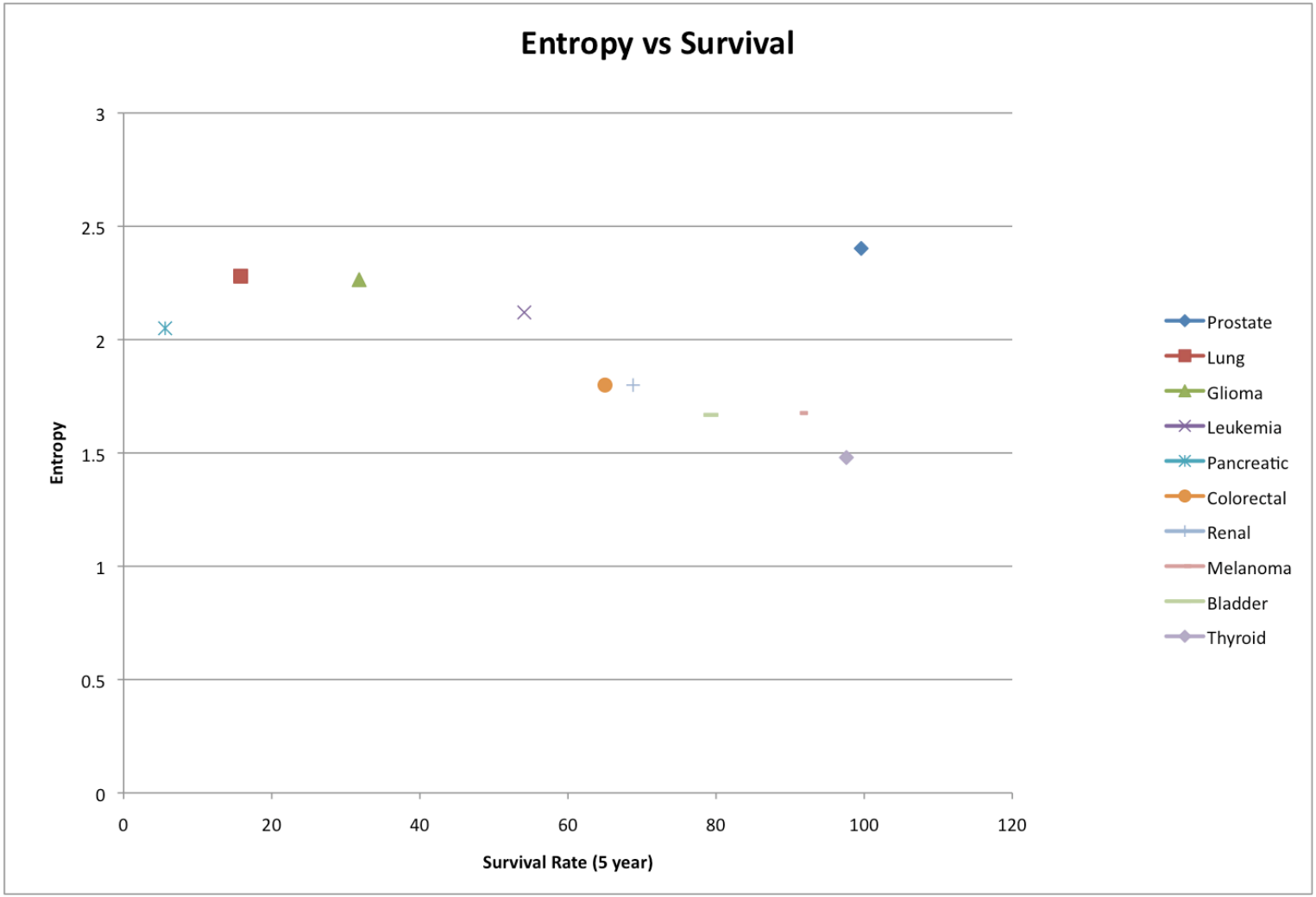
where *N* is the total number of nodes in the network and *p(k)* is the degree (number of incident lines) of node k (Wang, et al. 2006). In words, the degree-entropy provides a measure of the network's heterogeneity and complexity.

| cancer | SEER | H | nodes | edges | degree | B1 | B2 | B3 |
|---|---|---|---|---|---|---|---|---|
| AML | 23.6 | 2.0998 | 60 | 170 | 5.533 | 2322 | 6688 | 3728 |
| CML | 55.2 | 2.1607 | 73 | 185 | 5.041 | 2885 | 4193 | 9846 |
| colorectal | 63.6 | 1.7994 | 62 | 104 | 3.3548 | 3845 | 5900 | 1499 |
| glioma | 33.4 | 2.2646 | 65 | 189 | 5.8154 | 1956 | 3480 | 5159 |
| melanoma | 91.2 | 1.6761 | 71 | 281 | 7.4648 | 4893 | 5604 | 5595 |
| NSCL | 18 | 2.3584 | 54 | 124 | 4.6481 | 3845 | 11186 | 595 |
| renal | 69.5 | 1.7691 | 70 | 109 | 3.1143 | 2549 | 5981 | 5594 |
| SCL | 6.2 | 2.212 | 84 | 219 | 5.2262 | 4792 | 5747 | 595 |
| thyroid | 97.2 | 1.4798 | 29 | 49 | 3.379 | 3265 | 4893 | 3845 |
| bladder | 78.1 | 1.668 | 42 | 46 | 2.1905 | 5605 | 5604 | 5595 |
| endometral | 68.6 | 1.8352 | 52 | 87 | 3.2308 | 2885 | 105 | 5170 |
| basal | 91.4 | 1.8768 | 55 | 310 | 11.273 | 2932 | 1499 | 2735 |
| pancreatic | 5.5 | 2.0501 | 70 | 137 | 3.9143 | 3845 | 10928 | 3716 |
| prostate | 99.4 | 2.4025 | 89 | 295 | 6.6292 | 2885 | 2932 | 207 |

**Table 1. Cancer, survival probability, network statistics. Here, H stands for degree-entropy; nodes for the number of nodes; edges for the number of edges; degree for the average degree. The symbols B1, B2, B3 indicate the Entrez ID's for the top three betweenness centrality nodes, respectively.**

Table 1. Cancer survival probabilities and network statistics for 14 cancer types. The columns B1, B2, and B3 give the HGNC gene symbols[19] for the top three betweenness centrality nodes. Table reproduced from Breitkreutz *et al*[12].

| Cancer Type | B1 | B2 | B3 |
|---|---|---|---|
| Acute myeloid leukemia | FLT3 | SPI1 | JUP |
| Basal cell carcinoma | GSK3B | CTNNB1 | GLI1 |
| Bladder cancer | MAP2K2 | MAP2K1 | MAPK3 |
| Chronic myeloid leukemia | GRB2 | MDM2 | GAB2 |
| Colorectal cancer | KRAS | RALGDS | CTNNB1 |
| Endometrial cancer | GRB2 | ADARB2 | PDPK1 |
| Glioma | EGFR | IGF1R | PDGFRB |
| Melanoma | NRAS | MAP2K1 | MAPK3 |
| Non small-cell lung cancer | KRAS | RASSF1 | CCND1 |
| Pancreatic cancer | KRAS | RALBP1 | JAK1 |
| Prostate cancer | GRB2 | GSK3B | AKT1 |
| Renal cell carcinoma | GAB1 | RFC1 | MAPK1 |
| Small cell lung cancer | NFKBIA | PTK2 | CCND1 |
| Thyroid cancer | HRAS | NRAS | KRAS |


Entropy vs Survival
Entropy
3
2.5
2
1.5
1
0.5
0
0
20
40
60
80
100
120
Survival Rate (5 year)
Prostate
Lung
Glioma
Leukemia
Pancreatic
Colorectal
Renal
Melanoma
Bladder
Thyroid

# Degree-Entropy of PPI Networks is Correlated with percent 5-yr Survival

## Molecular signaling network complexity is correlated with cancer patient survivability

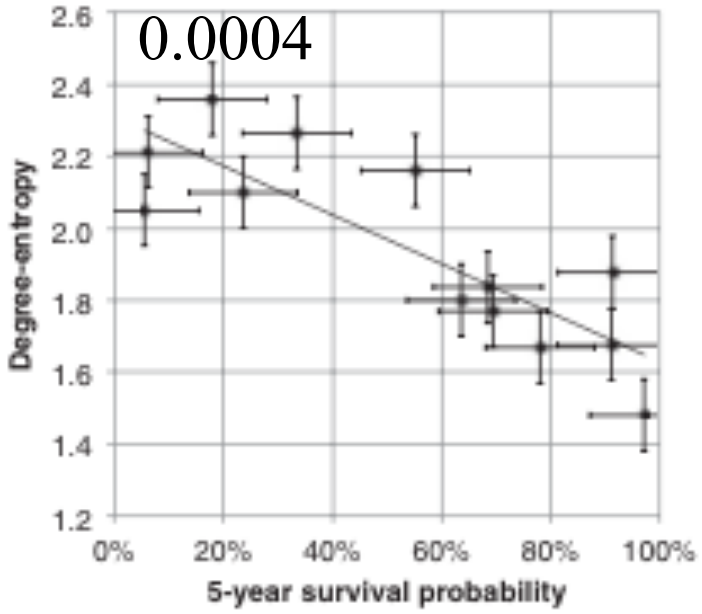
Dylan Breitkreutz[a,b], Lynn Hlatky[c], Edward Rietman[c], and Jack A. Tuszynski[a,b,1]

PMID: 22615392

PNAS, June 5, 2012, 9209-9212

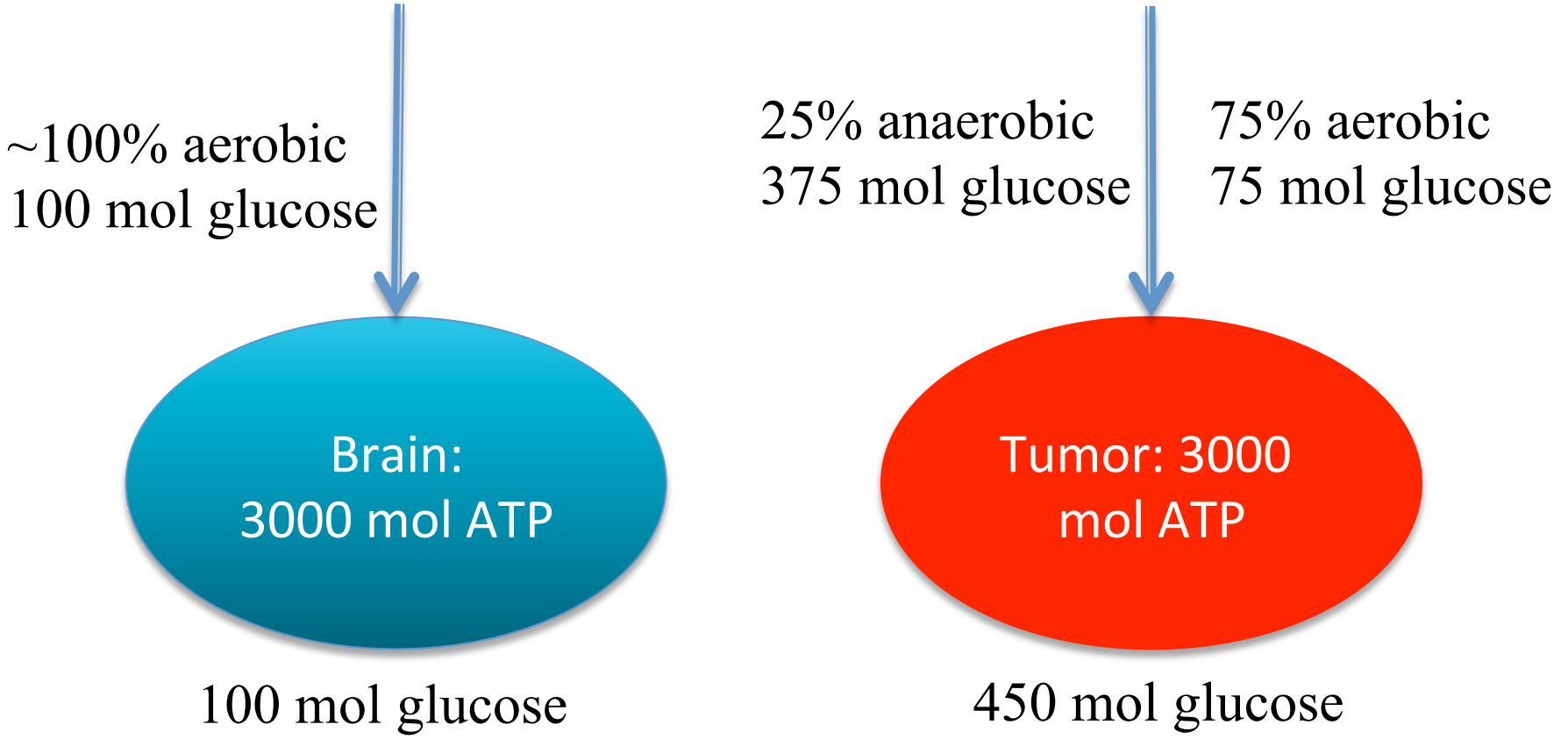
- Compute degree-entropy on each KEGG cancer network and then plot with survival.
- To confirm that the correlation was not an artifact we computed the entropy of a population of 1,000 Erdös-Rényi similar-sized random networks and 1,000 similar-sized Barabási scale-free networks for each of the cancer networks. We found statistically no correlation with survival.
- Considering that we are correlating two highly unrelated databases it is remarkable that we got such a good correlation.

R = -0.84, p = 0.0004



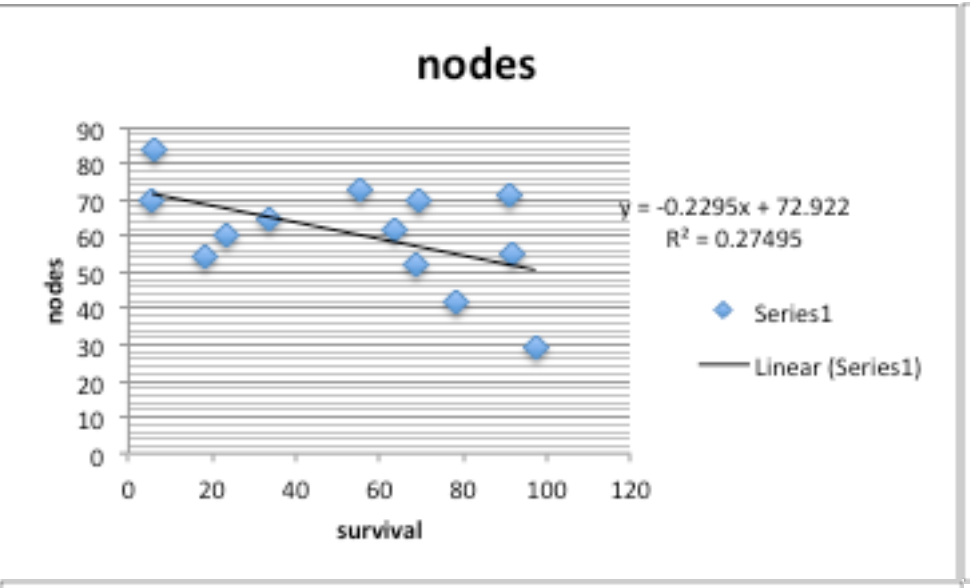
- A possible correlation between entropy and lethality is seen.
- These survival rates however, take into account all methods of treatment.
- To improve both the reliability and clarity of this correlation a few things are being done:
- Survival statistics of patients who refused treatment and those that only received chemotherapy will be used.
- Check how the random deletion of edges affects the entropy of each pathway.

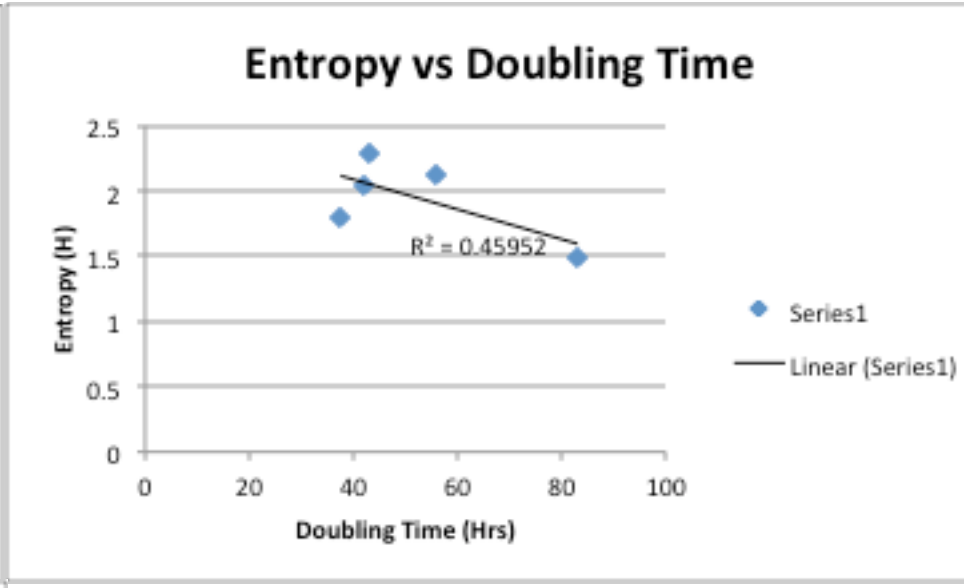
# Tumor consumes vast amounts of glucose to fulfill energy requirements

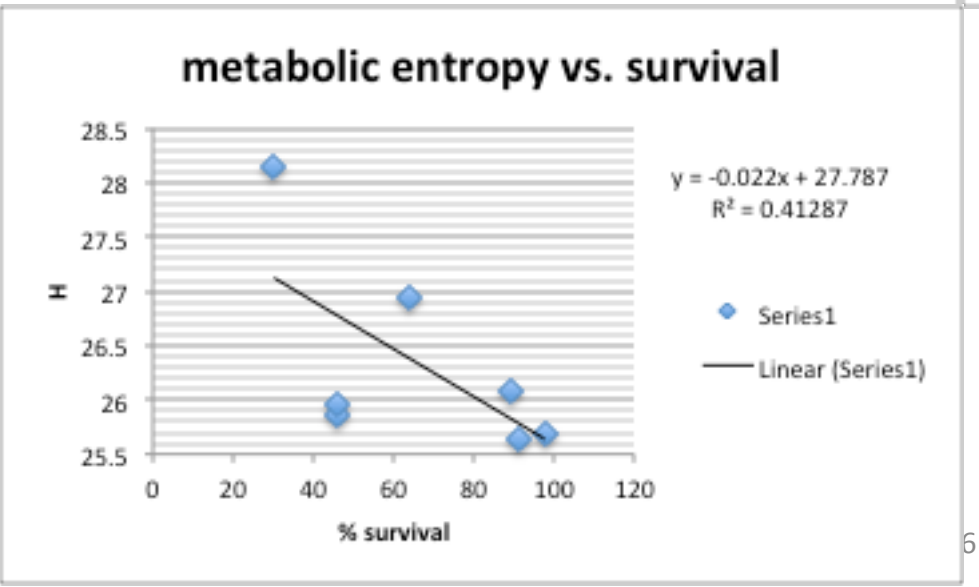



**Aerobic: 1 glucose = 30 ATP**
**Anaerobic: 1 glucose = 2 ATP**

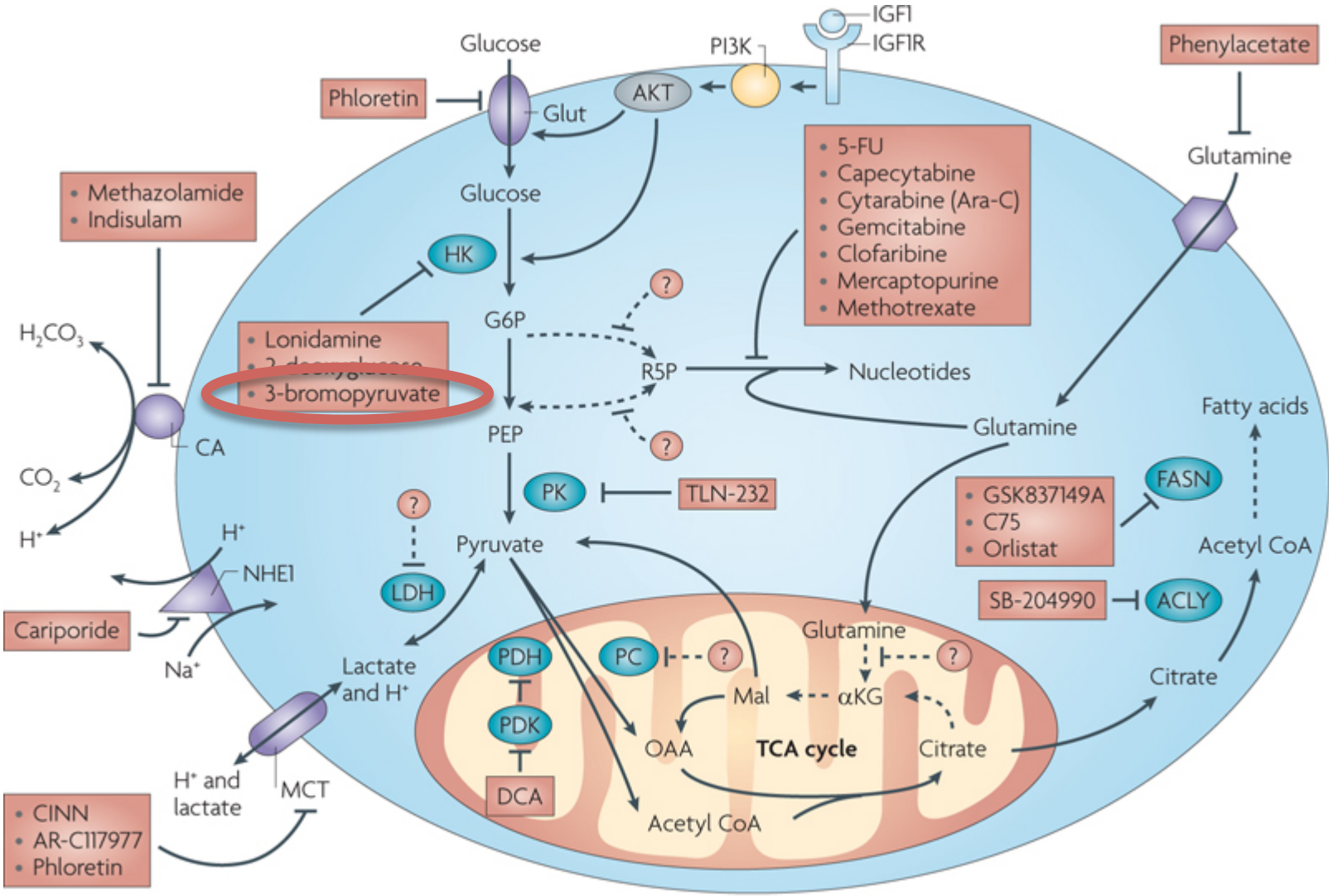
# Preliminary Work in Progress: Metabolic Entropy-Warburg Effect



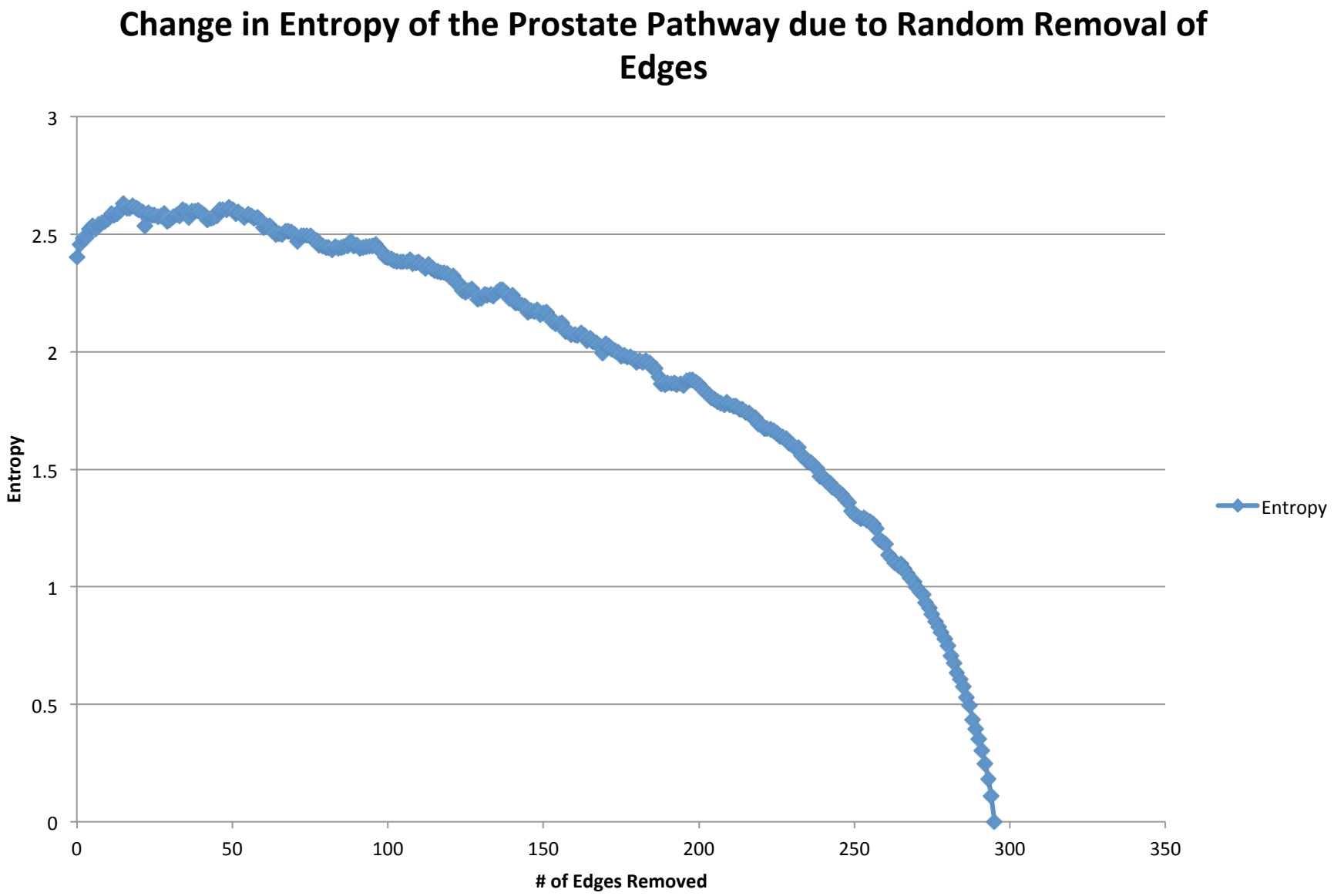
# Possible Targets

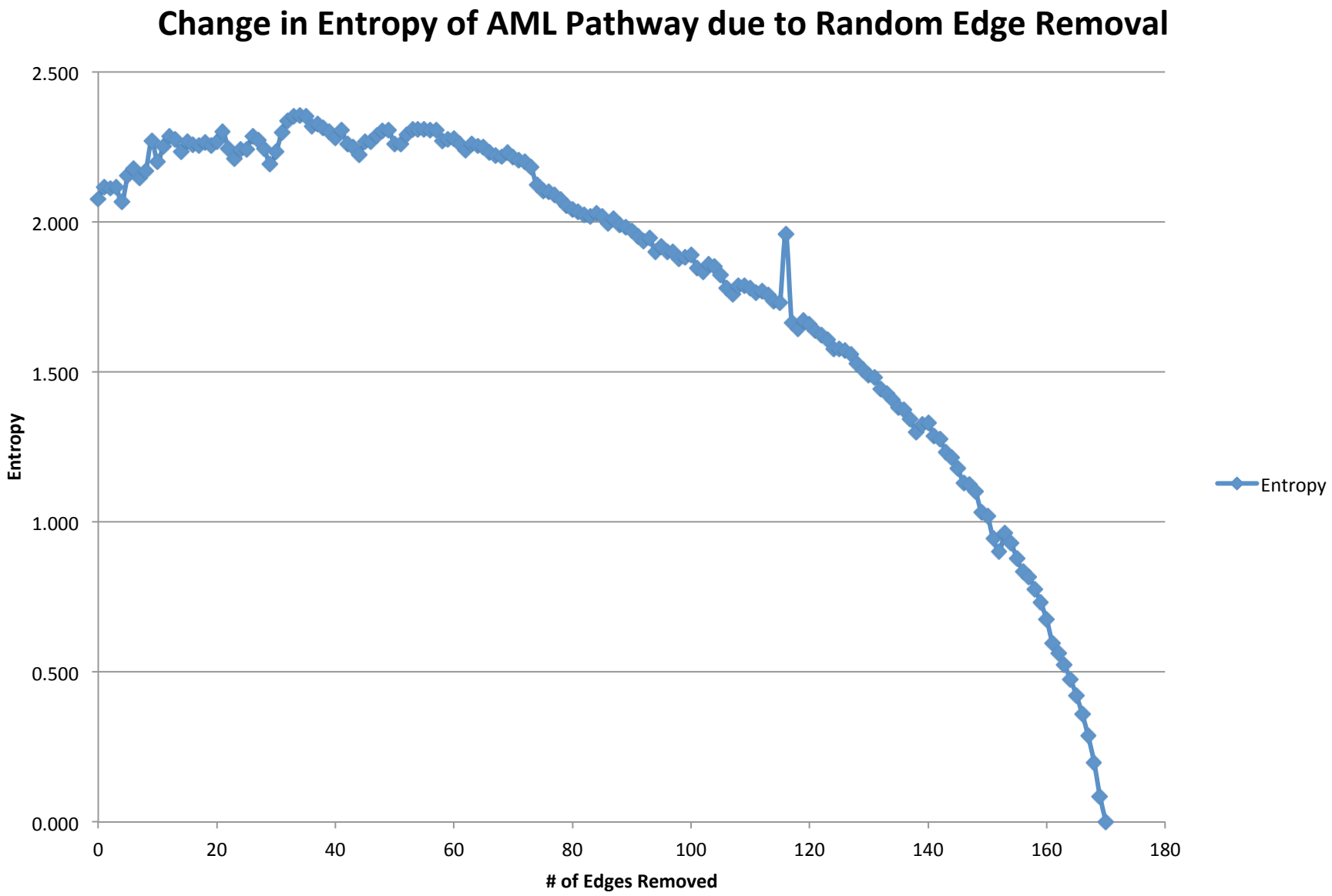


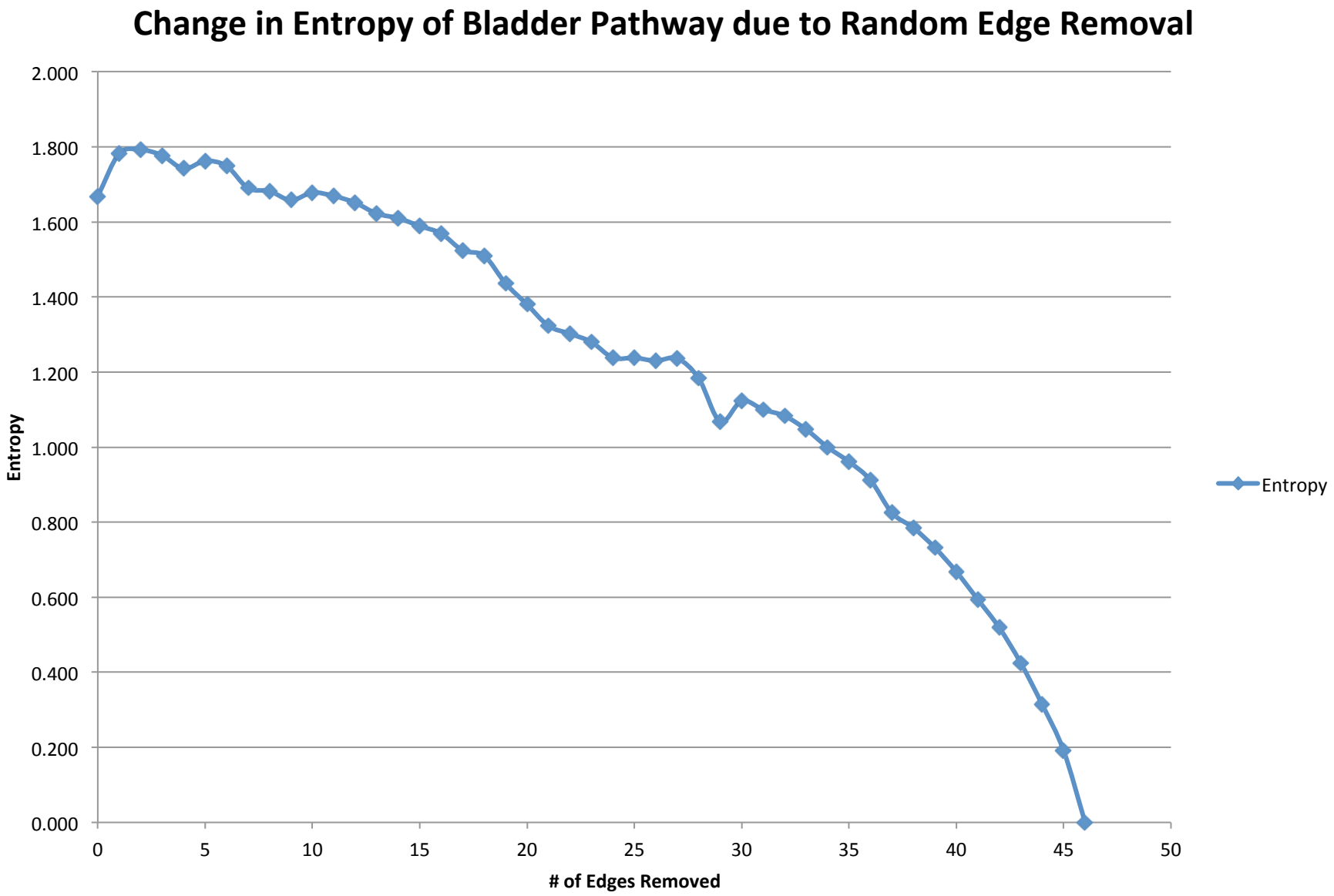

Tennant et al., Nature Review Cancer 10:267-277, 2010.

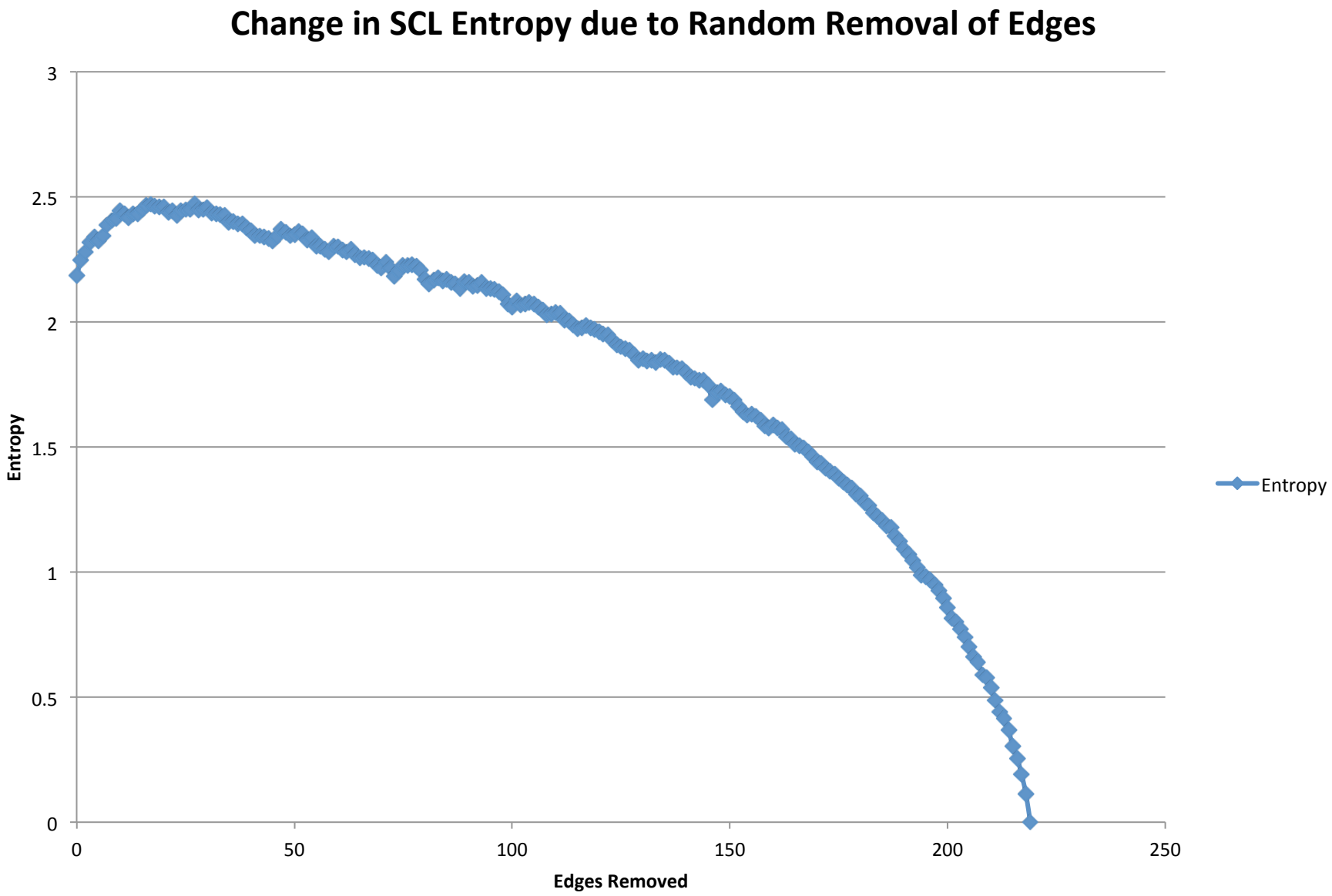
# Randomly “Drugged” Signaling Pathways

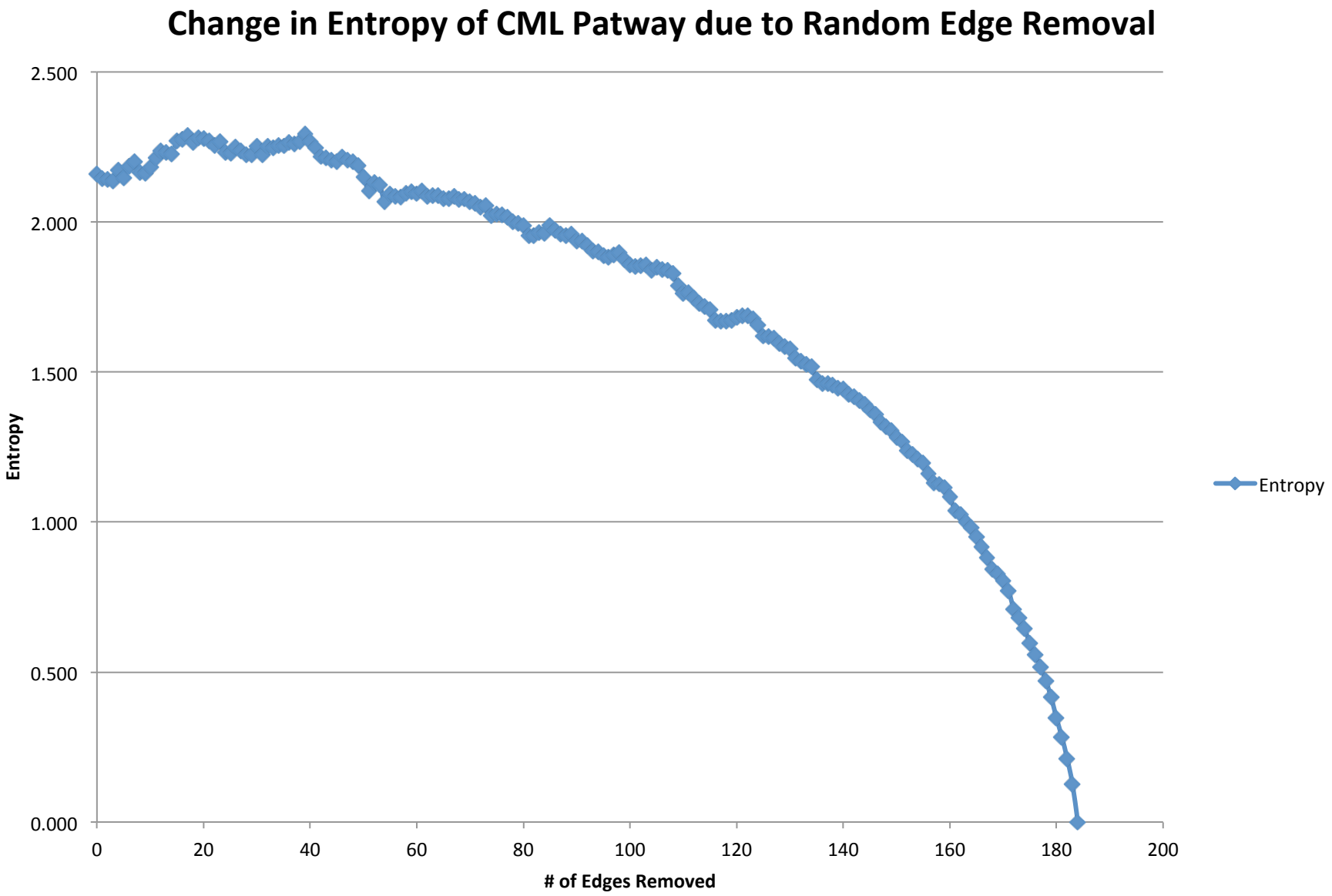
- Went through each of the pathways and altered them depending on the drugs that inhibited certain interactions.
- Inhibition was represented as the removal of certain parts of the graph that could only be reached by the inhibited interaction.
- The results were less than extraordinary, only a few pathways were altered by more than about $\Delta H=0.4$

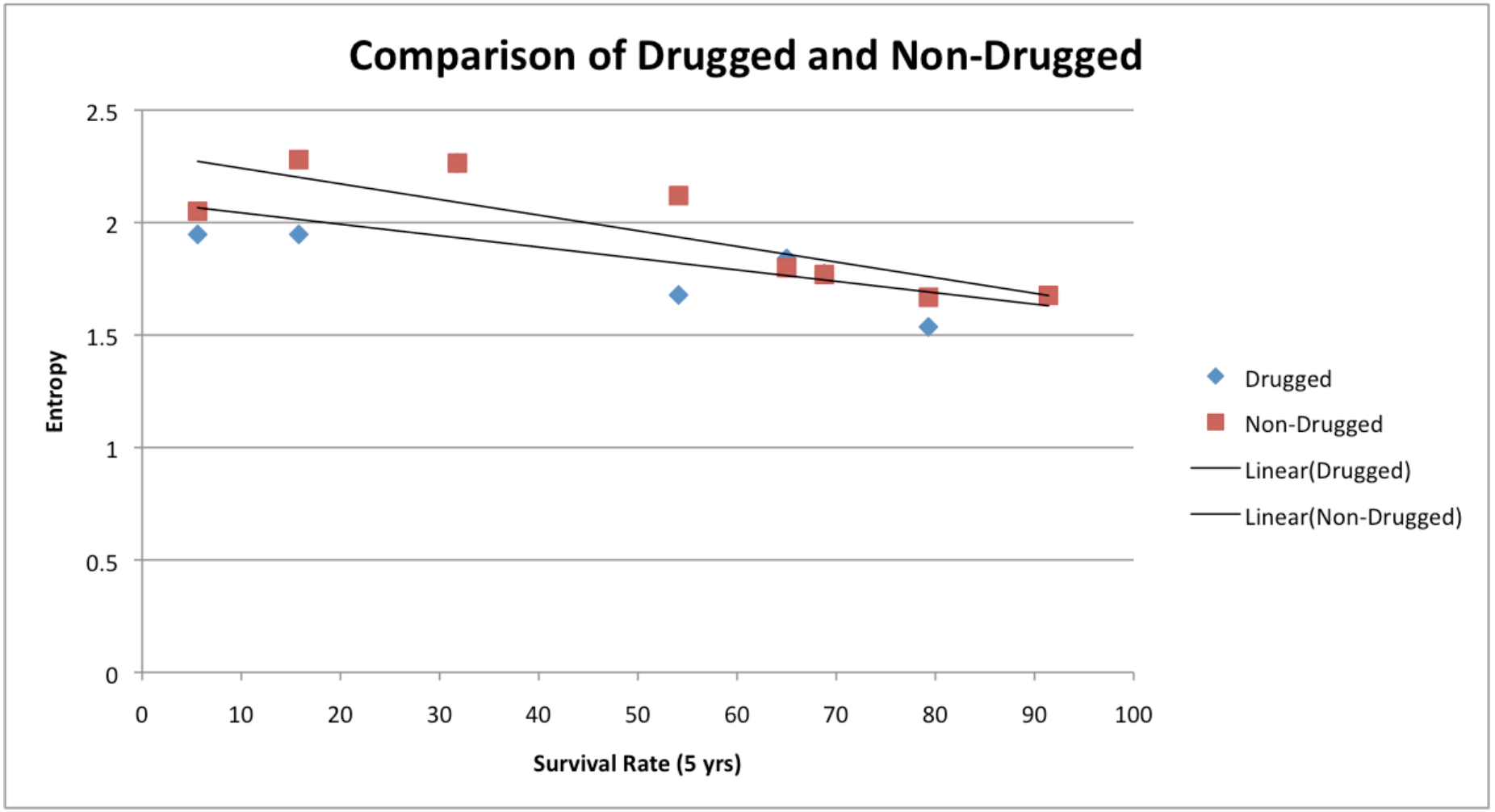

Change in Entropy of the Prostate Pathway due to Random Removal of Edges
Entropy
3
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
300
350
# of Edges Removed
Entropy


Change in Entropy of AML Pathway due to Random Edge Removal
2.500
2.000
1.500
1.000
0.500
0.000
Entropy
0
20
40
60
80
100
120
140
160
180
# of Edges Removed
Entropy


Change in Entropy of Bladder Pathway due to Random Edge Removal
2.000
1.800
1.600
1.400
1.200
1.000
0.800
0.600
0.400
0.200
0.000
Entropy
0
5
10
15
20
25
30
35
40
45
50
# of Edges Removed
Entropy


Change in SCL Entropy due to Random Removal of Edges
Entropy
3
2.5
2
1.5
1
0.5
0
0
50
100
150
200
250
Edges Removed
Entropy

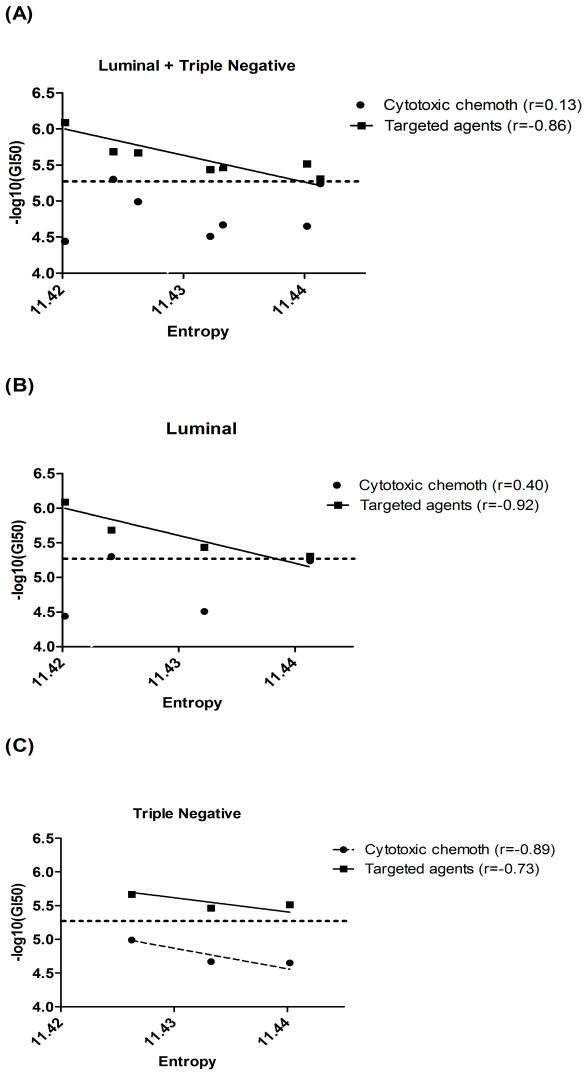

Change in Entropy of CML Patway due to Random Edge Removal
2.500
2.000
1.500
1.000
0.500
0.000
Entropy
0
20
40
60
80
100
120
140
160
180
200
# of Edges Removed
Entropy


Comparison of Drugged and Non-Drugged
Entropy
2.5
2
1.5
1
0.5
0
0
10
20
30
40
50
60
70
80
90
100
Survival Rate (5 yrs)
Drugged
Non-Drugged
Linear(Drugged)
Linear(Non-Drugged)

# Implications for Chemotherapy

- Three main avenues of application:
- The standard chemotherapeutic treatments can be investigated for target inhibition of pathway nodes
- Important nodes of the graphs may be ideal targets for new drugs.
- Combinations of several nodes can be selected for inhibition (subject to non-overlapping side effects)
- An accurate model of drug inhibition would allow for the development of new synergistic chemotherapy regimens.

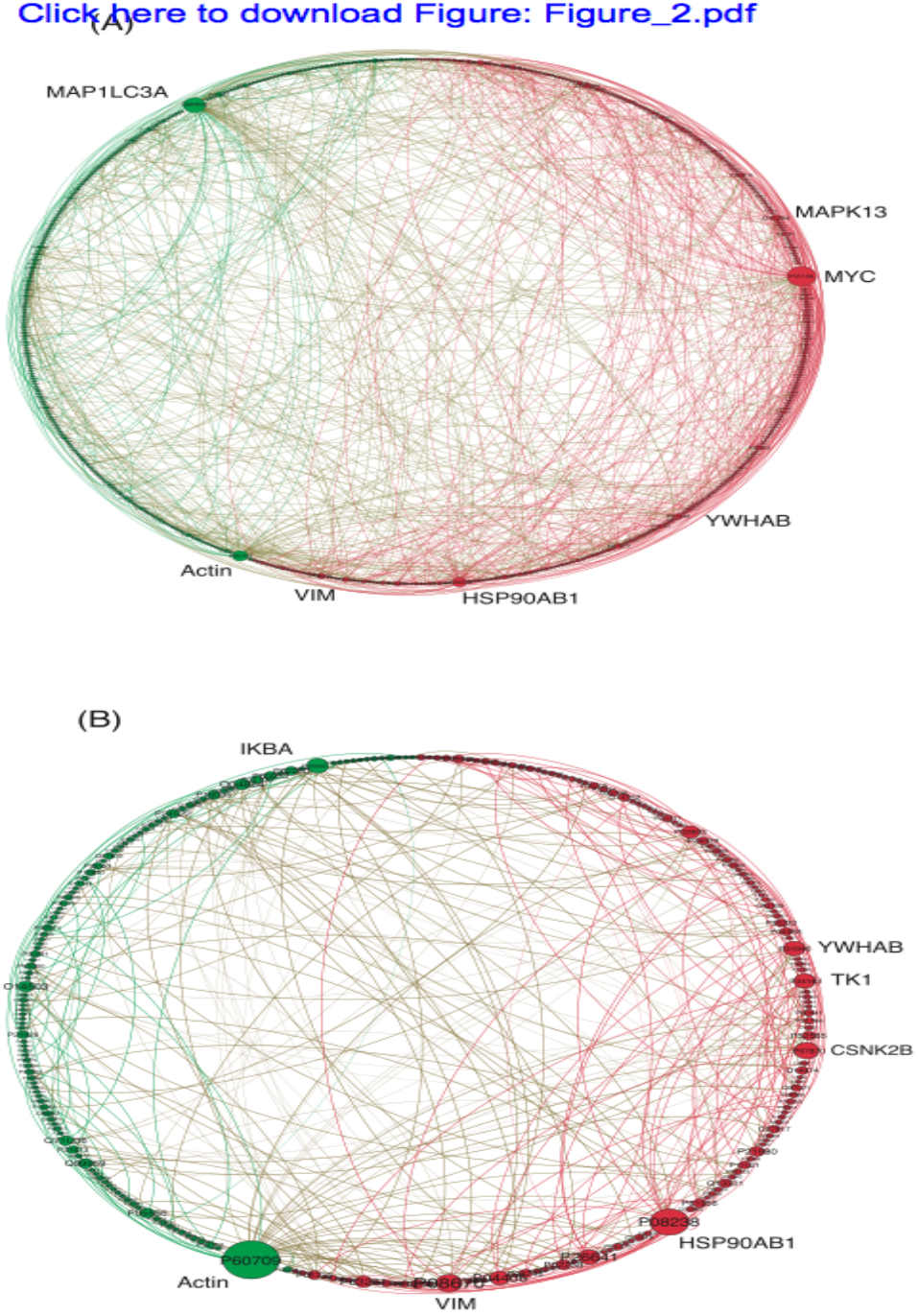
# Breatst cancer subtypes



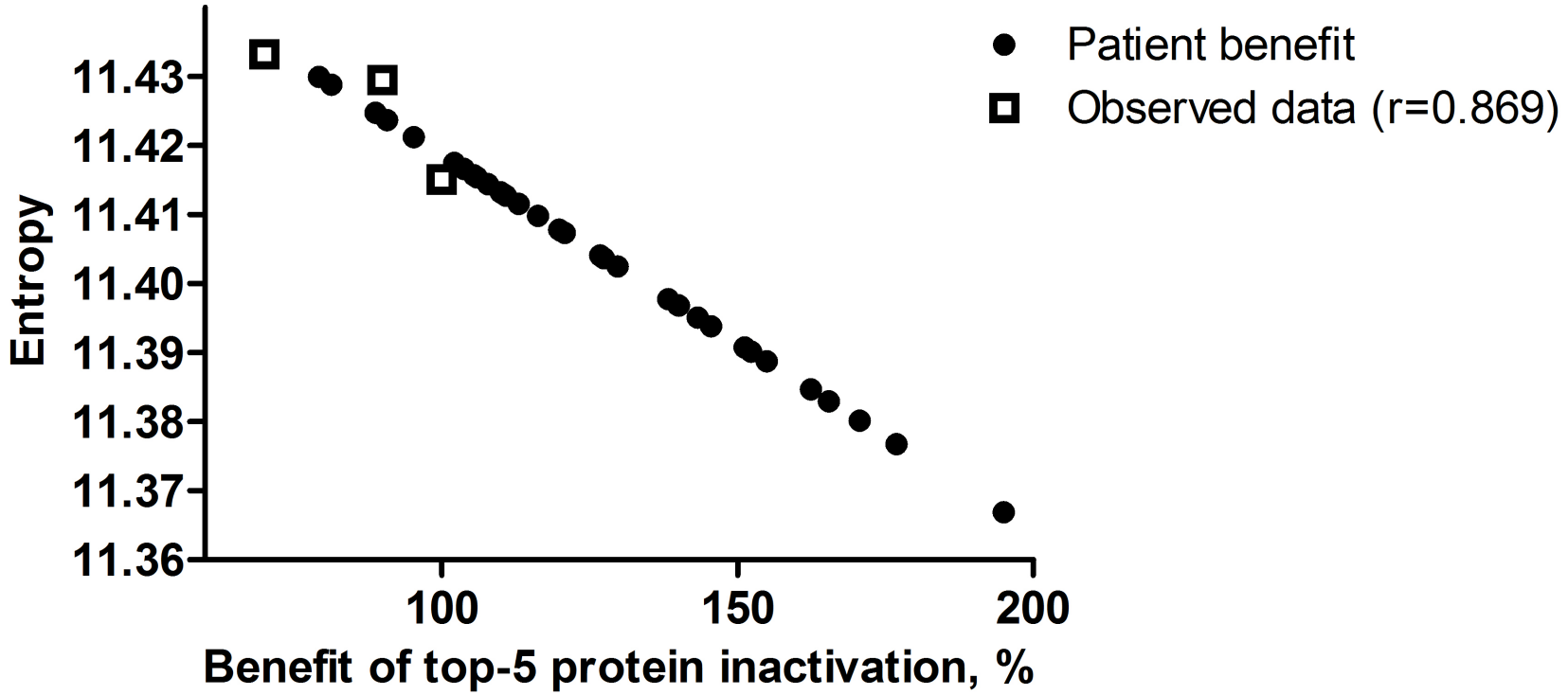
MDA-MB-231 vs. MCF-10A

Figure

Click here to download Figure: Figure_2.pdf



# Calculated patient benefit on the use of top 5 inhibitors

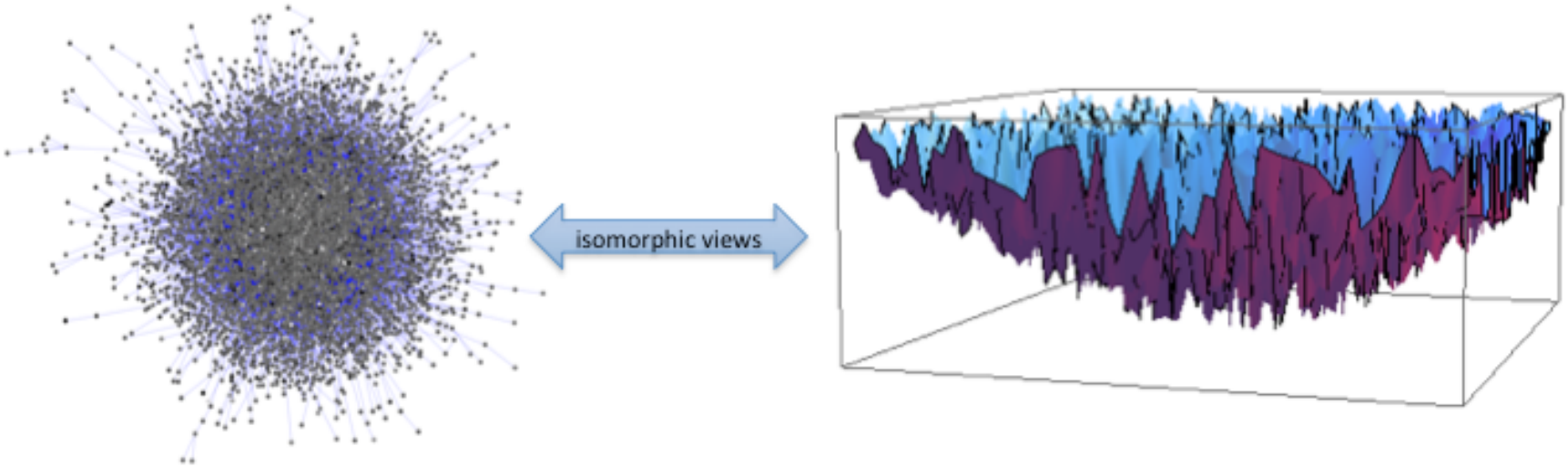


**Triple-Negative Breast Cancer Top Five targets**

- Epidermal growth factor receptor EGFR
- Heat shock protein HSP 90-beta HSP90AB1
- Mitogen-activated protein kinase 13 MAPK13
- 14-3-3 protein beta/alpha YWHAB
- Protein mago nashi homolog MAGOH

# Gibbs Free Energy:The Basic Idea and the Math

- The interaction energy between two molecules is known as the chemical potential
- A cell is a massive network of molecular species
- We can represent the cell as a PPI network with scalar numbers associated with each protein – the Gibbs free-energy
- The Gibbs free-energy for a particular protein is given by




network view

energy landscape view

Shading represents Gibbs energy

$$G_i = c_i \ln \frac{c_i}{\sum_j c_j}$$

where $c_i$ = normalized expression

# Gibbs Free-Energy on PPI Networks from mRNA Expression

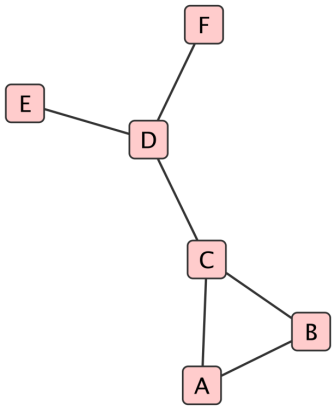



**Figure 1: Example of a small PPI network**. The nodes (A-F) represent individual proteins, the lines, called the edges, represent protein-protein interactions. No information about the directionality of the interaction is shown.

$$G_i = c_i \ln \frac{c_i}{\sum_{j=i} c_j}$$

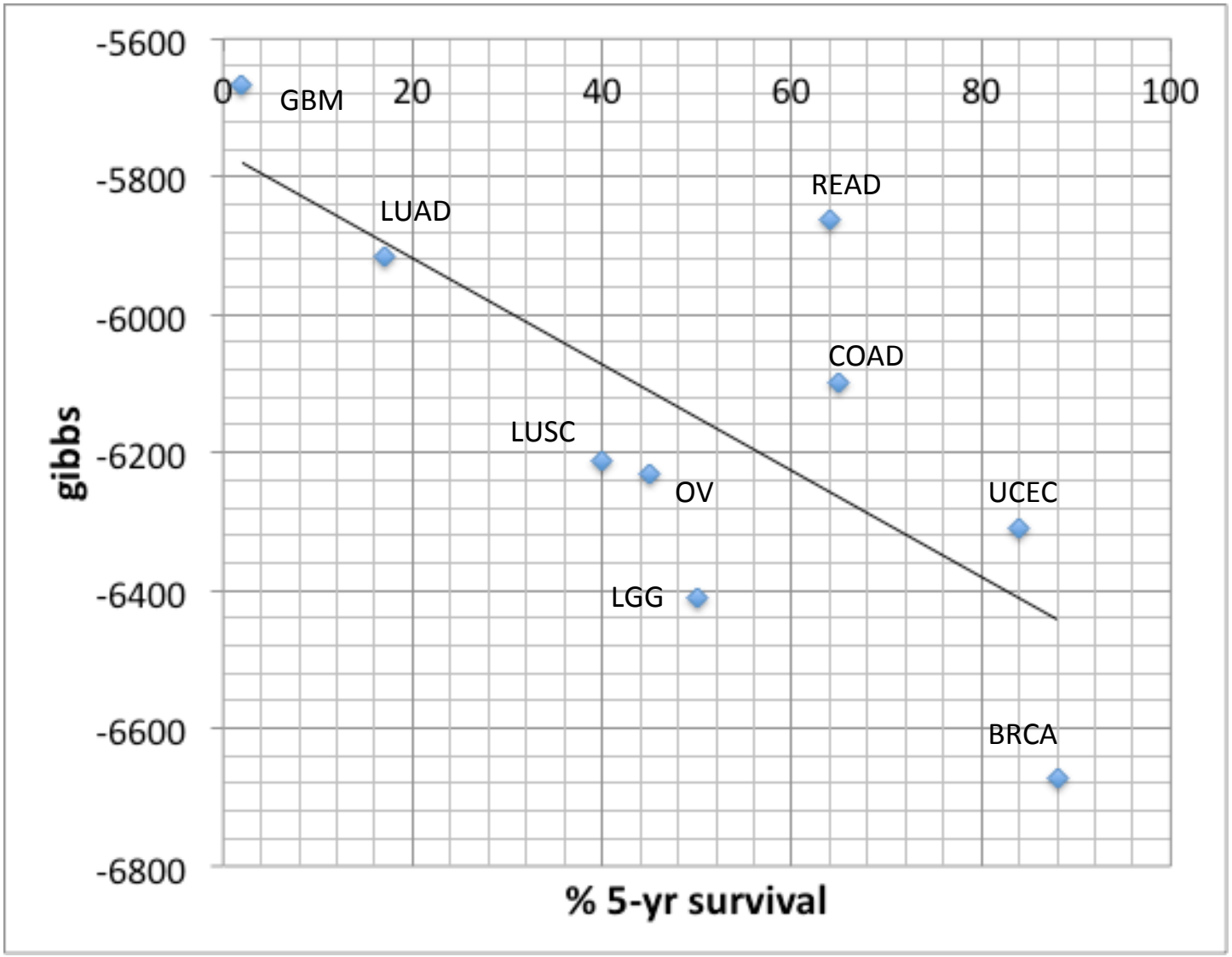
where $c_i$ = normalized expression, representing concentration

- Log2 transformed mRNA expression data are rescaled to be between 0 and 1.
- The most negative value would thus be rescaled to 0 and the most positive value would be rescaled to 1.
- These rescaled values are surrogates for protein concentration.

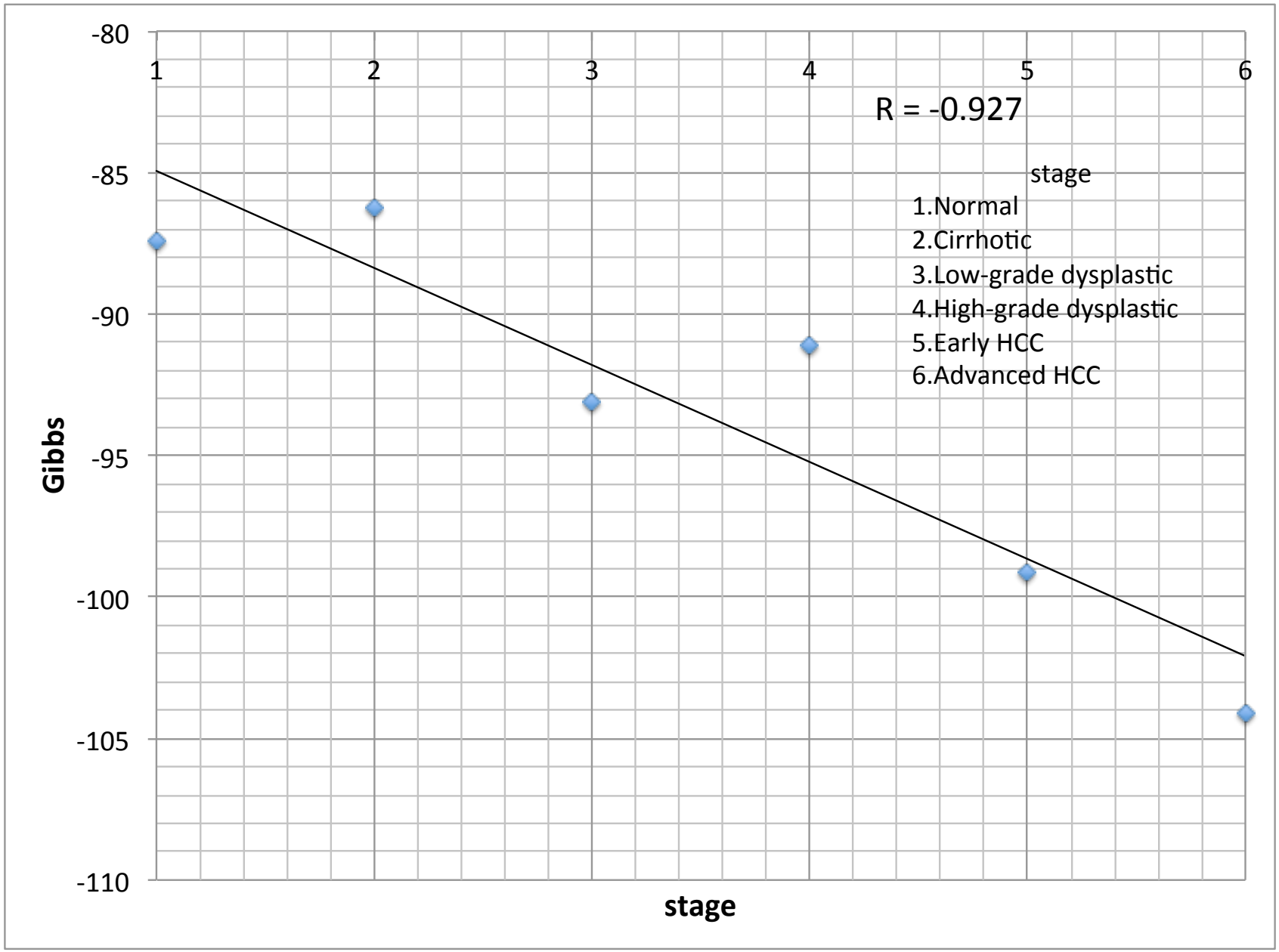
| TCGA name | cancer type | N | % 5-yr | Gibbs |
|---|---|---|---|---|
| KIRC | kidney renal clear cell | 72 | 68 | -5687 |
| KRIP | kidney relan papillary cell | 16 | 68 | -4944 |
| LGG | low grade glioma | 27 | 50 | -6411 |
| GBM | glioblastoma multiforme | 483 | 2 | -5668 |
| BRCA | breast invasive carcionma | 590 | 88 | -6674 |
| UCEC | uterin corpus endometrial | 54 | 84 | -6310 |
| OV | serous cystadenocarcinoma | 562 | 45 | -6233 |
| COAD | colon adenocarcinoma | 174 | 65 | -6099 |
| READ | rectum adenocarcinoma | 72 | 64 | -5861 |
| LUAD | lung adenocarcinoma | 32 | 17 | -5916 |
| LUSC | lung squamous cell | 155 | 40 | -6212 |

# 5 yr survival vs. Gibbs free energy on TCGA cancers

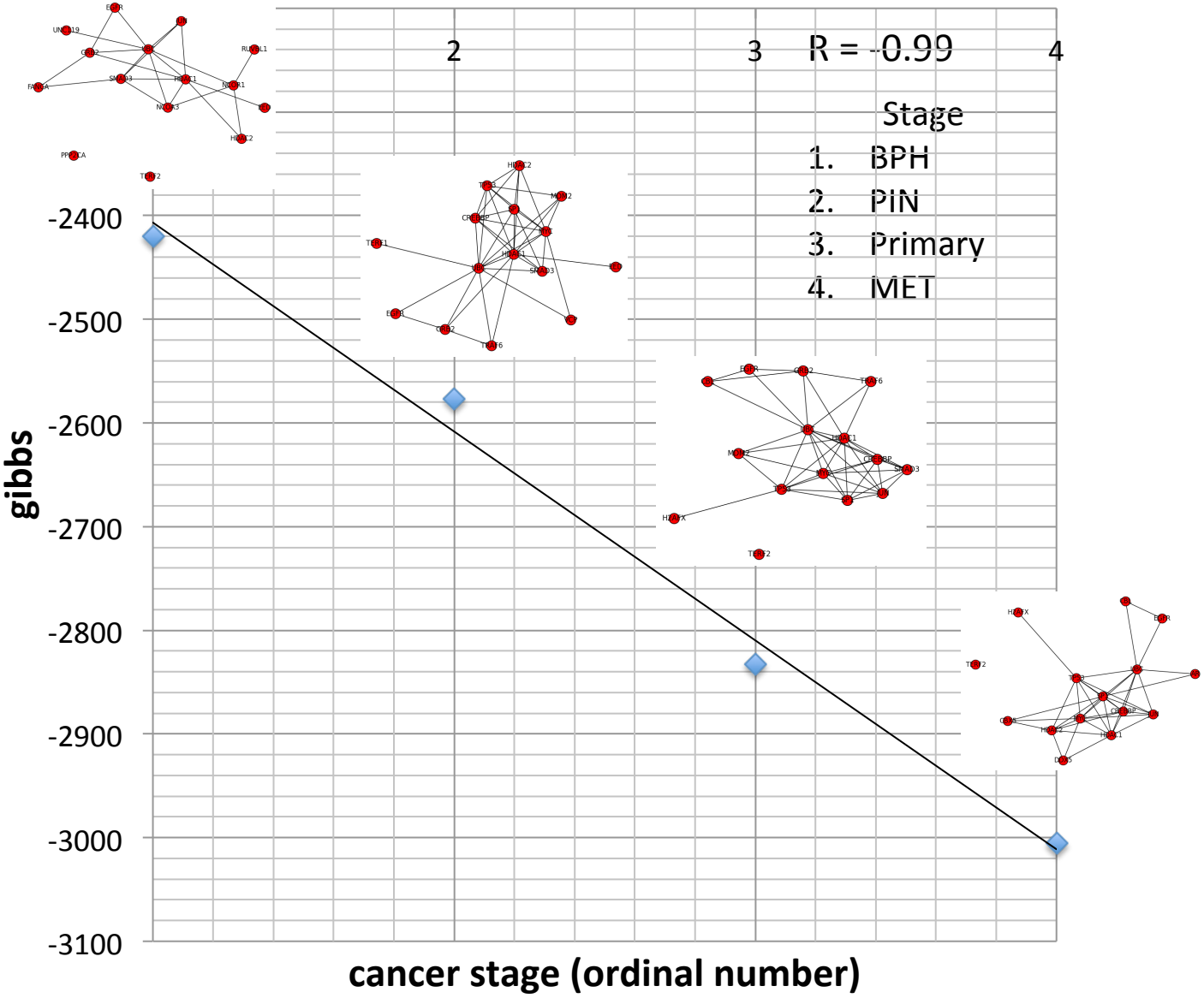
R correlation is -0.72 without KIRC, KRIP
R correlation is -0.21 with KIRC, KRIP
(KIRC, KRIP not shown)



# Liver (GEO) GSE6764

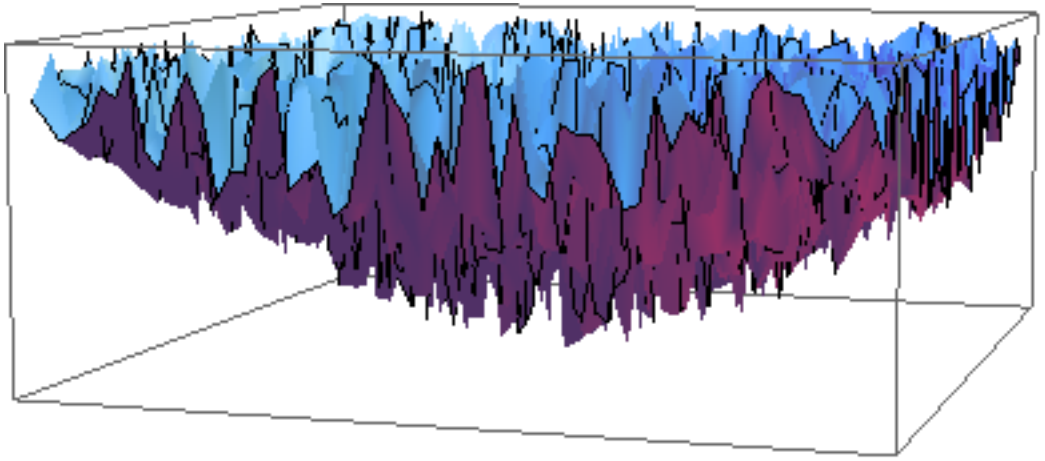


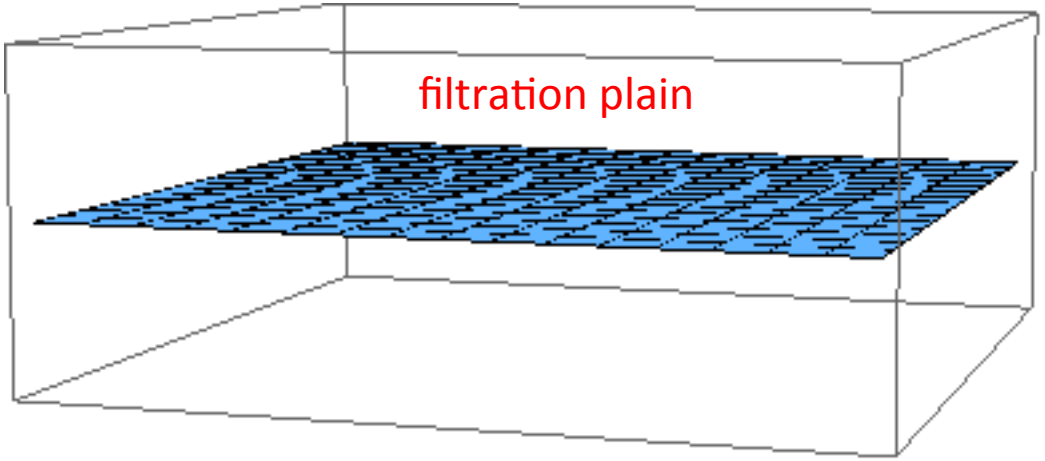
# Prostate (GEO)



# Filtration of Energy Landscape
## (another view of persistent homology)

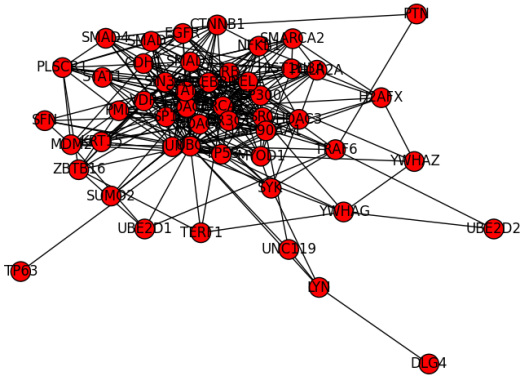
landscape to be “filtered”



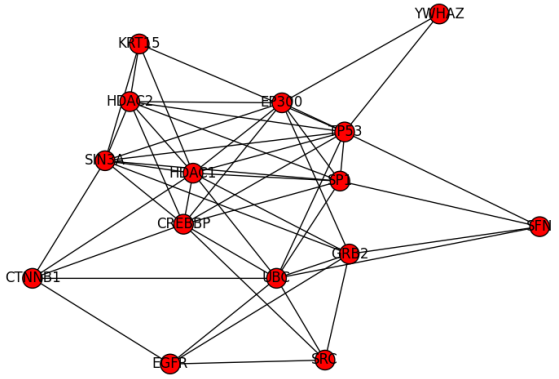



As the “filtration plain” moves up from the bottom, more-and-more nodes become exposed and larger-and-larger pathway networks come into view.

medium level threshold




low level threshold



# Gibbs Free-Energy Conclusion

- Using mRNA expression data as a surrogate for protein concentration we can calculate the Gibbs free-energy for each node in a PPI network
- Gibbs free-energy for different cancers correlates with cancer patient 5-yr survival.
- Gibbs free-energy also correlates with an ordinal scale representing cancer stage.
- These facts suggest the calculation of Gibbs free energy has captured a real thermodynamic measure of cancer.
- A Gibbs scalar function on each node allows us to calculate the Gibbs-homology for individual patients and thereby produce unique pathway networks for each patient at each stage in the cancer development.
- How to utilize this unique set of information is an ongoing research project.

# Future Projects:

- build a mathematical model with the presence of both bio-molecules and their inhibitors
- simulate the action of individual drugs as well as their combinations by setting coupled ODE's with respect to time (find parameter values!)
- show why some drug combinations are not effective in stopping cancer due to parallel pathways and redundancies
- predict the optimum efficacy of drug combinations as a function of scheduling and amplitudes

# Persistent homology

- The problem with the degree-entropy correlation is that removal of a node from the network (targeted inhibition) only changed the degree-entropy in the second decimal.
- Small changes in the dependent variable (e.g. entropy) will result in small changes in the independent variable (survival) – not a big improvement.
- We propose a different measure of complexity of the network based on topological properties of the graph: the Betti number

- Targeting a protein has an impact on the pathway that can be quantified by this number

# Persistent homology

- Graph -> filtrated simplicial complex -> Betti number

- k-dimensional homology group persisting from i to j

$$H_k^{i,j} = Z_k^i / (Z_k^i \cap B_k^j), \quad Z_k^i = \text{cycles}, \quad B_k^j = \text{boundaries}$$

$$\beta_k^{i,j} = \dim(H_k^{i,j})$$

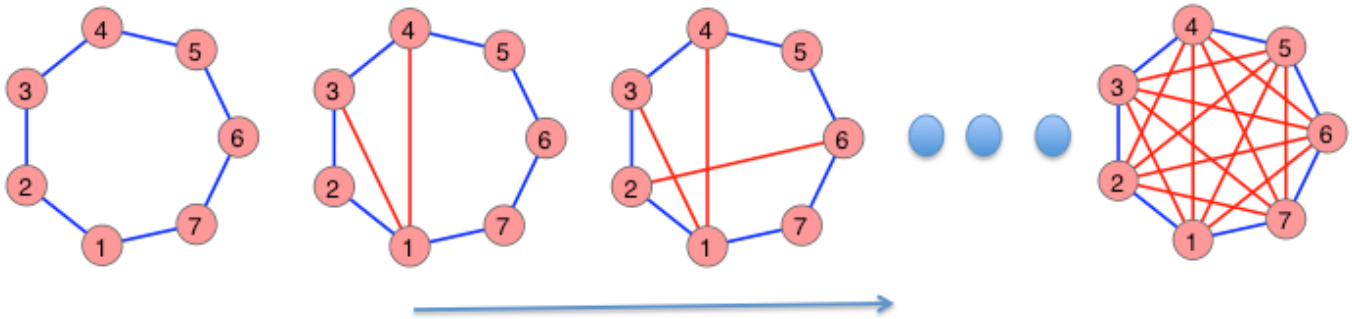
- The homology measure is called the Betti number.
- In network theory, homology measures the cycles in the network (rings of nodes).

# Betti number in graph theory

- In topological graph theory the first Betti number of a graph *G* with *n* vertices, *m* edges and *k* connected components equals m-n+k
- The "zero-th" Betti number of a graph is simply the number of connected components *k*.
- In graph theory, a **connected component** of an undirected graph is a subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph

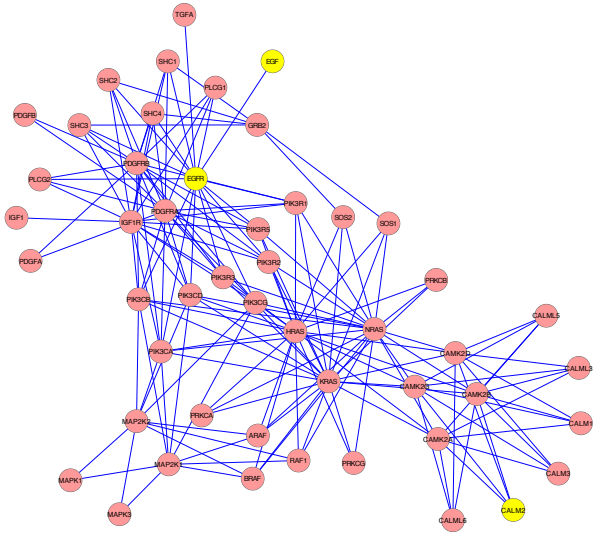
# Betti Numbers:
# A Topological Measure of Network Complexity

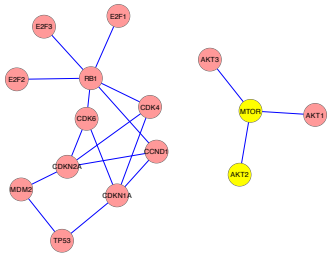
The Betti measure can be best thought of with respect to holes in the network – or cycles.




- The Jplex algorithm (http://www.math.duke.edu/~hadams/jplex/index.html )builds potentially rather large matrices of edges and vertices that are operated on to find rings.
- The software was originally developed to study “point clouds” of data in hyperspace.
- The algorithm is designed for “generic” “topological objects” not just networks.
- As we walk around the ring and drop in “virtual” edges we see that the largest ring now gets smaller. Until finally we have no more ring.
- How much connectivity is required to “kill” the ring?

# Glioma

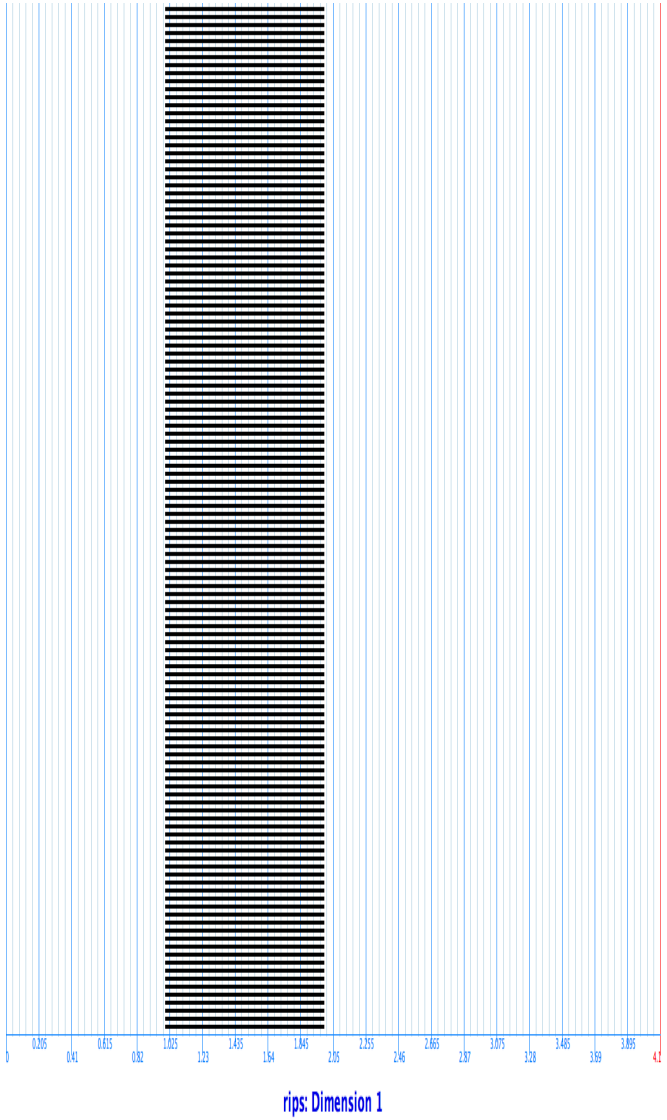






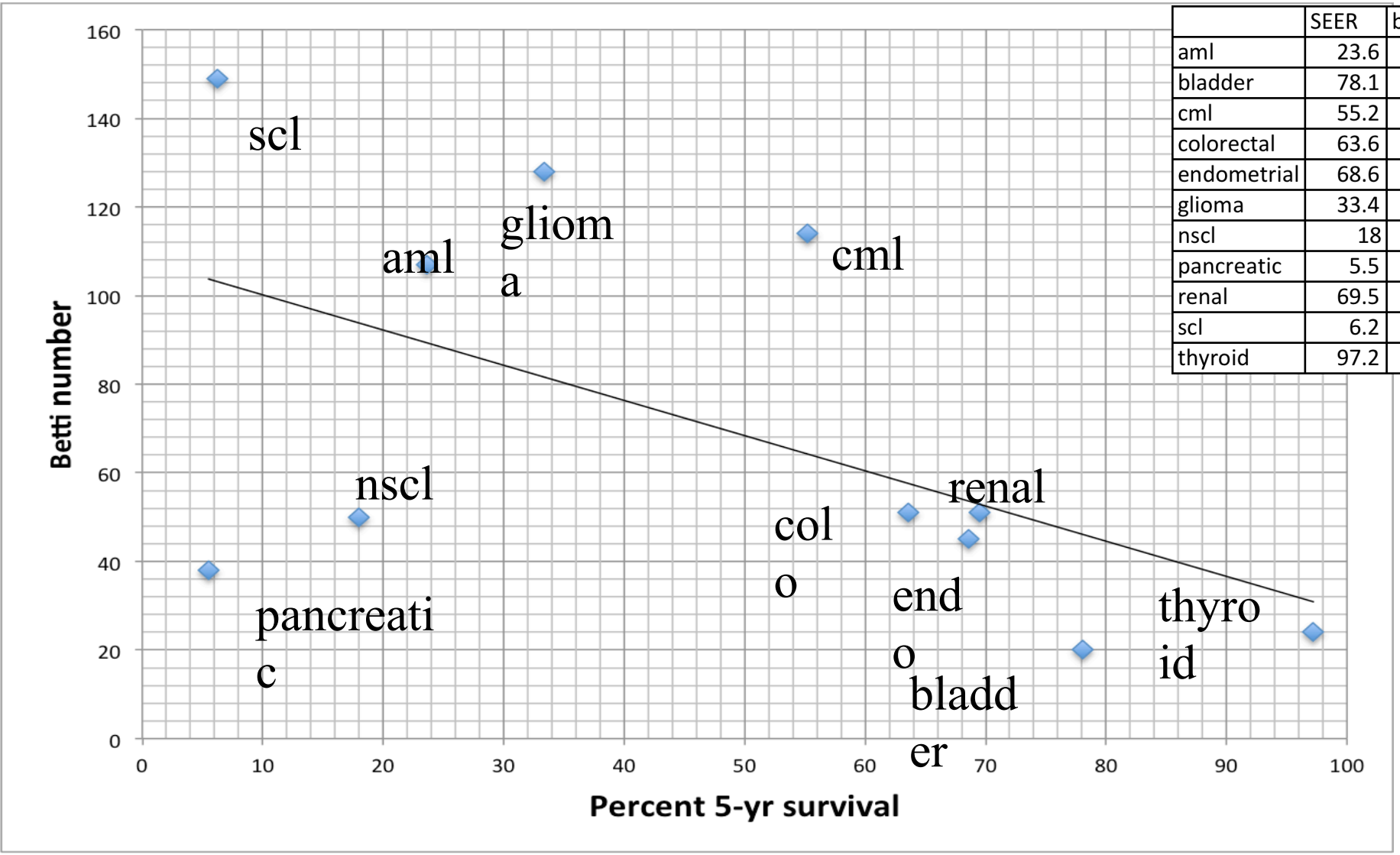
$$\beta_1^{1,2} = 128$$



# Cancer Network Homology
# Mapping to Cancer Patient Survival

R = -0.551

| | SEER | betti |
|---|---|---|
| aml | 23.6 | 107 |
| bladder | 78.1 | 20 |
| cml | 55.2 | 114 |
| colorectal | 63.6 | 51 |
| endometrial | 68.6 | 45 |
| glioma | 33.4 | 128 |
| nscl | 18 | 50 |
| pancreatic | 5.5 | 38 |
| renal | 69.5 | 51 |
| scl | 6.2 | 149 |
| thyroid | 97.2 | 24 |



# Equivalent Targets

| | aml | bladder | cml | colorectal | endometrial | glioma | nscl | pancreatic | renal | scl | thyroid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SEER | 23.6 | 78.1 | 55.2 | 63.6 | 68.6 | 33.4 | 18 | 5.5 | 69.5 | 6.2 | 97.2 |
| nominal betti | 107 | 20 | 114 | 51 | 45 | 128 | 50 | 38 | 51 | 149 | 24 |
| best betti | 95 | 15 | 101 | 41 | 35 | 109 | 37 | 29 | 34 | 131 | 17 |
| equivalent targets | HRAS | MAPK3 | AKT1 | AKT3 | PDPK1 | HRAS | KRAS | KRAS | HIF1A | ITGA3 | HRAS |
| | FLT3 | MAPK1 | AKT2 | AKT2 | ILK | NRAS | | | EPAS1 | ITGA6 | NRAS |
| | NRAS | | AKT3 | AKT1 | | KRAS | | | | ITGA2B | KRAS |
| | KRAS | | | | | | | | | ITGB1 | |
| | | | | | | | | | | ITGA2 | |
| | | | | | | | | | | ITGAV | |

# Suggested Drug Targets

| | aml | bladder | cml | colorectal | endometrial | glioma | nscl | pancreatic | renal | scl | thyroid |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SEER | 23.6 | 78.1 | 55.2 | 63.6 | 68.6 | 33.4 | 18 | 5.5 | 69.5 | 6.2 | 97.2 |
| nominal betti | 107 | 20 | 114 | 51 | 45 | 128 | 50 | 38 | 51 | 149 | 24 |
| suggested drug targets and Betti number result | IKBKB/103 | EGFR/20 | HDAC1/113 | JUN/49 | EGFR/38 | PDGFRB/113 | EGFR/42 | EGFR/38 | JUN/50 | IKBKB/146 | RET/22 |
| | AKT1/96 | RAF1/16 | HDAC2/113 | AKT1/41 | AKT1/41 | PDGFRA/113 | AKT1/47 | IKBKB/38 | AKT1/44 | AKT1/139 | NTRK1/22 |
| | RAF1/103 | MAPK3/15 | IKBKB/111 | RAF1/50 | RAF1/41 | EGFR/115 | RAF1/43 | AKT1/36 | RAF1/50 | NFKB1/136 | MAPK3/23 |
| | FLT3/95 | MAPK1/15 | AKT1/101 | MAPK3/48 | MAPK3/43 | RAF1/124 | MAPK3/48 | RAF1/37 | MAPK3/47 | PTGS2/148 | MAPK1/23 |
| | RARA/105 | ERBB2/20 | ABL1/102 | MAPK1/48 | CCND1/42 | MAPK1/127 | MAPK1/48 | STAT1/33 | MAPK1/47 | CCND1/145 | RXRG/23 |
| | MAPK3/106 | CCND1/18 | RAF1/110 | CCND1/49 | MAPK1/43 | MAPK3/127 | ERBB2/42 | MAPK3/38 | BRAF/50 | | RXRA/23 |
| | NFKB1/105 | BRAF/16 | MAPK1/113 | BRAF/50 | BRAF/41 | AKT1/128 | CCND1/47 | MAPK1/38 | | | RXRB/23 |
| | MAPK1/106 | | MAPK3/113 | BCL2/51 | | CCND1/126 | BRAF/43 | NFKB1/30 | | | CCND1/21 |
| | CCND1/107 | | NFKB1/114 | | | BRAF/124 | | ERBB2/38 | | | BRAF/21 |
| | KIT/96 | | CCND1/112 | | | | | CCND1/38 | | | |
| | STAT5B/103 | | STAT5B/112 | | | | | BRAF/37 | | | |
| | BRAF/103 | | BRAF/110 | | | | | | | | |

# Double Inhibition Allows for Further Reduction in Betti Number and thus Complexity

| | AML | Bladder | CML | Colorectal | Endometrial |
|---|---|---|---|---|---|
| SEER | 23.6 | 78.1 | 55.2 | 63.6 | 68.6 |
| Nominal Betti | 107 | 20 | 114 | 51 | 45 |
| Min single | 95 | 15 | 101 | 41 | 35 |
| Min double | 83 | 11 | 88 | 31 | 26 |
| Double targets | HRAS FLT3 | HRAS MAPK3 | AKT1 AKT2 | AKT1 AKT2 | ILK PDPK1 |
| | HRAS NRAS | ARAF MAPK3 | AKT1 AKT3 | AKT1 AKT3 | |
| | FLT3 NRAS | RAF1 MAPK3 | AKT2 AKT3 | AKT2 AKT3 | |
| | HRAS KRAS | NRAS MAPK3 | | | |
| | FLT3 KRAS | HRAS MAPK1 | | | |
| | NRAS KRAS | ARAF MAPK1 | | | |
| | | RAF1 MAPK1 | | | |
| | | NRAS MAPK1 | | | |
| | | MAPK3 KRAS | | | |
| | | MAPK1 KRAS | | | |
| | | MAPK3 BRAF | | | |
| | | MAPK1 BRAF | | | |

# Double Inhibition Allows for Further Reduction in Betti Number and thus Complexity (continued)

| | Glioma | NSCL | Pancreatic | Renal | SCL | Thyroid |
|---|---|---|---|---|---|---|
| SEER | 33.4 | 18 | 5.5 | 69.5 | 6.2 | 97.2 |
| Nominal Betti | 128 | 50 | 38 | 51 | 149 | 24 |
| Min single | 109 | 37 | 29 | 34 | 131 | 17 |
| Min double | 90 | 29 | 21 | 26 | 113 | 10 |
| Double targets | NRAS HRAS | EGFR KRAS | NFKB1 KRAS | HIF1A GAB1 | ITGA2 ITGA3 | HRAS NRAS |
| | NRAS KRAS | ERBB2 KRAS | RELA KRAS | GAB1 EPAS1 | ITGA2 ITGA6 | HRAS KRAS |
| | HRAS KRAS | | | | ITGA3 ITGA6 | NRAS KRAS |
| | | | | | ITGA2 ITGA2B | |
| | | | | | ITGA3 ITGA2B | |
| | | | | | ITGA6 ITGA2B | |
| | | | | | ITGA2 ITGB1 | |
| | | | | | ITGA3 ITGB1 | |
| | | | | | ITGA6 ITGB1 | |
| | | | | | ITGA2B ITGB1 | |
| | | | | | ITGA2 ITGAV | |
| | | | | | ITGA3 ITGAV | |

# Betti suggests targets for AML

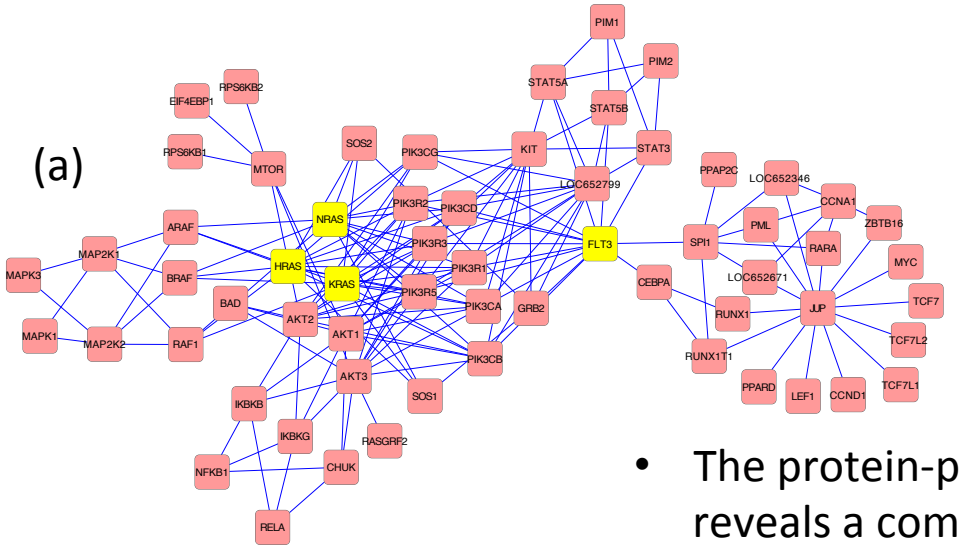
| | AML |
|---|---|
| SEER | 23.6 |
| nominal betti | 107 |
| Target 1 | HRAS/95 |
| Target 2 | FLT3/95 |
| Target 3 | NRAS/95 |
| Target 4 | KRAS/95 |
| | |
| drug targets | IKBKB/103 |
| drug targets | AKT1/96 |
| drug targets | RAF1/103 |
| drug targets | FLT3/95 |
| drug targets | RARA/105 |
| drug targets | MAPK3/106 |
| drug targets | NFKB1/105 |
| drug targets | MAPK1/106 |
| drug targets | CCND1/107 |
| drug targets | KIT/96 |
| drug targets | STAT5B/103 |
| drug targets | BRAF/103 |
| | |
| double knockouts | HRAS ,FLT3/83 |
| double knockouts | NRAS, FLT3/83 |
| double knockouts | KRAS, FLT3/83 |

Betti suggestions

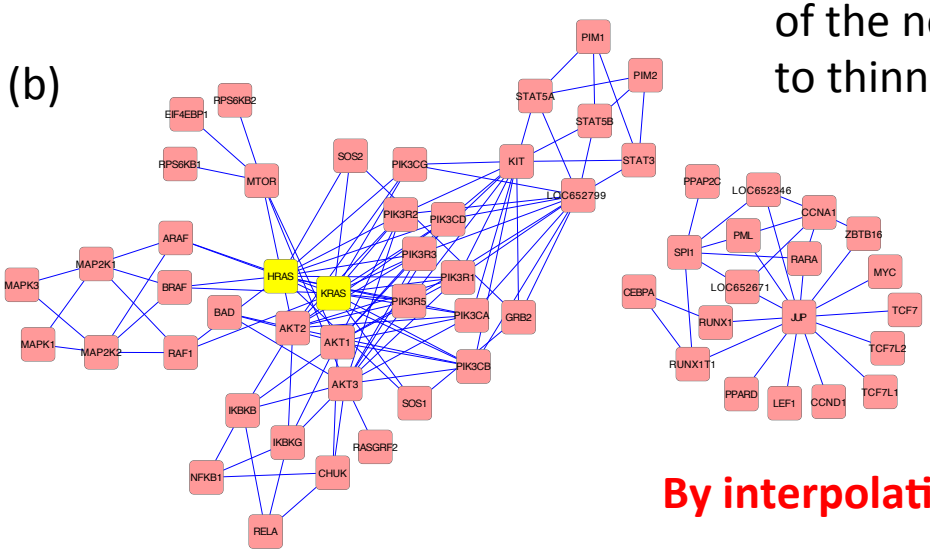
Existing drugs for this cancer

Drug combination, via Betti

# Example: Targeting the AML network




- The protein-protein interaction network (PPI) for AML reveals a complex single component network (Panel A)
- The network breaks into two separate components by targeted elimination of the connecting FLT3 protein (Panel B).
- The effect of simultaneous elimination of NRAS and FLT3 (shown in Panel B) leads not only to break down of the network into two separate components, but also to thinning of the main network.



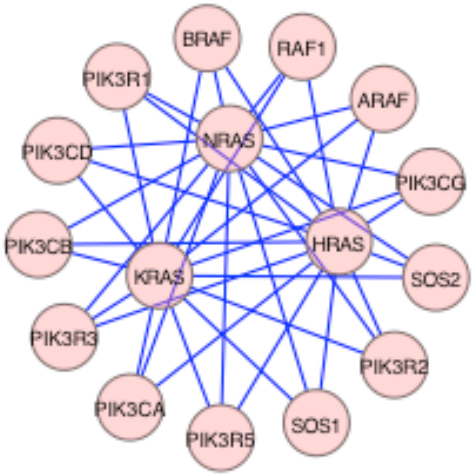

$$\beta_1^{1,2} = 107$$

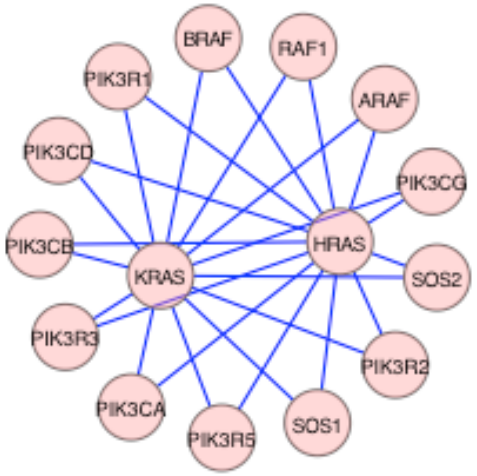
$$\Delta\beta_1^{1,2} = 83$$

**By interpolation …. About 7% improvement in survival.**

# AML Network Breakup by Targeting

(a)




(b)

- Analysis of the RAS family of proteins (Panel A) indicates that each of respective proteins (HRAS, KRAS and NRAS) is equivalent in importance and interconnected with similar neighbors.
- Thus the elimination any one of the RAS proteins leads a thinning of the major component (as shown in the previous slide).

# Persistent-Homology Summary

- Persistent-homology, i.e. Betti number quantifies complexity of a cancer protein-protein interaction network
- Betti number correlates with survival
- Sensitive to protein inhibition
- Identifies already known targets for a given cancer.
- Could suggest use of existing drugs approved for a given cancer in a different one, as well as new drugs development
- Drug combination
- Could yield a powerful numerical tool for clinical use. Rationalize which targeted therapy to try first

- We are currently undertaking a retrospective clinical study of pediatric patients who have had biological-agent therapies.