Protein Interaction Networks and their use in Drug Discovery

Matteo Pellegrini
Protein Pathways Inc.

Outline

❑Introduction

❑Methods to infer protein couplings

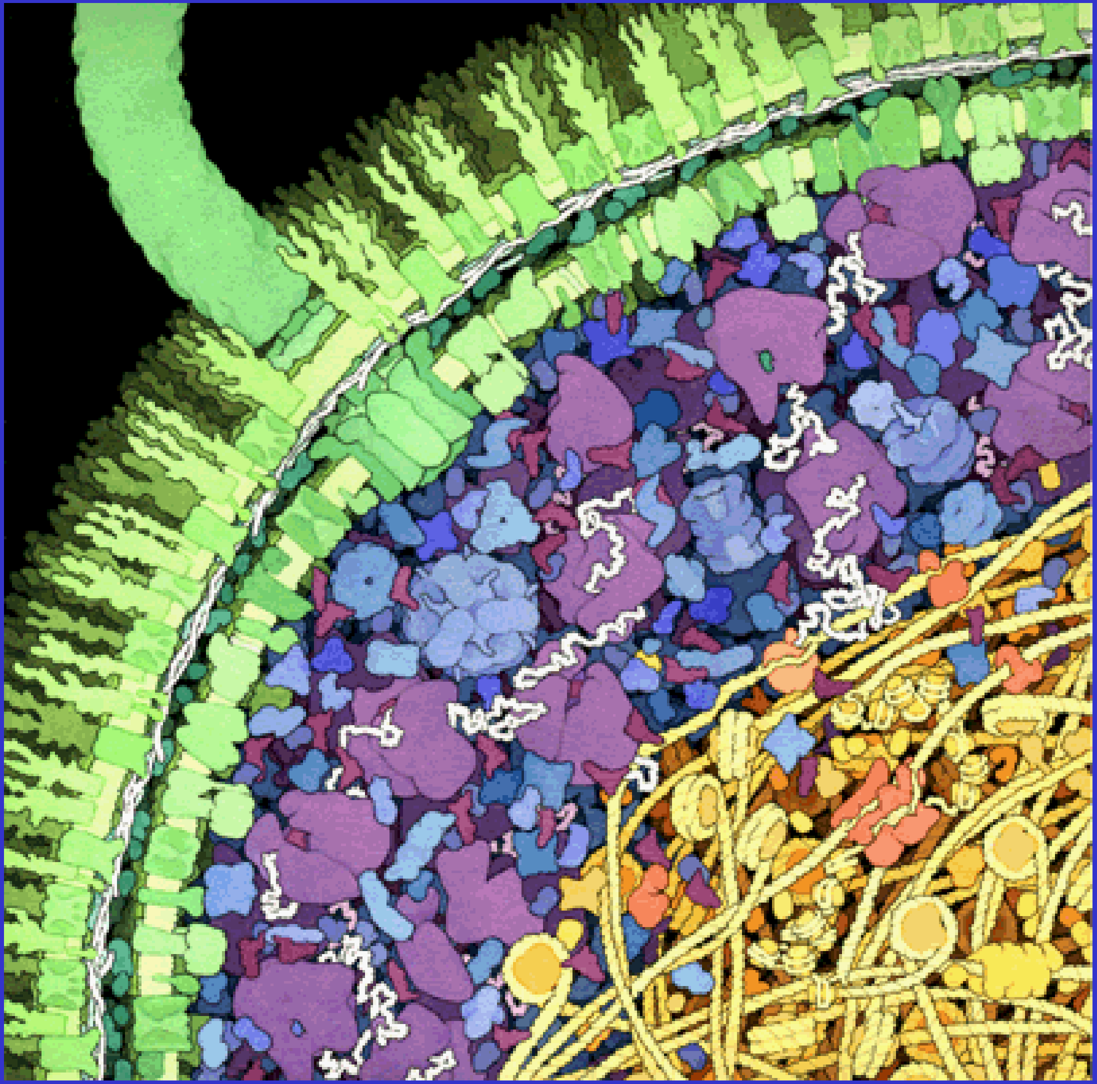
- Direct detection of protein-protein interactions
- Analysis of co-evolution
- Analysis of expression data

❑Properties of protein networks

❑Applications of protein networks

- Identifying genes involved in osteoporosis

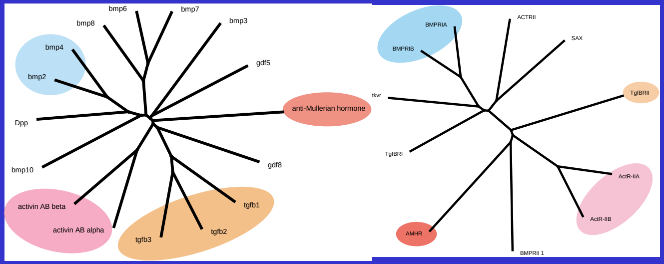
Cells Contain High Concentrations of Proteins that Participate in a Multitude of Interactions



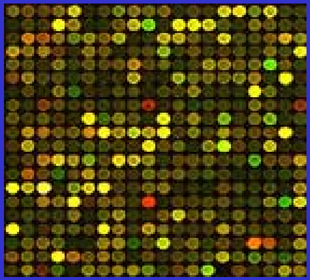
Escherichia coli drawn to molecular scale by David Goodsell

Qualitative Protein Networks May be Reconstructed From Varied Data

Direct measurements of protein interactions



Analysis of co-evolving genes



Analysis of expression microarrays

Qualitative Protein Networks May be Reconstructed From Varied Data

Direct measurements of protein interactions

Experimental Methods for Detection of Protein-Protein Physical Interactions

Physical Interactions:

- Two hybrid
- Co-purification
- Protein Fragment Complementarity Assays
- Protein Chips

Database of Interacting Proteins (DIP)

Number of proteins	6963
Number of organisms	113
Number of interactions	18059
Number of unique experiments	21722
Number of articles	2110

Qualitative Protein Networks May be Reconstructed From Varied Data

Analysis of co-evolving genes

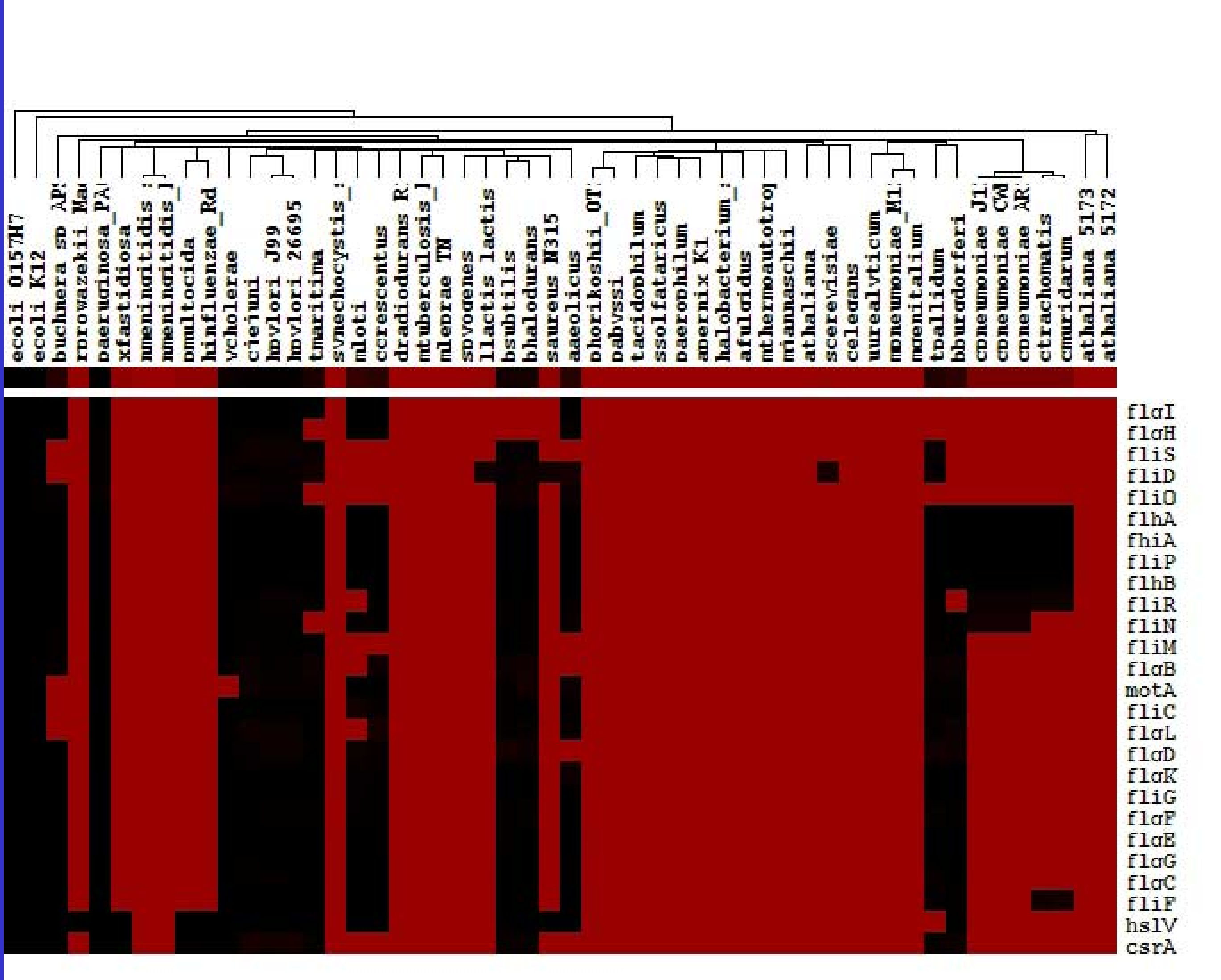
Co-Evolving Genes

Despite extensive genomic rearrangements that occur during evolution, certain pairs of genes are preserved

- together
- nearby
- fused

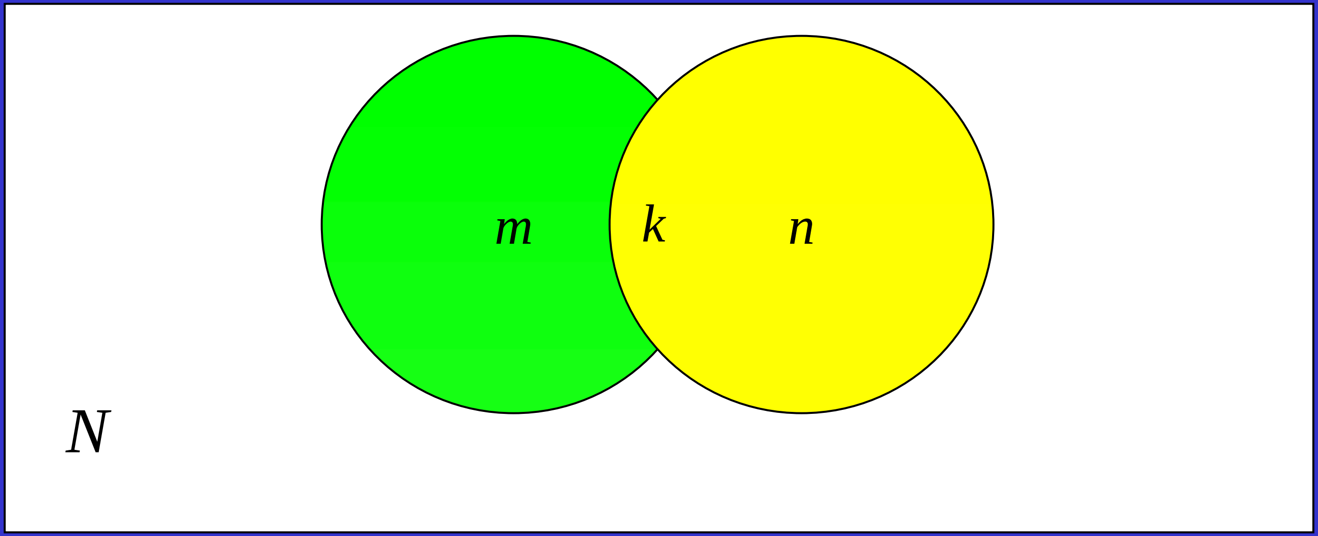
in many genomes, and are thus considered to co-evolve

Example of Co-evolving Genes: Phylogenetic Profiles*



*Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*. 96(8):4285-8,. 1999

What is the Probability that two Genes Co-evolve?



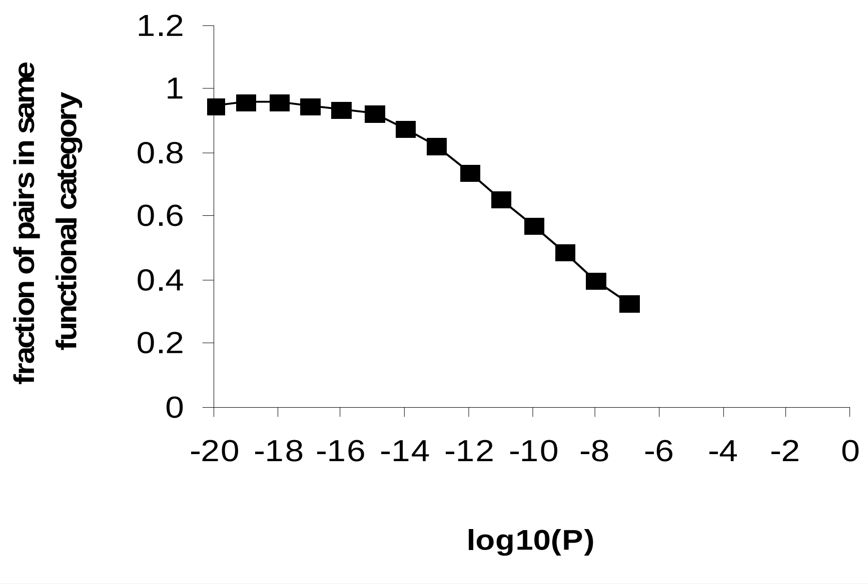
$$P(k \mid n, m, N) = \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$$

Co-evolving proteins are often found in the Same Functional Category

Information Storage and Processing
Translation, ribosomal structure and biogenesis
Transcription
DNA replication, recombination and repair

Cellular processes
Cell division and chromosome partitioning
Posttranslational modification, protein turnover, chaperones
Cell envelope biogenesis, outer membrane
Cell motility and secretion
Inorganic ion transport and metabolism
Signal transduction mechanisms

Metabolism
Energy production and conversion
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Nucleotide transport and metabolism
Coenzyme metabolism
Lipid metabolism
Secondary metabolites biosynthesis, transport and catabolism

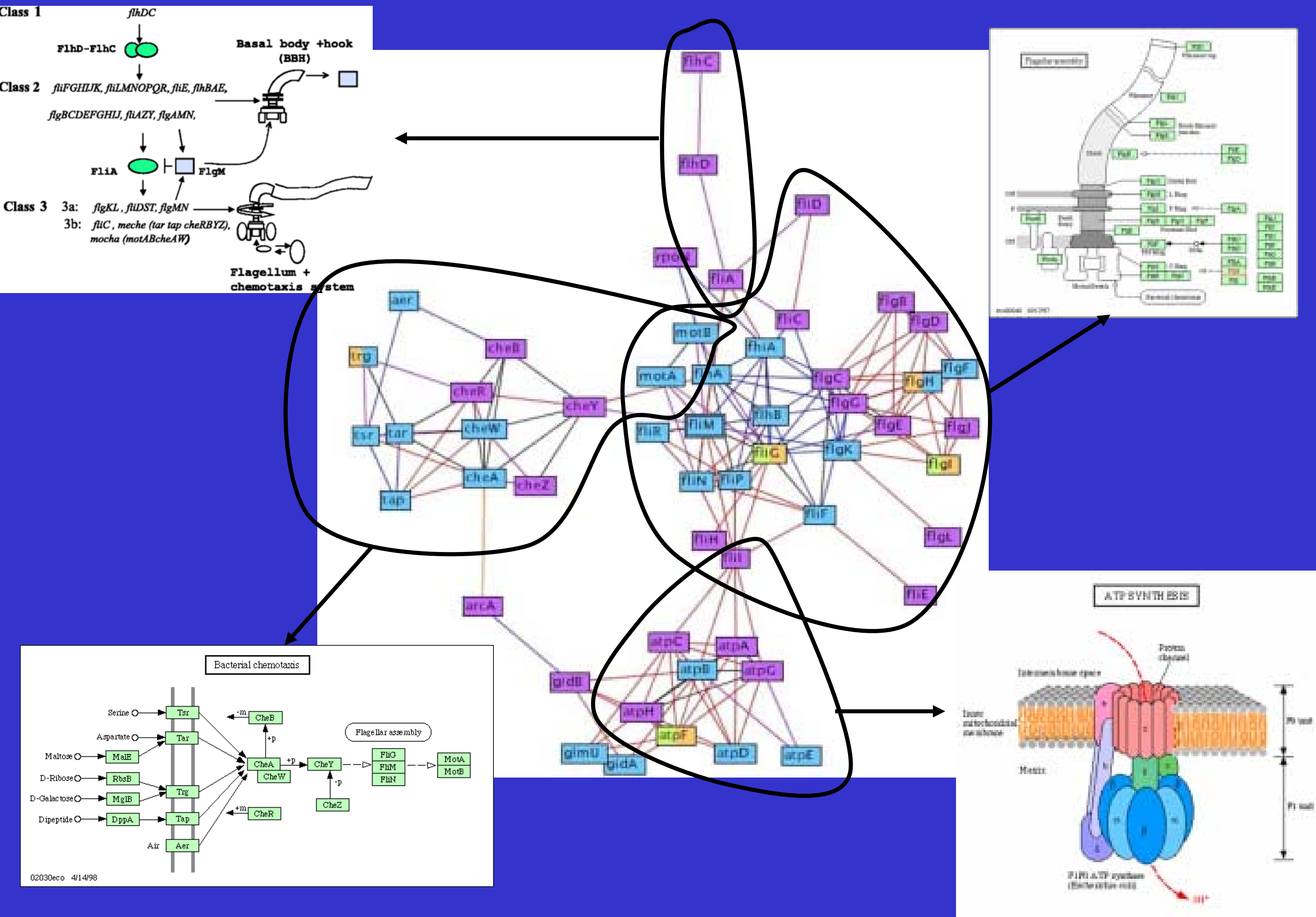


Inferring Co-evolution Using Additional Methods

Method	Description
Gene fusion*	Pairs of genes that are fused in some organism
Gene Neighbor	Pairs of genes that are coded nearby in multiple organisms
Operon reconstruction	Reconstruction of bacterial operons based on intergenic spacing

*Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. (1999). Detecting protein function and protein-protein interactions from genome sequences. Science, 285, 751-753.

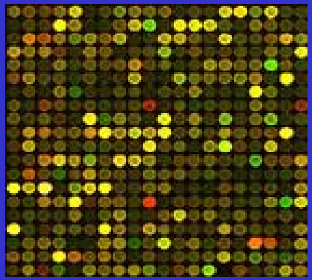
Bacterial Co-evolution Network Using All Four Methods



Sources of Flagella Protein Interactions

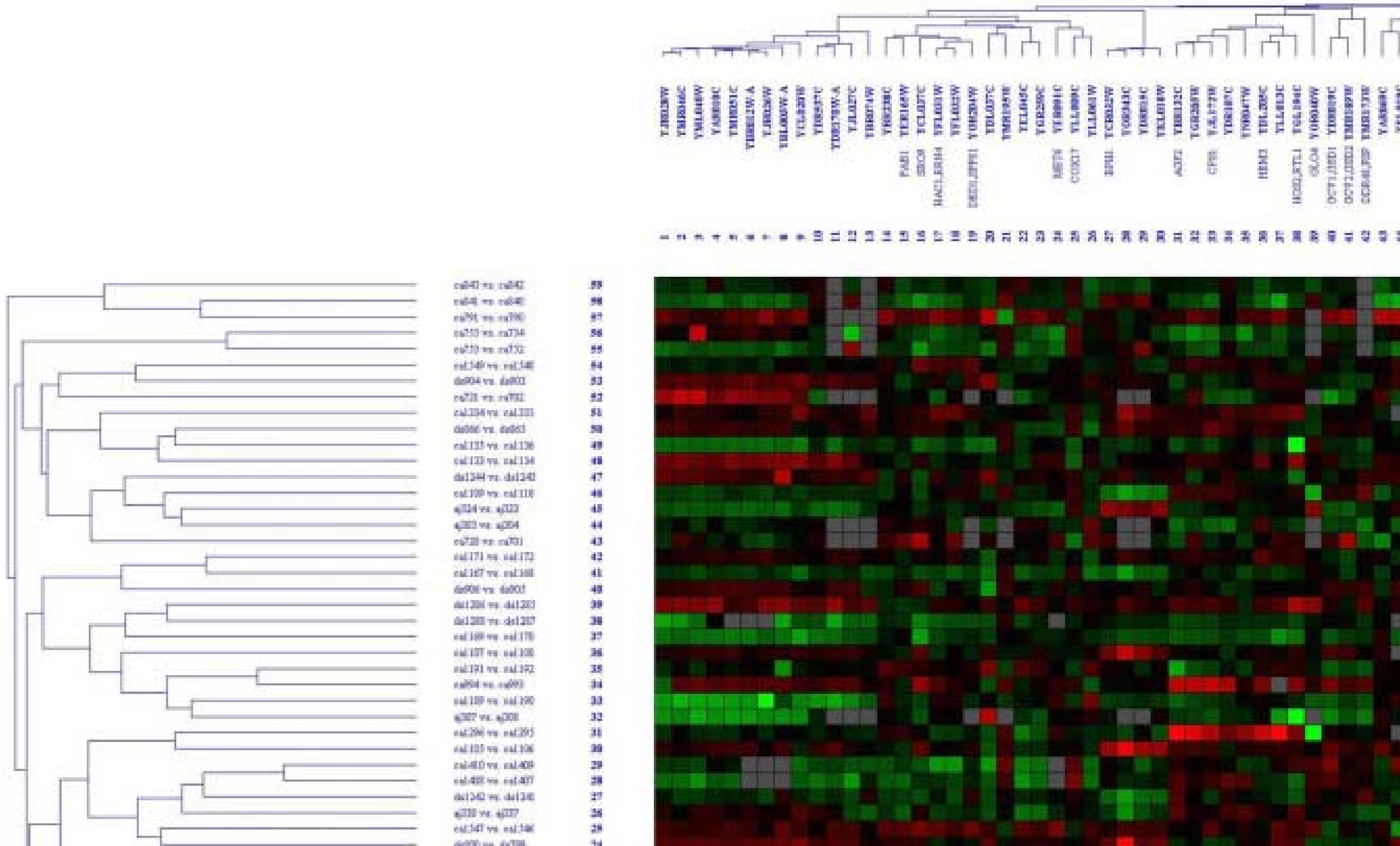
Phylogenetic profile	29
Gene Neighbor	85
Rosetta Stone	1
From DIP	18
Multiple	34

Qualitative Protein Networks May be Reconstructed From Varied Data



Analysis of expression microarrays

Measuring Correlations in Expression Levels Reveals Co-transcribed Genes



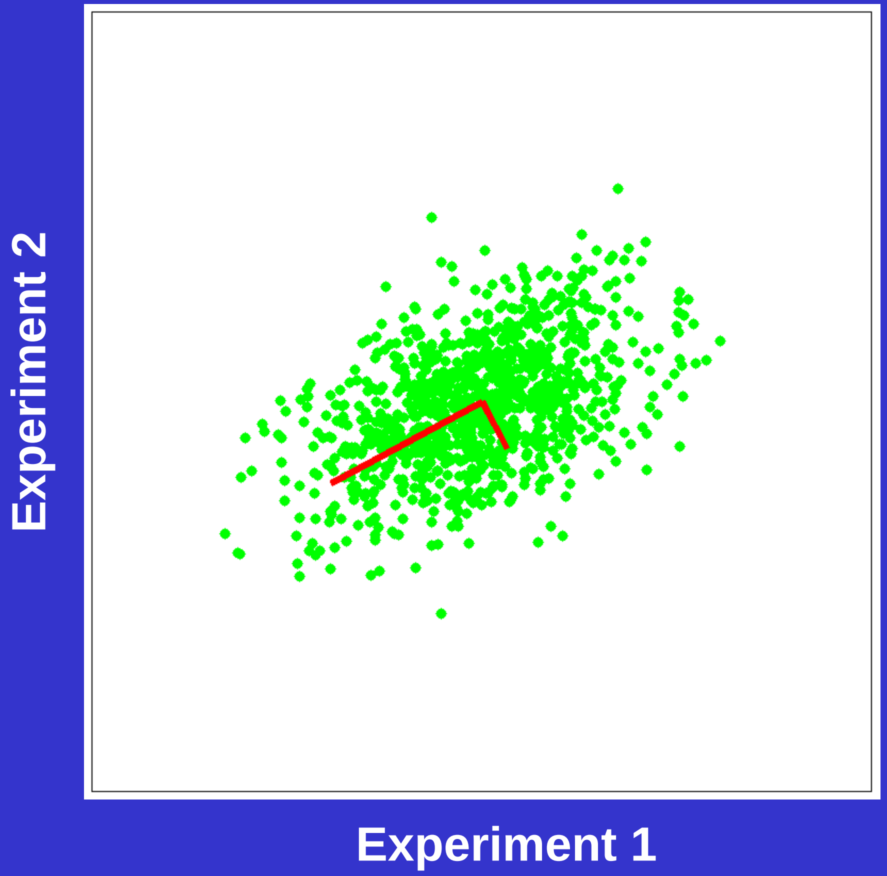
Degree of Correlation

The degree of correlation between two genes is computed by the Pearson correlation coefficient:

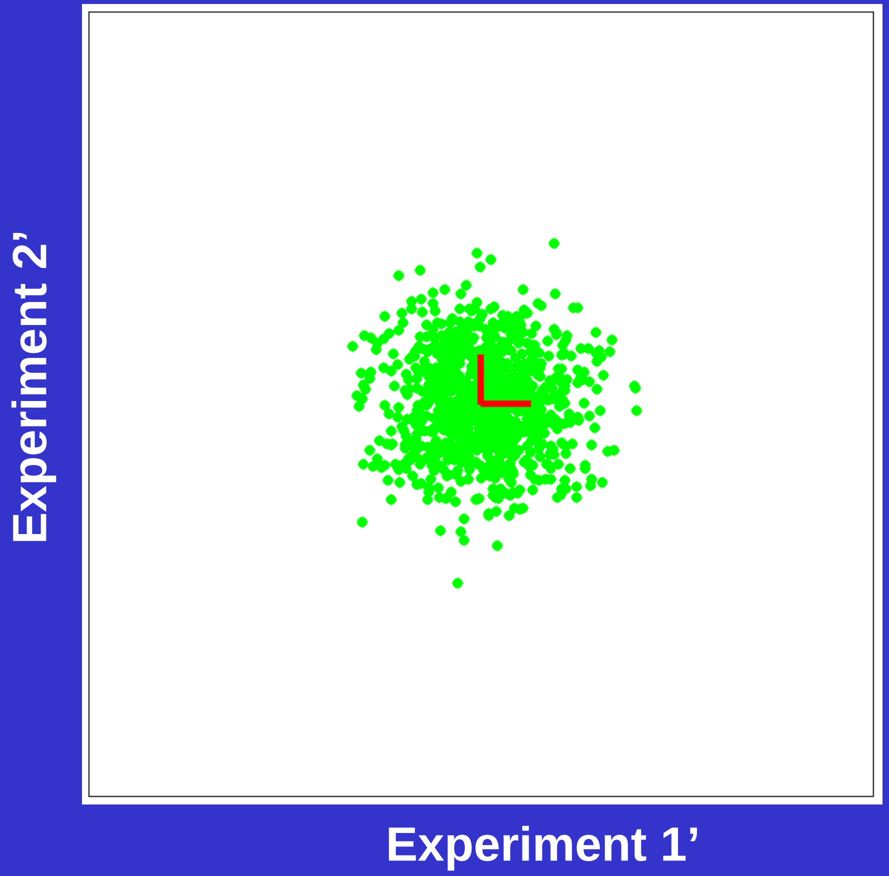
$$C_{ij} = \frac{\sum_n (x_i^n - \bar{x}_i)(x_j^n - \bar{x}_j)}{\sqrt{\sum_n (x_i^n - \bar{x}_i)^2 \sum_n (x_j^n - \bar{x}_j)^2}}$$

Improving the Measurement of Expression Correlations

Correlated Experiments

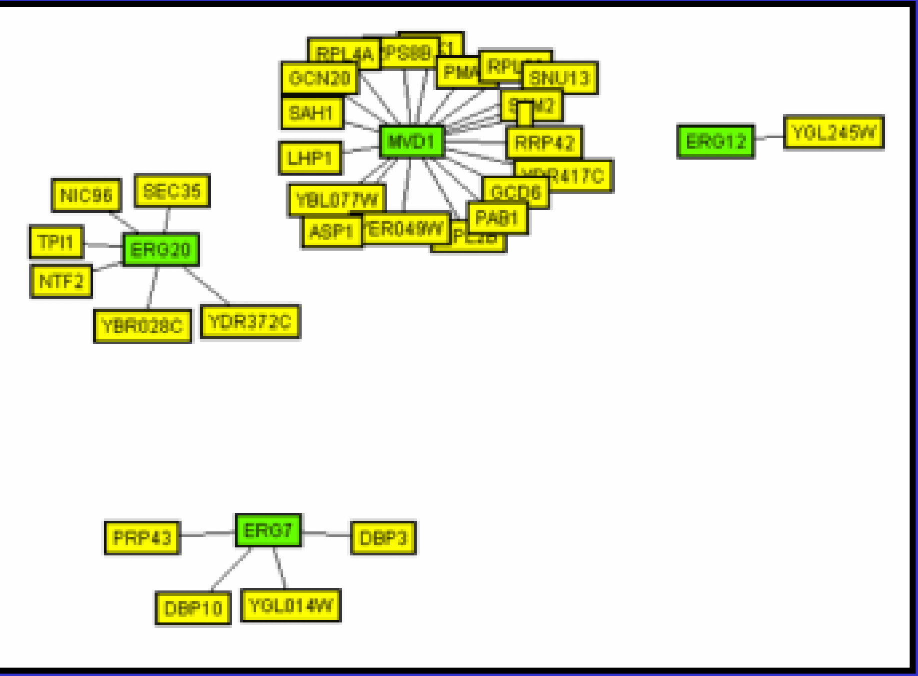


Uncorrelated Experiments

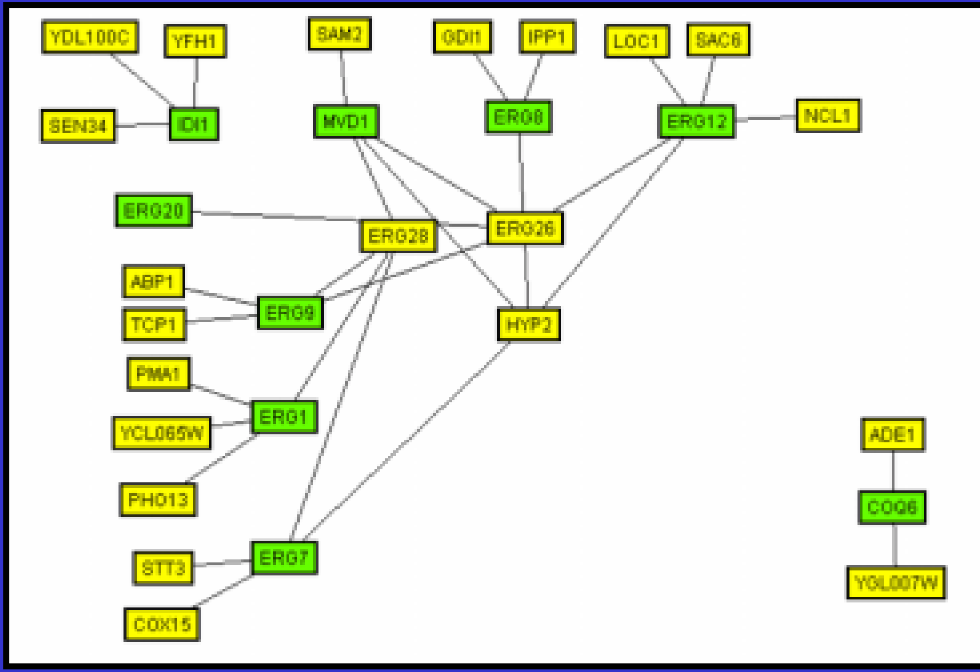


Ergosterol Pathway Reconstruction is Enhanced by Decorrelating Expression Data

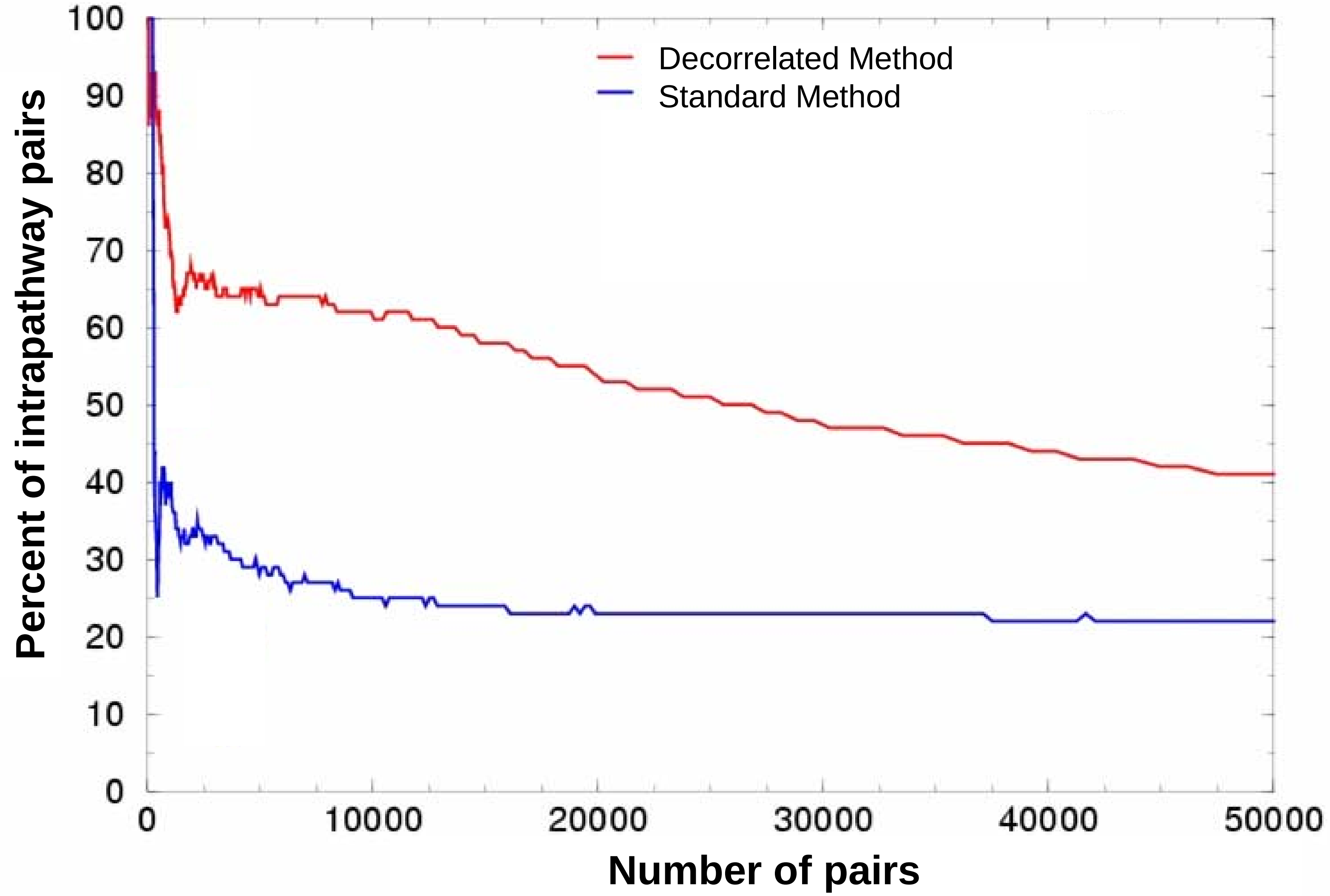
Correlated Experiments



Uncorrelated Experiments

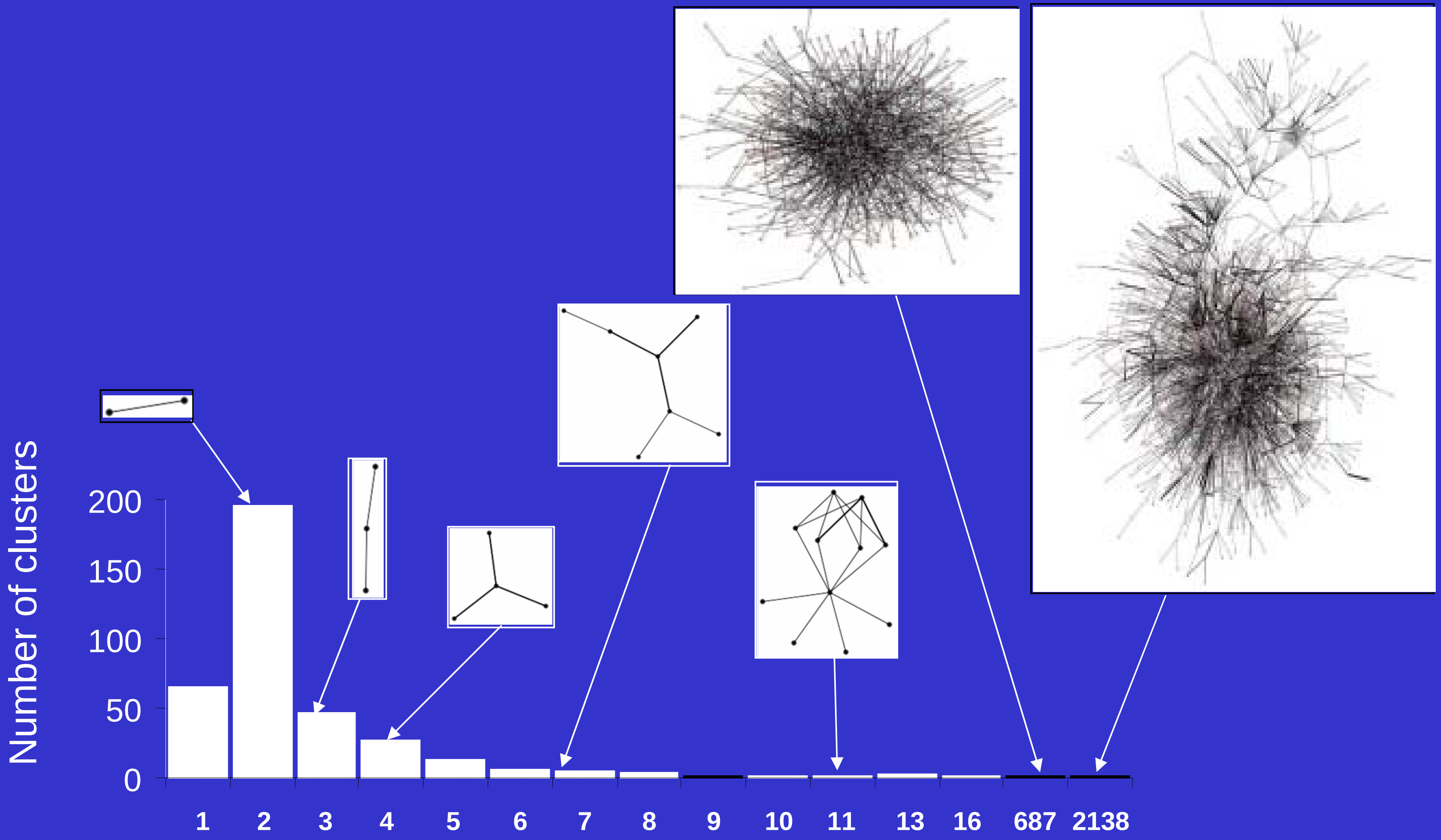


Decorrelation Method Reveals Additional Pathway Relationships



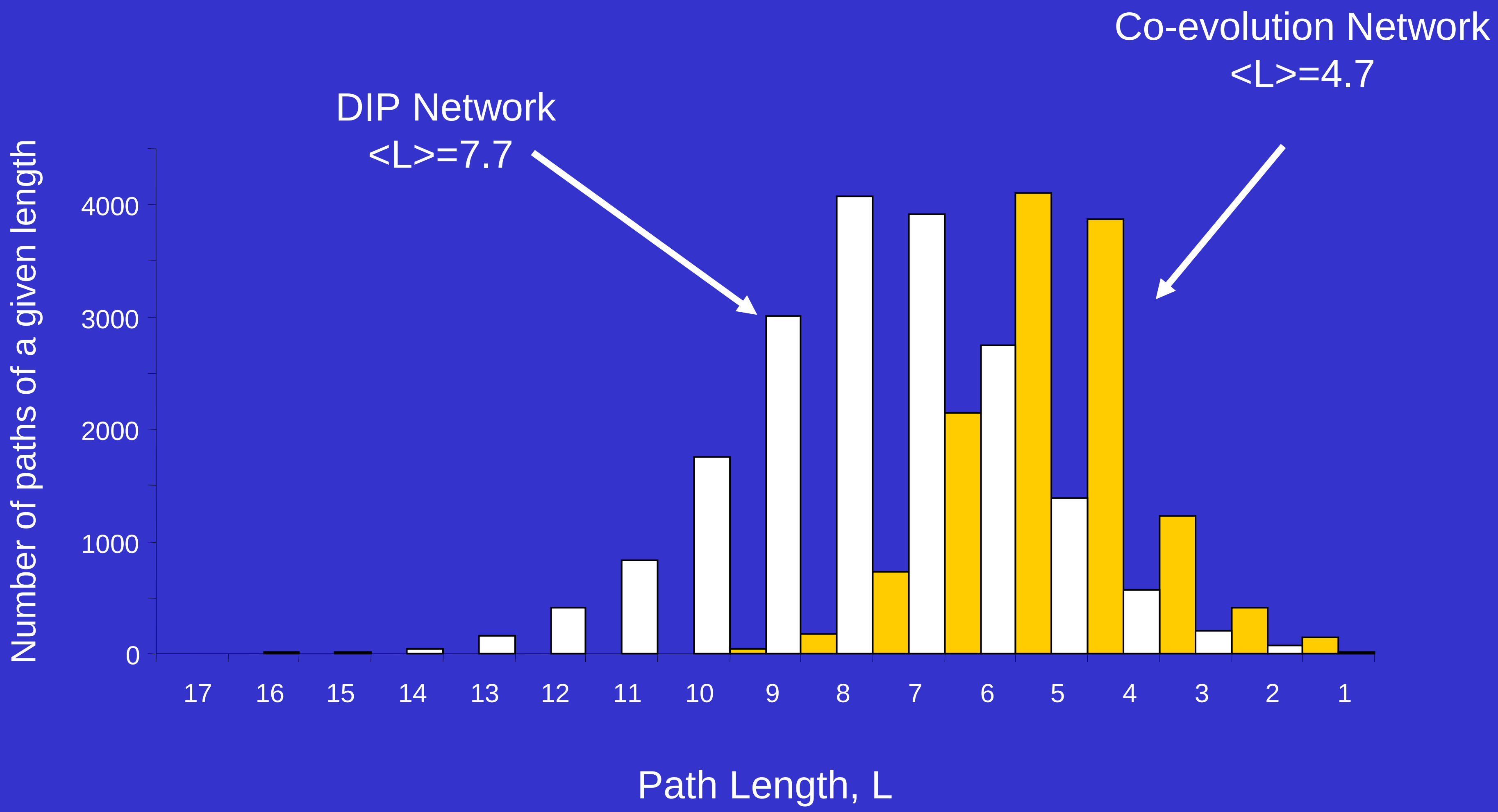
Properties of Protein Networks

DIP Contains 11,000 Interactions Within 370 Networks

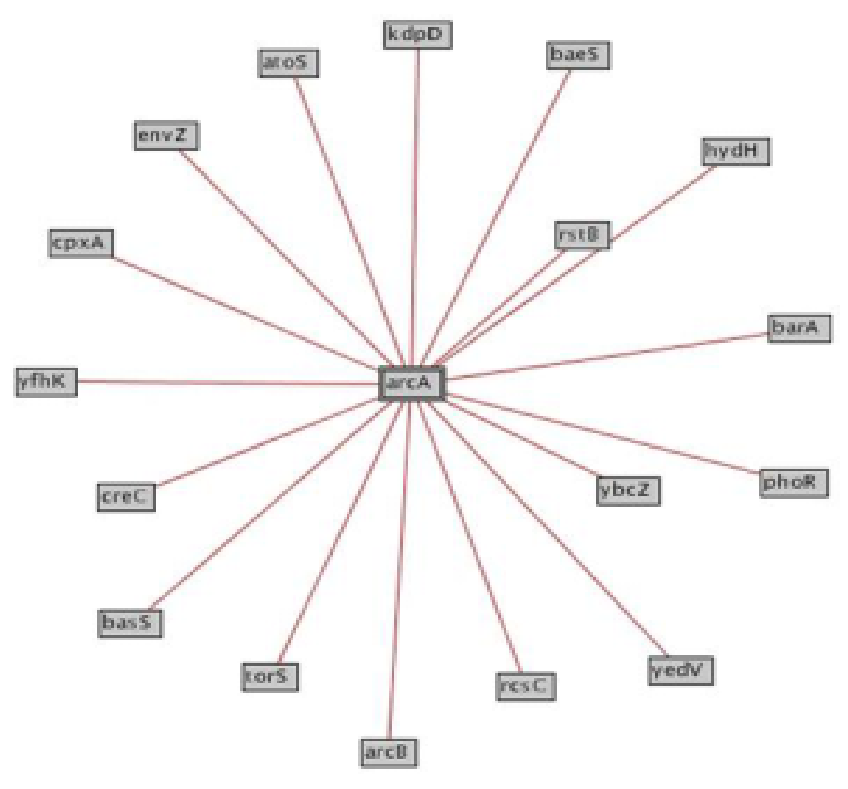


Number of proteins per network

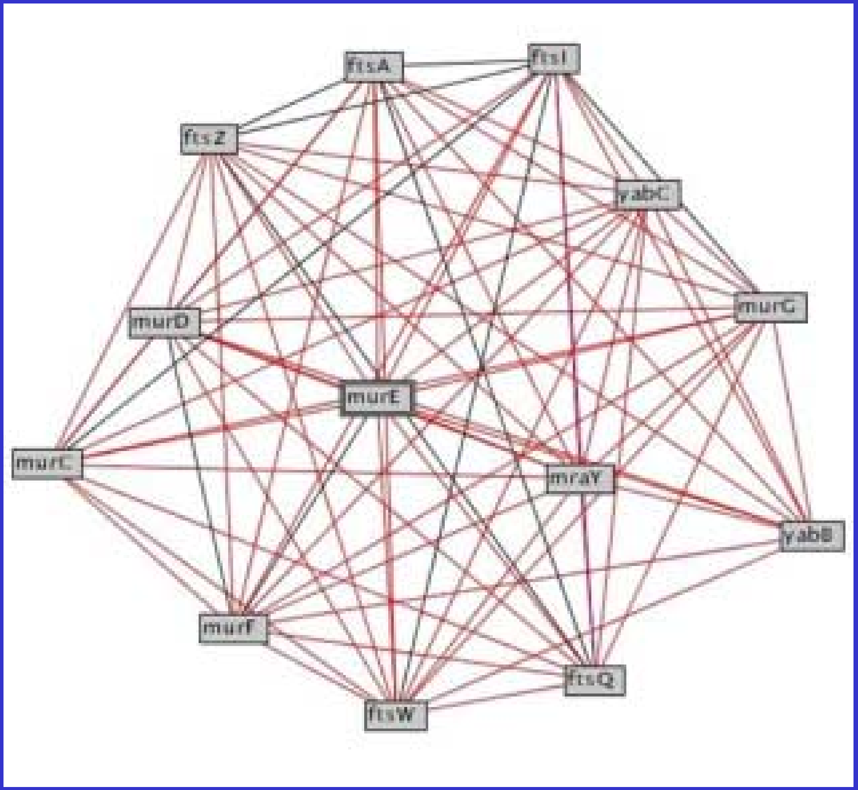
Path Lengths of Linked Proteins



Protein Networks Have High Local Connectivity

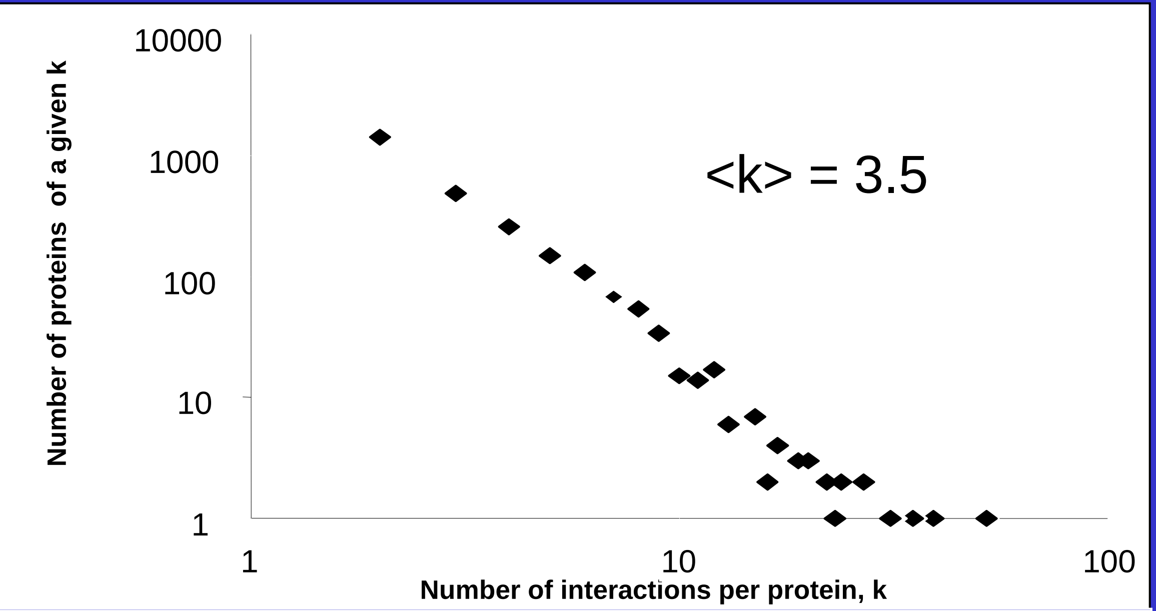


Low Connectivity

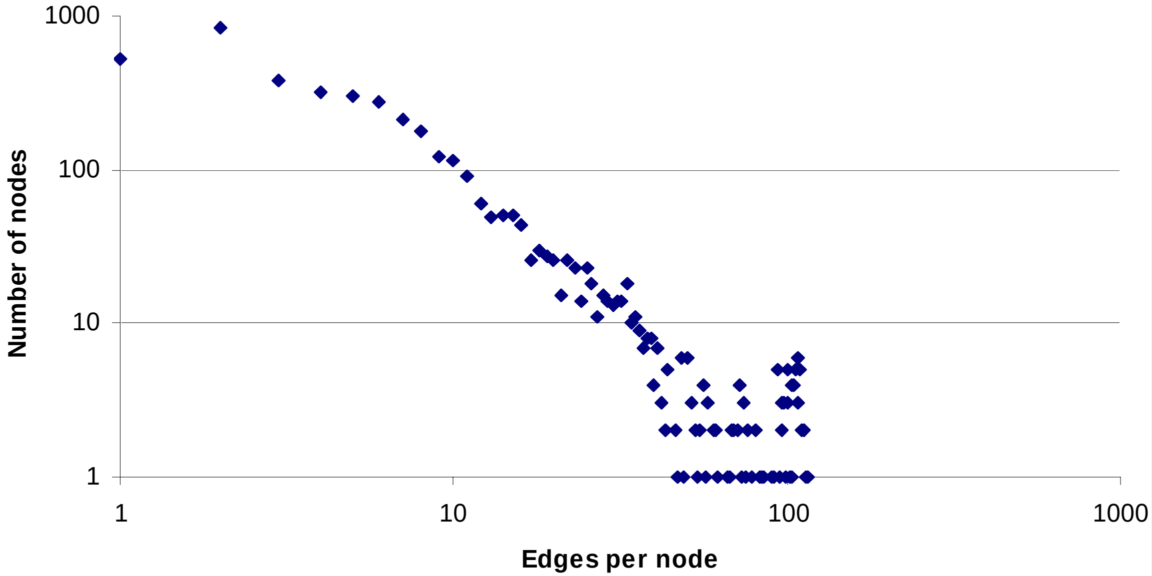


High Connectivity

Networks are Scale-Invariant

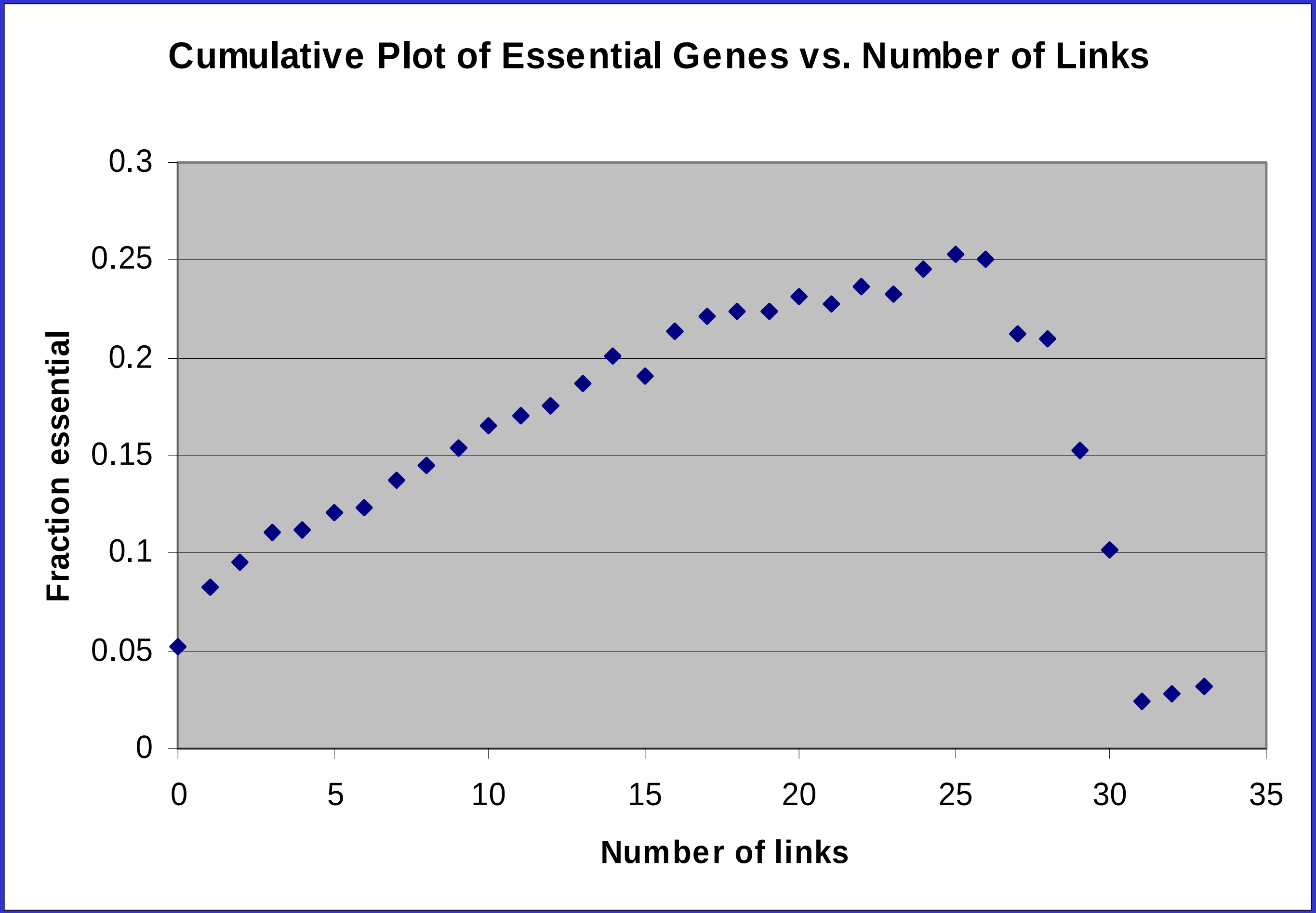


DIP



Co-evolution

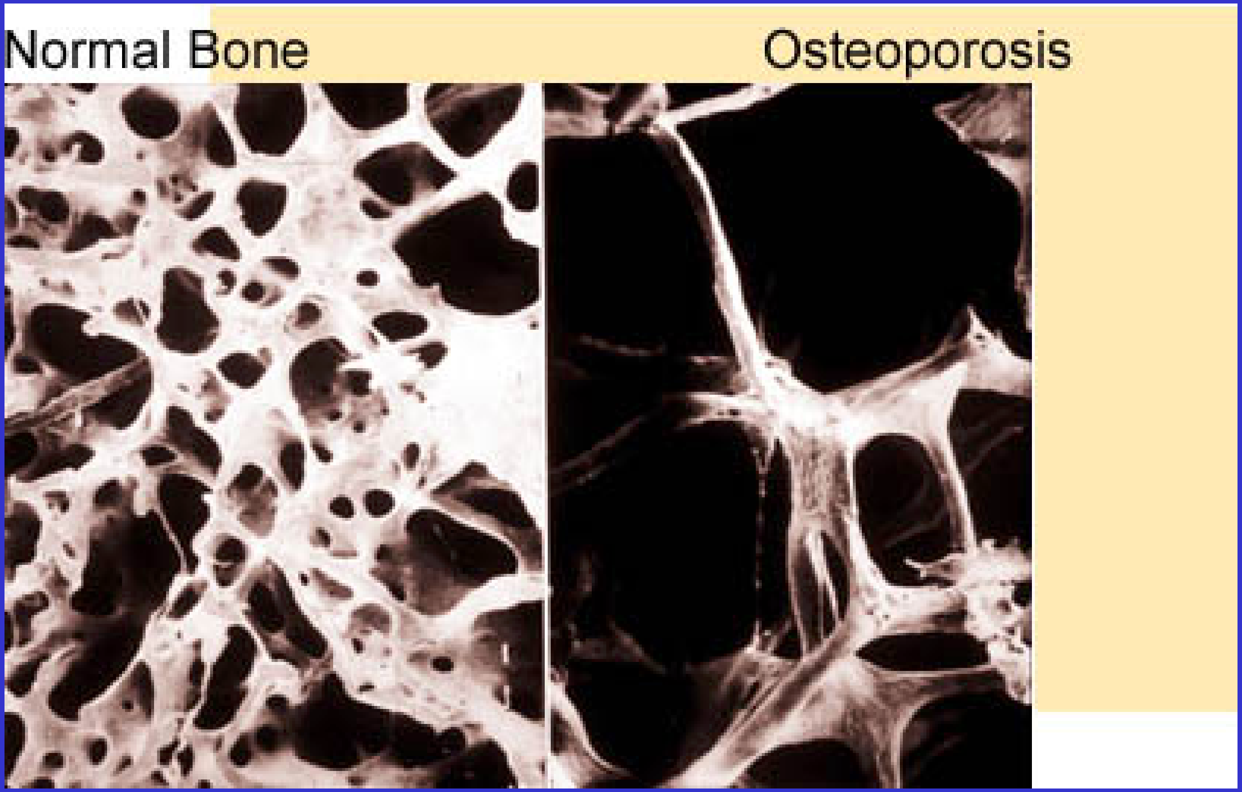
Network Hubs Tend to be Essential



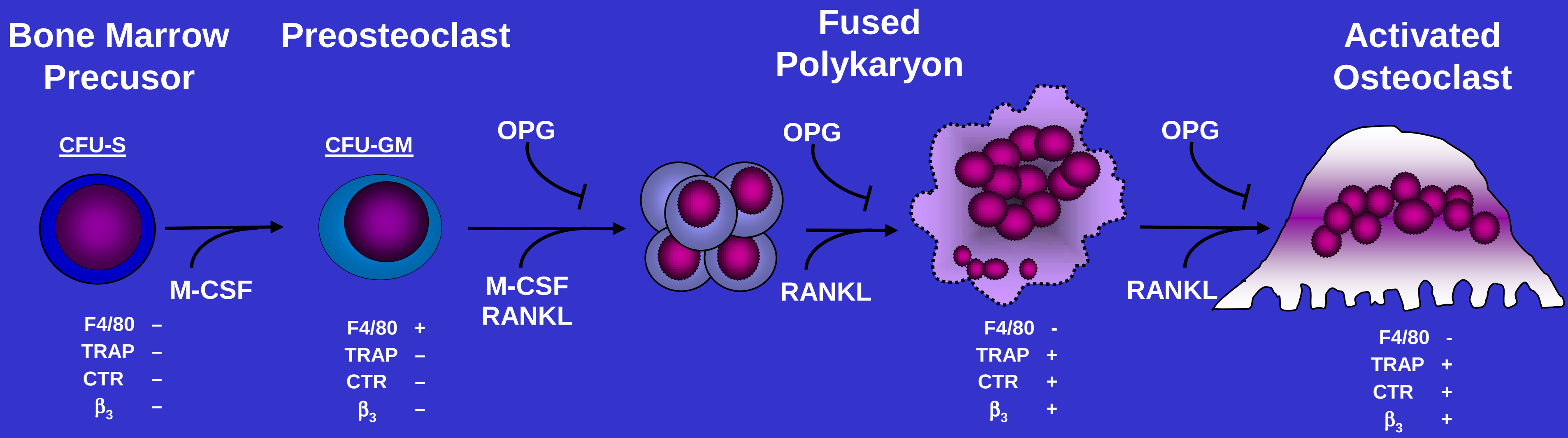
Treating Osteoporosis by Discovering Genes Involved in Osteoclast Differentiation

Osteoporosis

Osteoporosis is a major public health threat for an estimated 44 million Americans



Osteoporosis Cell Based Assay



Bone Marrow Precusor

CFU-S

M-CSF

F4/80 –
TRAP –
CTR –
β_3 –

Preosteoclast

CFU-GM

F4/80 +
TRAP –
CTR –
β_3 –

OPG

M-CSF
RANKL

Fused Polykaryon

OPG

RANKL

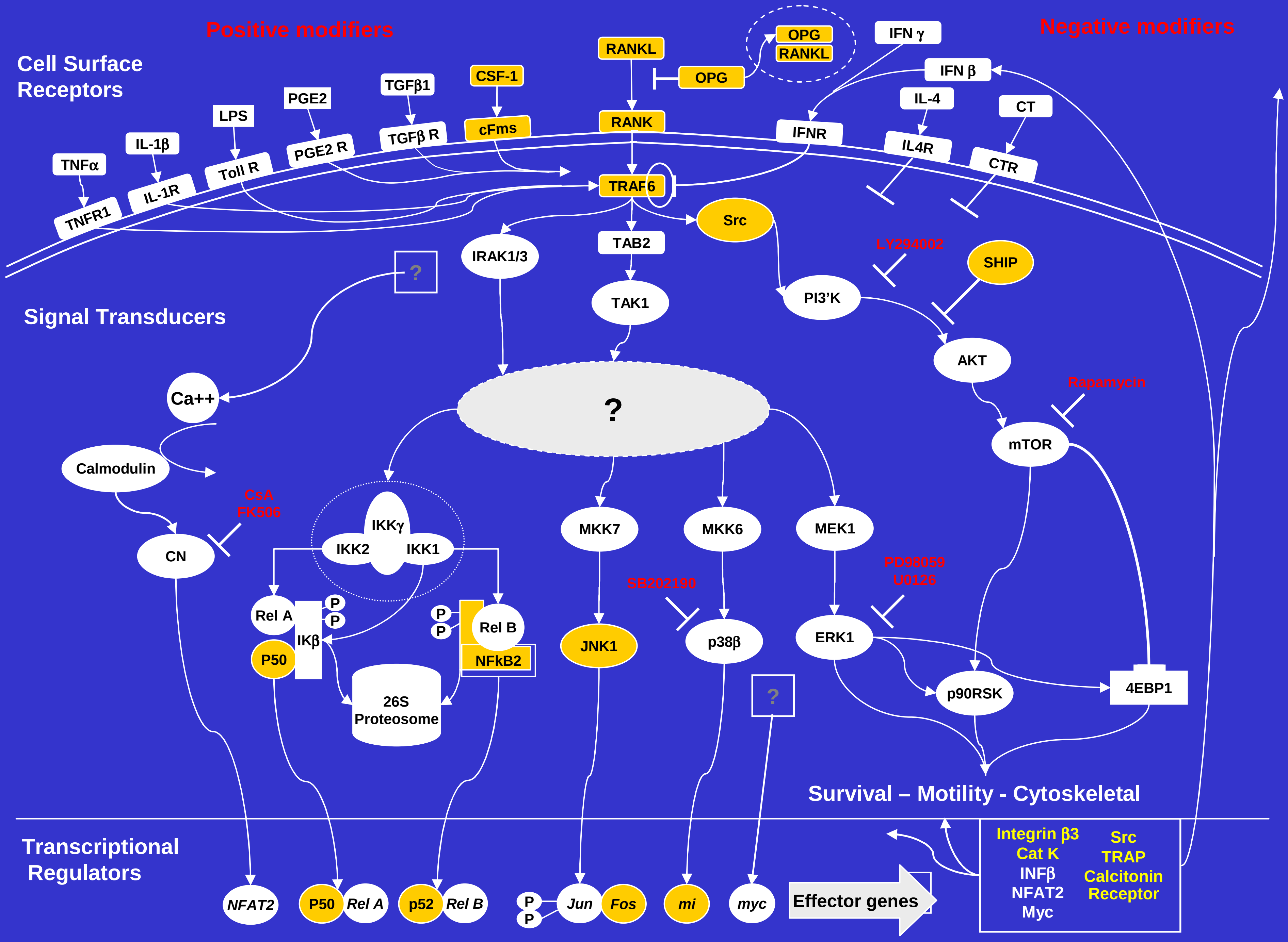
F4/80 -
TRAP +
CTR +
β_3 +

Activated Osteoclast

OPG

RANKL

F4/80 -
TRAP +
CTR +
β_3 +



Positive modifiers

Negative modifiers

Cell Surface Receptors

TNFα
TNFR1
IL-1β
IL-1R
LPS
Toll R
PGE2
PGE2 R
TGFβ1
TGFβ R
CSF-1
cFms
RANKL
RANK
OPG
OPG
RANKL
IFN γ
IFN β
IFNR
IL-4
IL4R
CT
CTR
TRAF6
Src

Signal Transducers

?
IRAK1/3
TAB2
TAK1
PI3’K
LY294002
SHIP
AKT
Rapamycin
mTOR
Ca++
Calmodulin
CN
CsA
FK506
?
IKKγ
IKK2
IKK1
Rel A
P
P
IKβ
P50
P
P
Rel B
NFkB2
26S Proteosome
MKK7
MKK6
MEK1
SB202190
PD98059
U0126
JNK1
p38β
ERK1
?
p90RSK
4EBP1

Survival – Motility - Cytoskeletal

Transcriptional Regulators

NFAT2
P50
Rel A
p52
Rel B
P
P
Jun
Fos
mi
myc

Effector genes

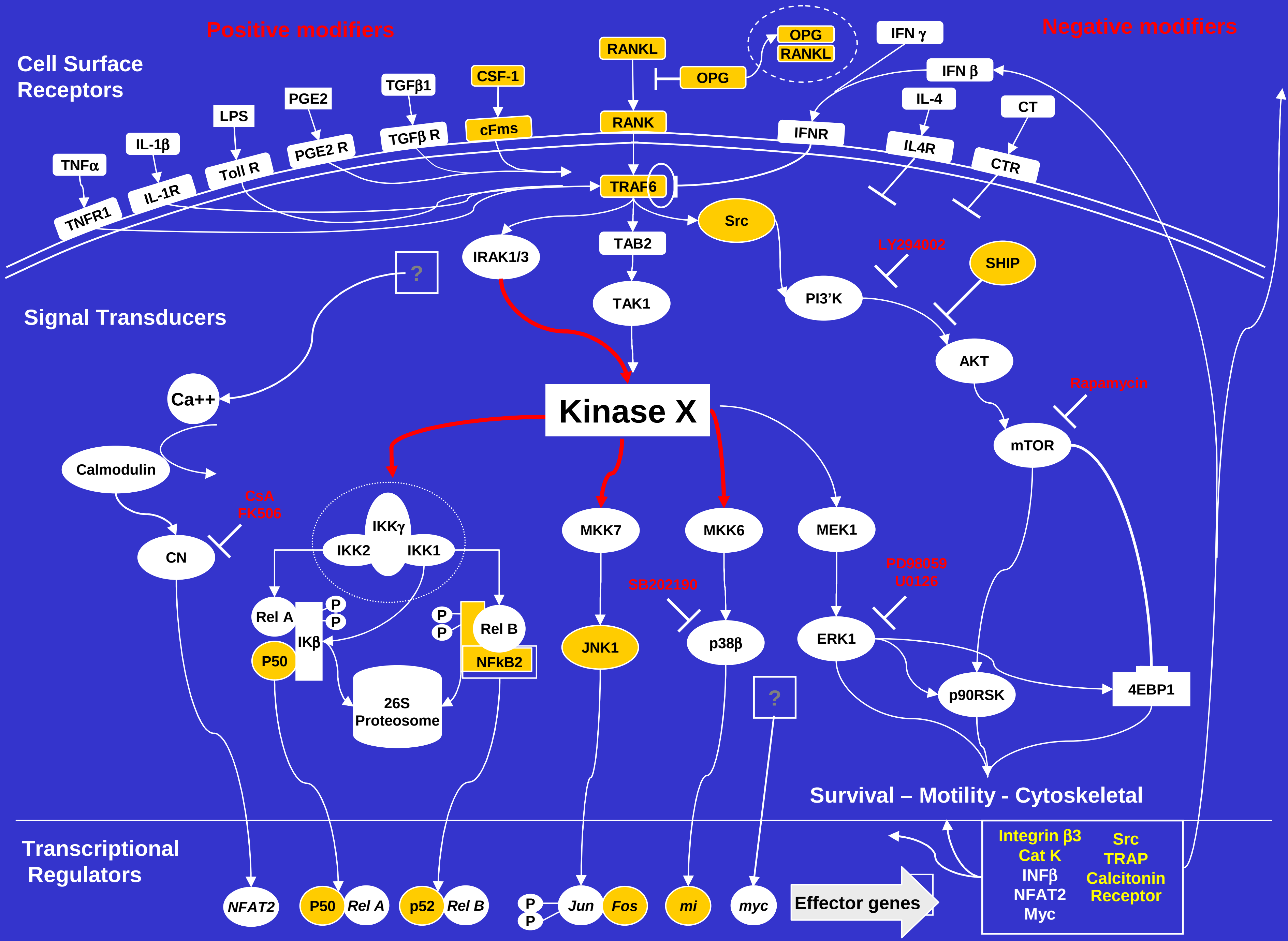
Integrin β3
Cat K
INFβ
NFAT2
Myc

Src
TRAP
Calcitonin Receptor

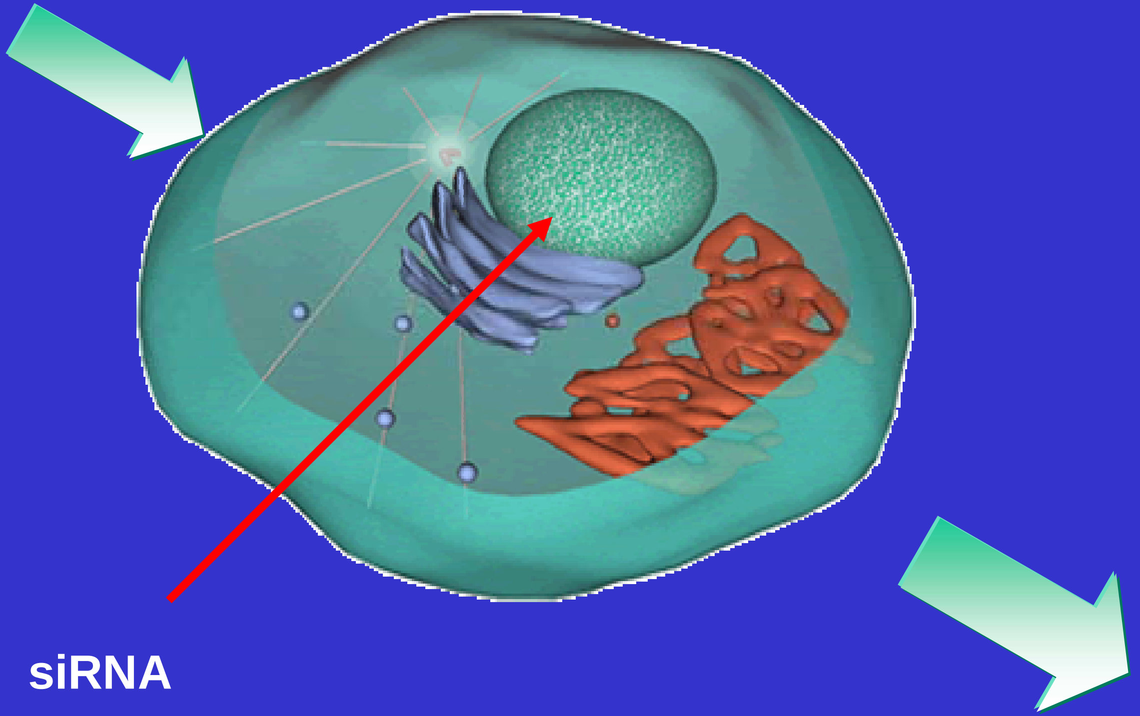
Can we identify other genes that control osteoclastogenesis and osteoclast activation?

Human Network

- Links are generated from the analysis of literature, co-evolution and co-expression
- 200,000 links between 20,000 human proteins
- Links are 70% accurate in recapitulating known pathway associations
- 30,000 links between 7000 proteins are supported by multiple methods


Positive modifiers
Negative modifiers
Cell Surface Receptors
TNFα
TNFR1
IL-1β
IL-1R
LPS
Toll R
PGE2
PGE2 R
TGFβ1
TGFβ R
CSF-1
cFms
RANKL
RANK
OPG
OPG
RANKL
IFN γ
IFN β
IFNR
IL-4
IL4R
CT
CTR
TRAF6
Src
IRAK1/3
TAB2
TAK1
PI3'K
LY294002
SHIP
AKT
Rapamycin
mTOR
Signal Transducers
?
Ca++
Calmodulin
CN
CsA
FK506
Kinase X
IKKγ
IKK2
IKK1
Rel A
P
P
IKβ
P50
26S Proteosome
P
P
Rel B
NFkB2
MKK7
MKK6
MEK1
SB202190
PD98059
U0126
JNK1
p38β
ERK1
p90RSK
4EBP1
?
Survival – Motility - Cytoskeletal
Transcriptional Regulators
NFAT2
P50
Rel A
p52
Rel B
P
P
Jun
Fos
mi
myc
Effector genes
Integrin β3
Cat K
INFβ
NFAT2
Myc
Src
TRAP
Calcitonin Receptor

Experimental Validation of Computational Predictions



Osteoclast Precursors

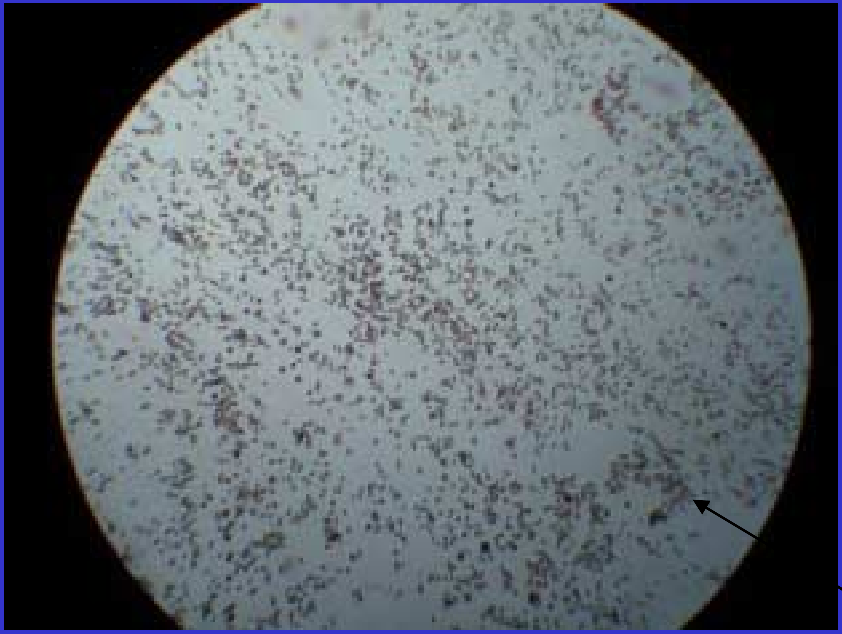
RANKL

siRNA

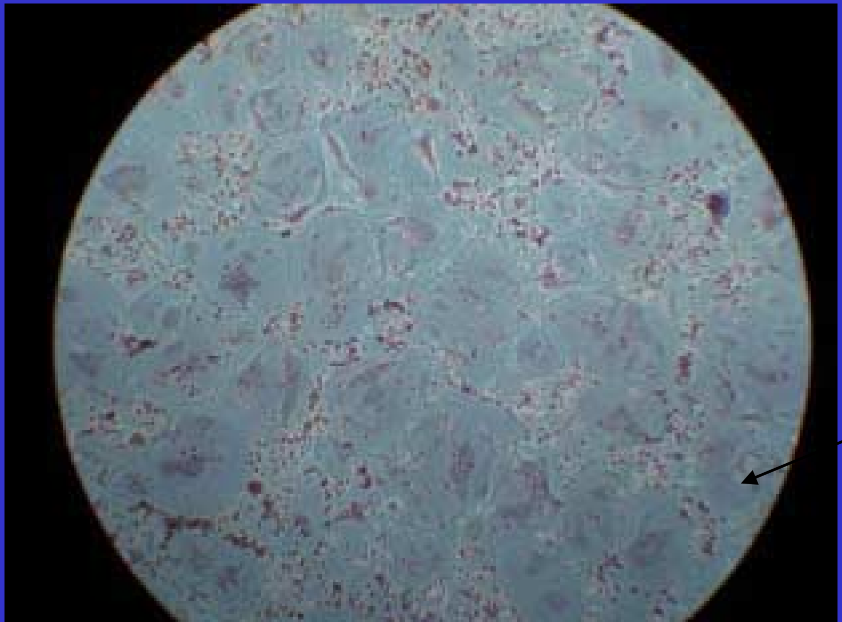
TRAP
CatK
CR
MNC

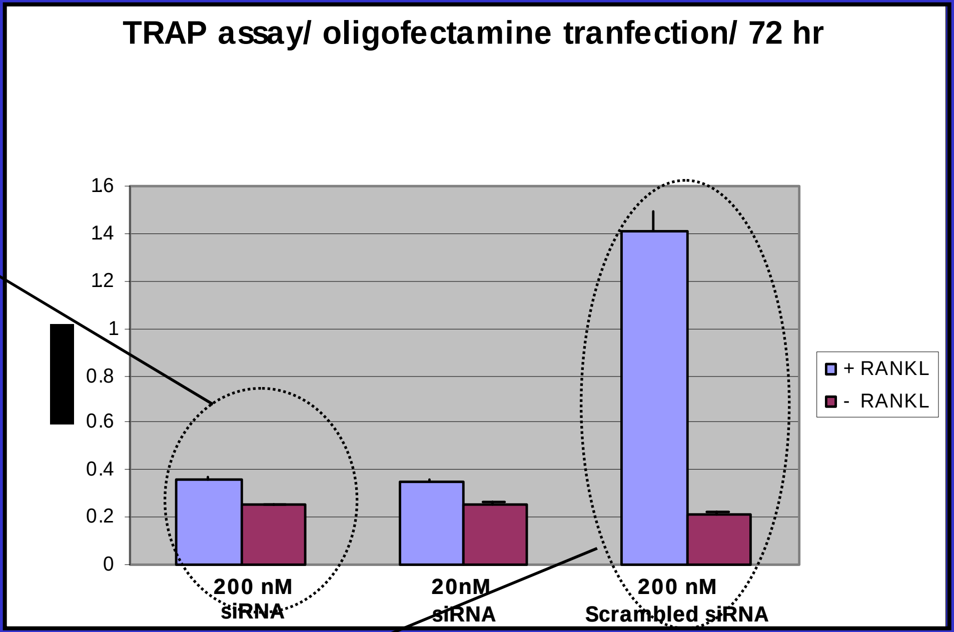
Inhibition of Kinase X blocks osteoclast formation in RAW cells

INR siRNA



Scrambled siRNA

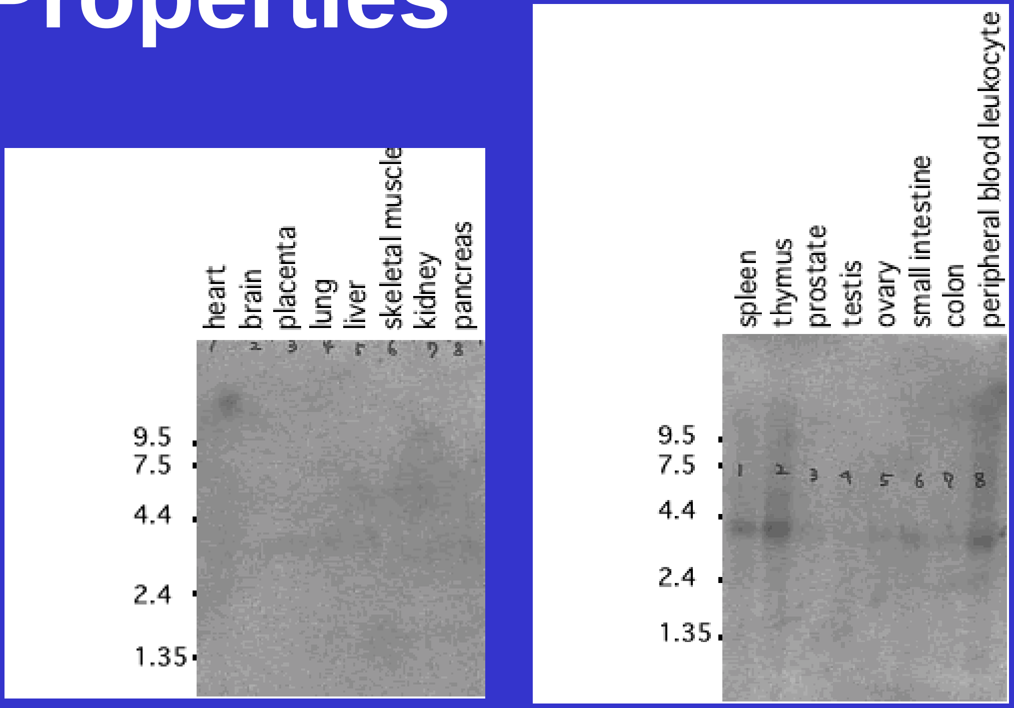




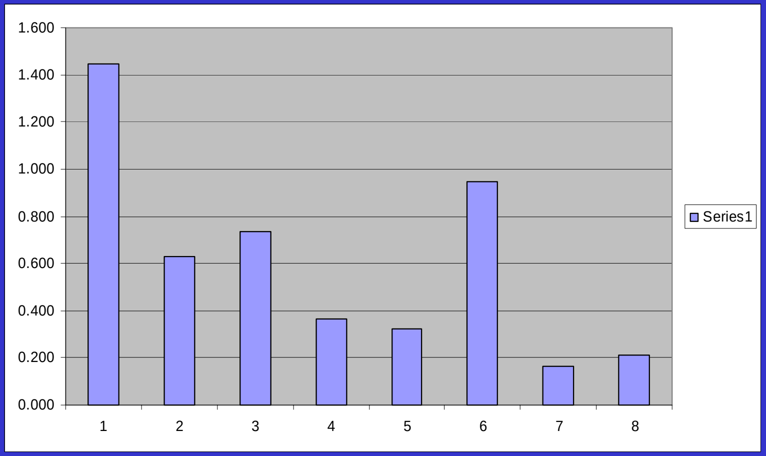
TRAP assay/ oligofectamine tranfection/ 72 hr

Kinase X Properties

Narrow Tissue Distribution



Induced by RANK



4h 1d 1d 1d 1d 4d 8d 8d

Conclusions

- We have developed computational techniques for constructing networks of protein functional relationships

- These techniques search for pairs of proteins that are correlated across genomes,expression arrays or abstracts

- The networks may be used to assign genes to functional categories, predict essential genes and discover genes that regulate cellular differentiation

Acknowledgements

Protein Pathways

Christos Andronis
Bill Boyle
Peter Bowers
Joseph Fierro
Seenu Kothakota
Kelly Oliner
Darin Taverna
Michael Thompson
Steve Wickert

UCLA

David Eisenberg
Todd Yeates
Lukasz Salwinski
Ioannis Xenarios

UT Austin

Edward Marcotte