Bioinformatics 2

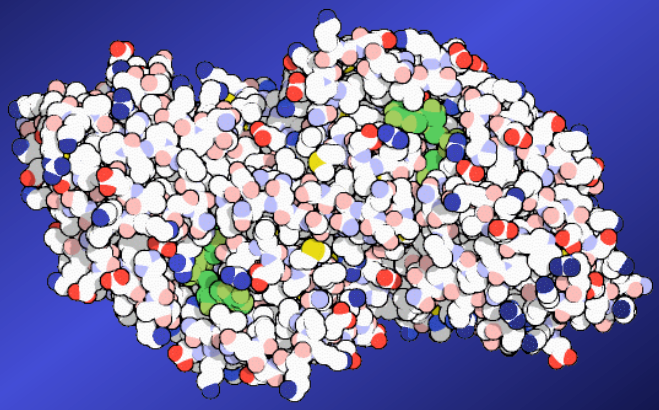
Protein (Interaction) Networks

Armstrong, 2009

- Biological Networks in general
- Metabolic networks
- Briefly review proteomics methods
- Protein-Protein interactions
- Protein Networks
- Protein-Protein interaction databases
- An example

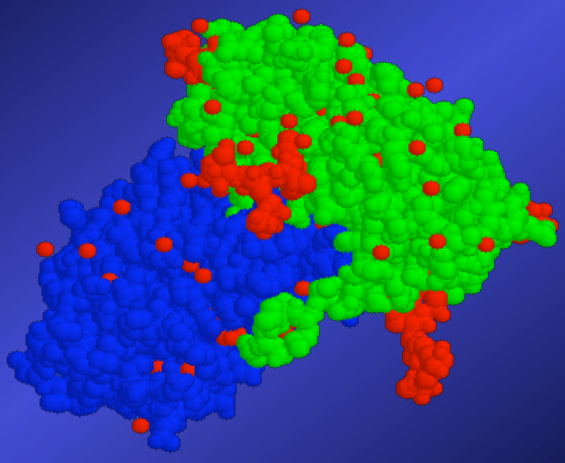
Armstrong, 2009

alcohol dehydrogenase



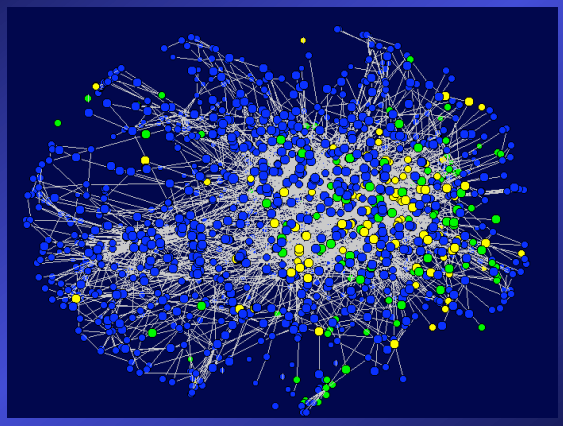
Armstrong, 2009

ricin (A and B)



Armstrong, 2009

synaptic proteome



Armstrong, 2009

Biological Networks

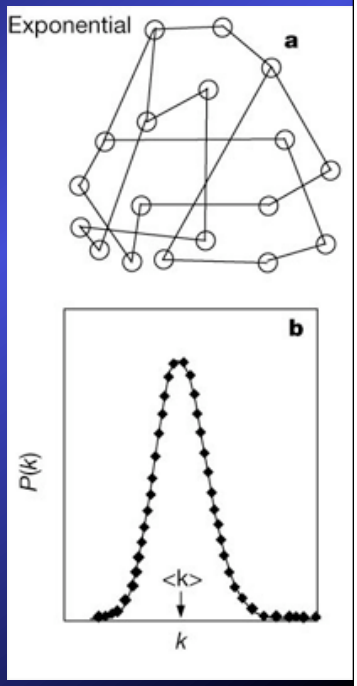
- Genes - act in cascades
- Proteins - form functional complexes
- Metabolism - formed from enzymes and substrates
- The CNS - neurons act in functional networks
- Epidemiology - mechanics of disease spread
- Social networks - interactions between individuals in a population
- Food Chains

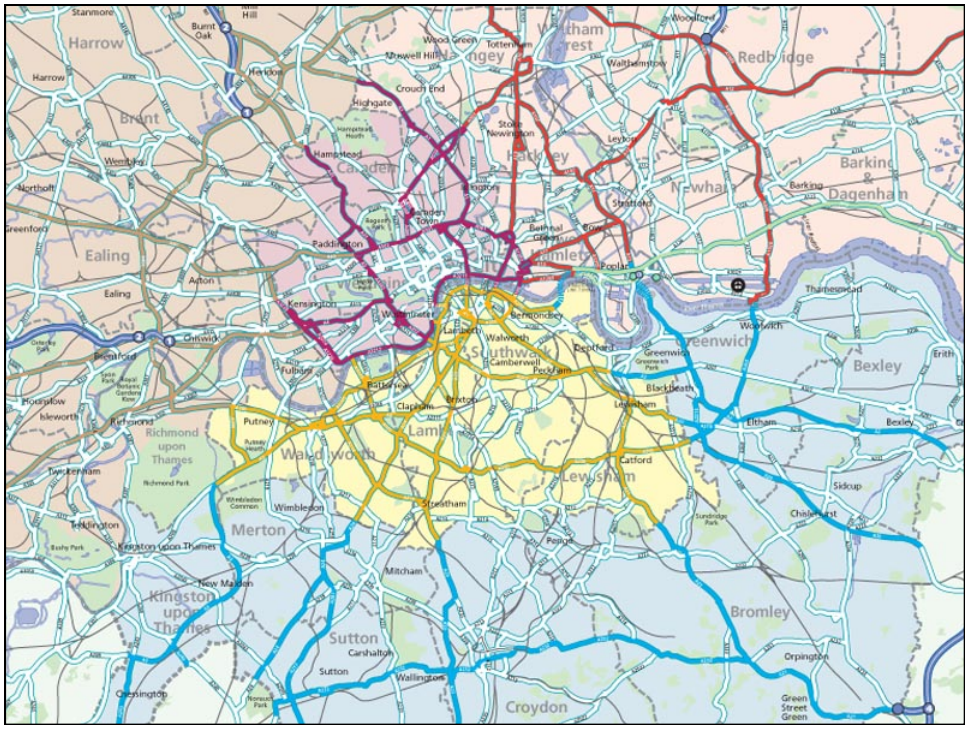
Armstrong, 2009

Large scale organisation

- First networks in biology generally modeled using classic random network theory.
- Each pair of nodes is connected with probability *p*
- Results in model where most nodes have the same number of links <*k*>
- The probability of any number of links per node is $P(k) \approx e^{-k}$

Armstrong, 2009

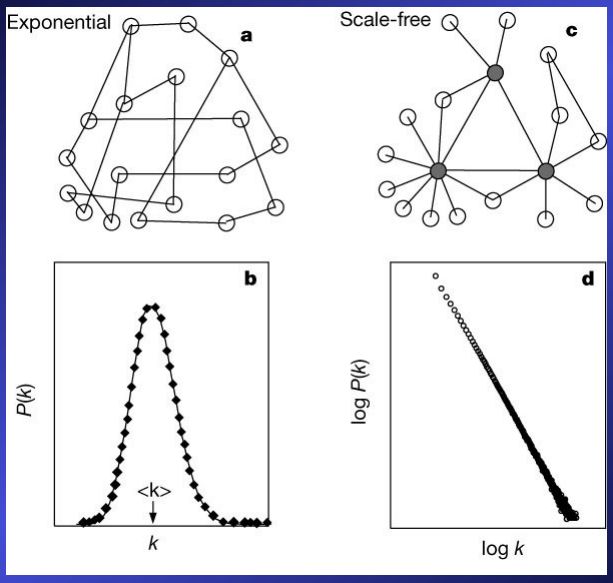




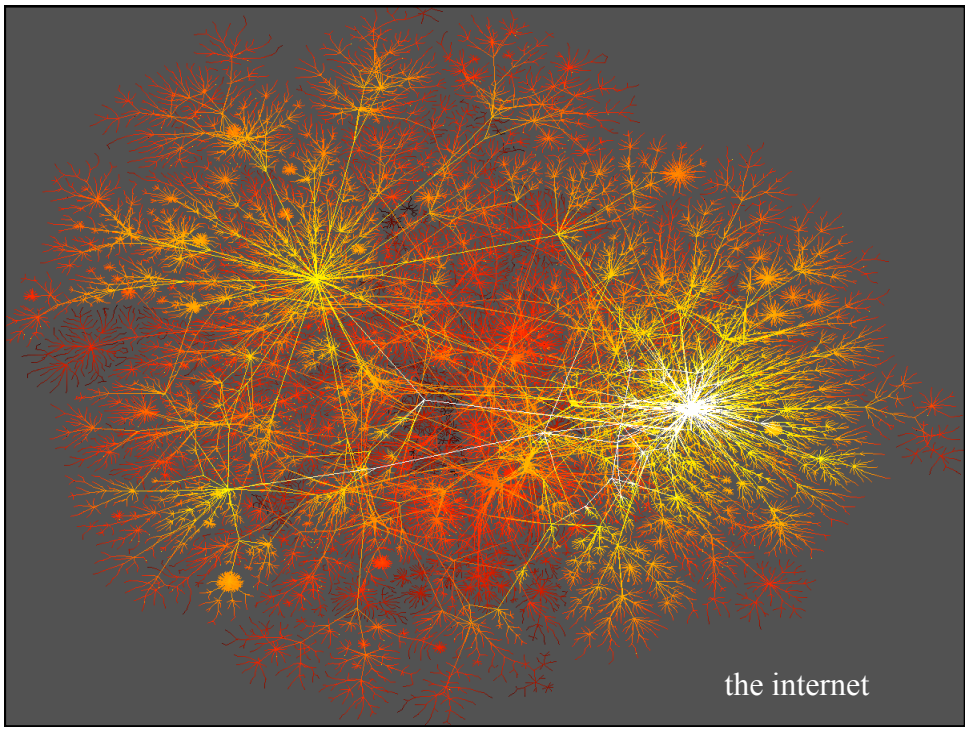
Non-biological networks

- Research into WWW, internet and human social networks observed different network properties
  - 'Scale-free' networks
  - P(k) follows a power law: P(k)≈$k^{-\gamma}$
  - Network is dominated by a small number of highly connected nodes - hubs
  - These connect the other more sparsely connected nodes

Armstrong, 2009



Armstrong, 2009



Small worlds

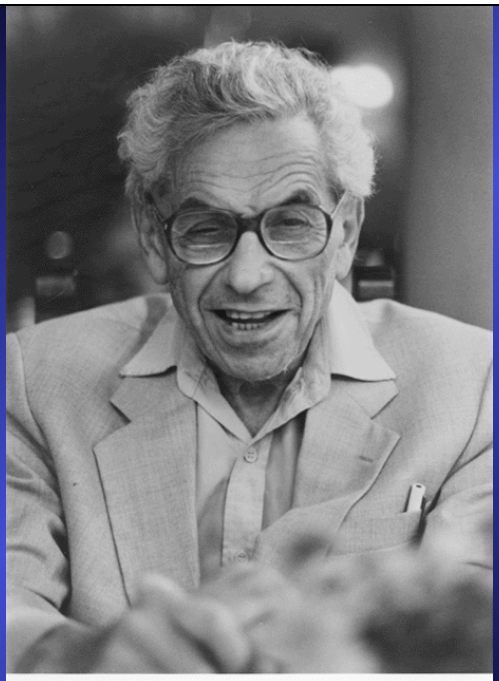
- General feature of scale-free networks
  - any two nodes can be connected by a relatively short path
  - average between any two people is around 6
    - What about SARS???
  - 19 clicks takes you from any page to any other on the internet.

6 degrees of separation..?

- Stanley Milgram's work in late 1960's
- Sent letters to people in Nebraska
- Target unknown person in Massachusetts
- Average 6 'jumps' to reach target

(only 5% got there)

Armstrong, 2009

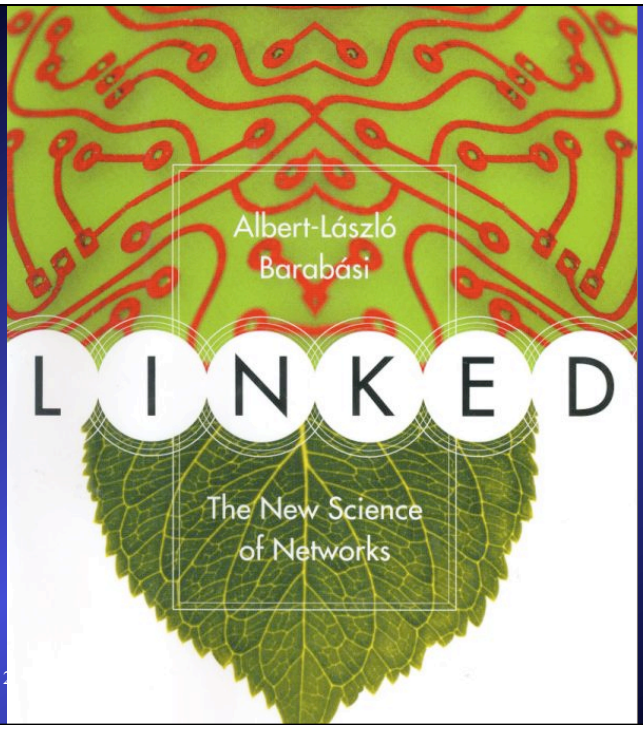


Paul Erdős, the most prolific mathematician who ever lived, has no home and no job, but he has wandered the world for over fifty years, inspiring other mathematicians. From the documentary N is a Number: A Portrait of Paul Erdős © 1993 by George Csicsery

Armstrong, 2009

http://oracleofbacon.org/
(real centre is Rod Steiger)

Armstrong, 2009



Armstrong, 2

Biological organisation

Jeong et al., 2000 The large-scale organisation of metabolic networks. Nature 407, 651-654

- Pioneering work by Oltvai and Barabasi
- Systematically examined the metabolic pathways in 43 organisms
- Used the WIT database
  - 'what is there' database
  - http://wit.mcs.anl.gov/WIT2/
  - Genomics of metabolic pathways

Armstrong, 2009

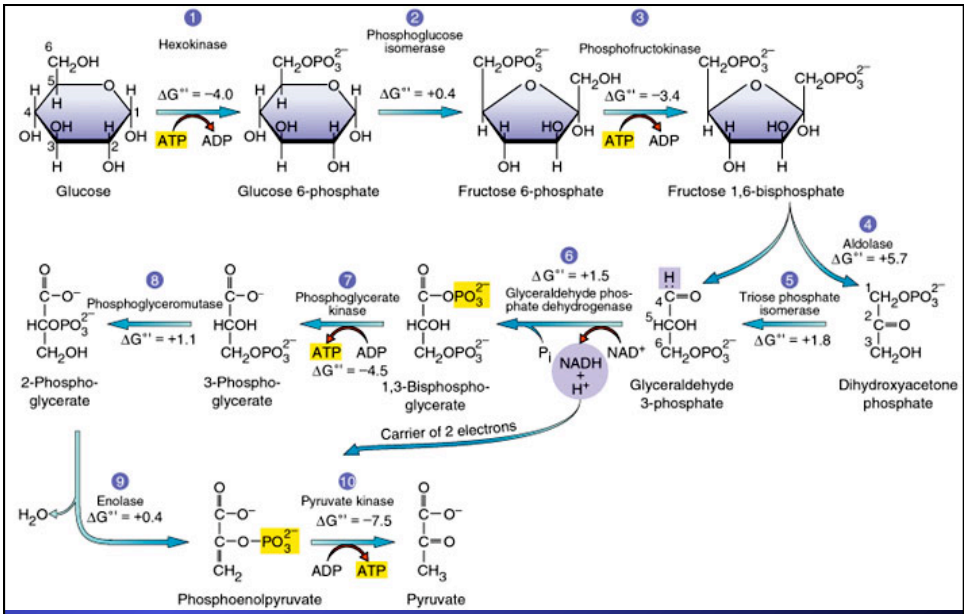
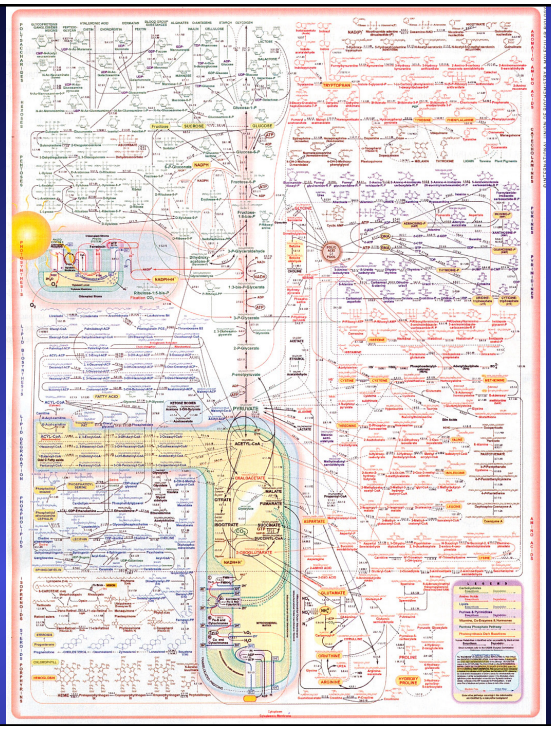


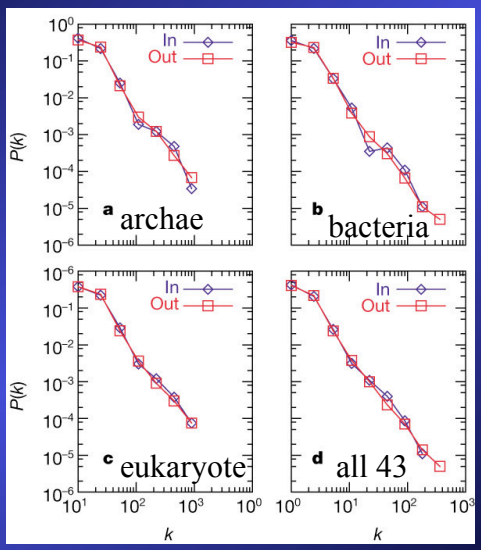
Image taken from http://fig.cox.miami.edu/~cmallery/255/255atp/255makeatp.htm

Armstrong, 2009



Armstrong, 2009

Using metabolic substrates as nodes

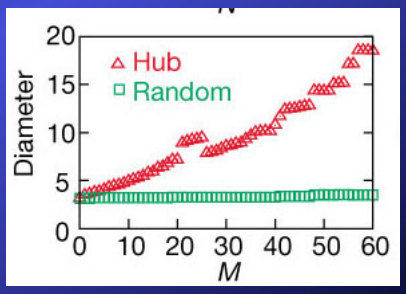


=scale free!!!

Armstrong, 2009

Random mutations in metabolic networks

- Simulate the effect of random mutations or mutations targeted towards hub nodes.
  - Measure network diameter
  - Sensitive to hub attack
  - Robust to random



Armstrong, 2009

Consequences for scale free networks

- Removal of highly connected hubs leads to rapid increase in network diameter
  - Rapid degeneration into isolated clusters
  - Isolate clusters = loss of functionality
- Random mutations usually hit non hub nodes
  - therefore robust
- Redundant connectivity (many more paths between nodes)

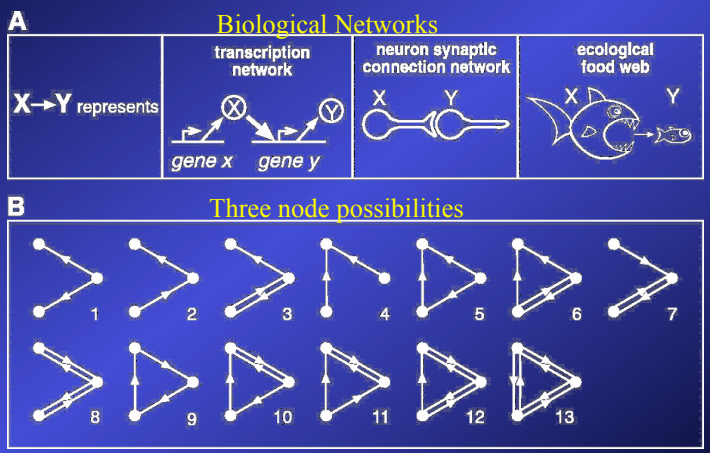
Armstrong, 2009

Network Motifs

- Do all types of connections exist in networks?
- Milo et al studied the transcriptional regulatory networks in yeast and E.Coli.
- Calculated all the three and four gene combinations possible and looked at their frequency

Armstrong, 2009

Milo et al. 2002 Network Motifs: Simple Building Blocks of Complex Networks. Science 298: 824-827



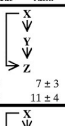
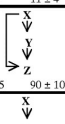
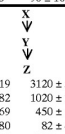
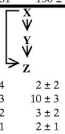
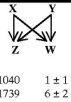
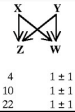
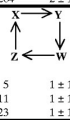
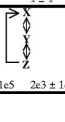
Armstrong, 2009

Gene sub networks

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41

Heavy bias in both yeast and E.coli towards these two sub network architectures

Armstrong, 2009

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

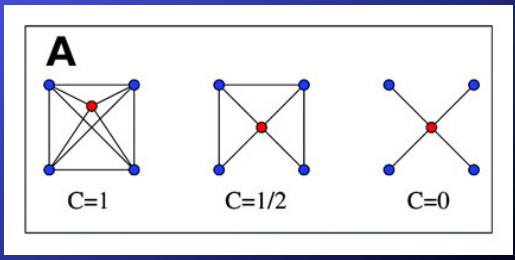
Armstrong

What about known complexes?

- OK, scale free networks are neat but how do all the different functional complexes fit into a scale free proteome arrangement?
  - e.g. ion channels, ribosome complexes etc?
- Is there substructure within scale free networks?
  - Examine the clustering co-efficient for each node.

Clustering co-efficients and networks.

- $C_i = 2n/k_i(k_i-1)$
- n is the number of direct links connecting the k_i nearest neighbours of node i
- A node at the centre of a fully connected cluster has a C of 1



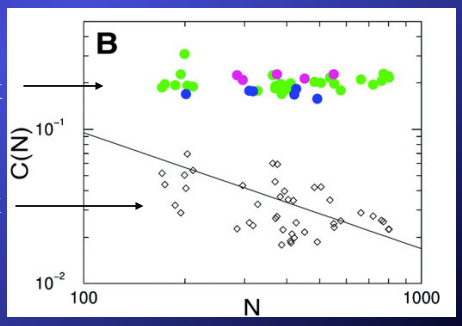
Clustering co-efficients and networks.

Ravasz et al.,(2002) Hierarchical Organisation of Modularity in Metabolic Networks. Science 297, 1551-1555

- The modularity (ave C) of the metabolic networks is an order of magnitude higher than for truly scale free networks.

Metabolic network

Non modular network

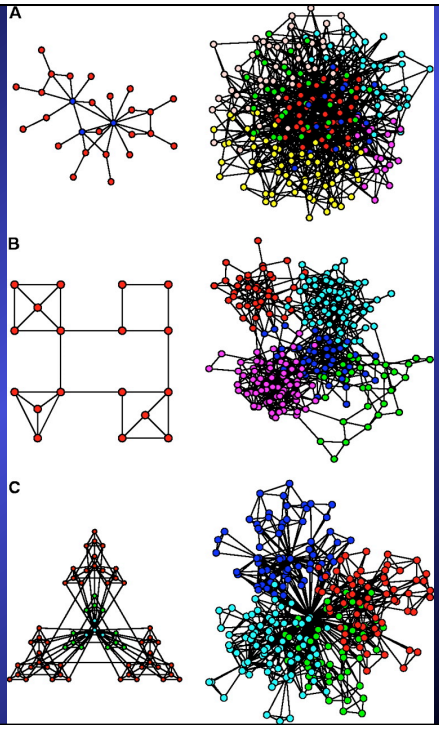


Armstrong, 2009

No modularity
Scale-free

Highly modular
Not scale free

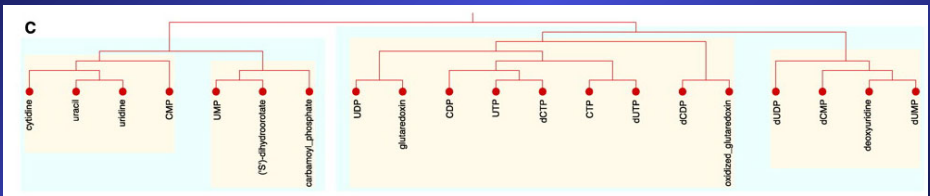
Hierarchical network
Scale-free



Armstrong, 2009

Clustering on C

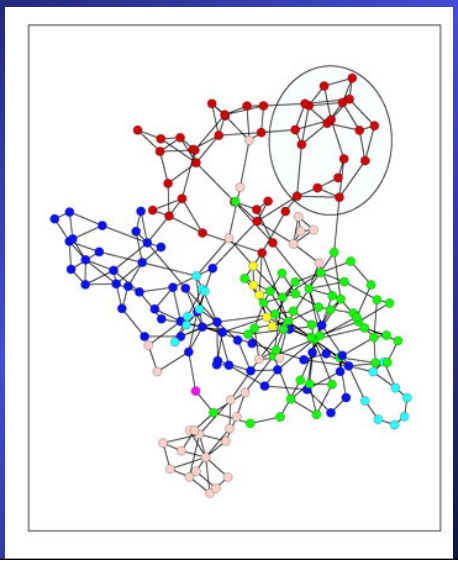
- Clustering on the basis of C allows us to rebuild the sub-domains of the network



- Producing a tree can predict functional clustered arrangements.

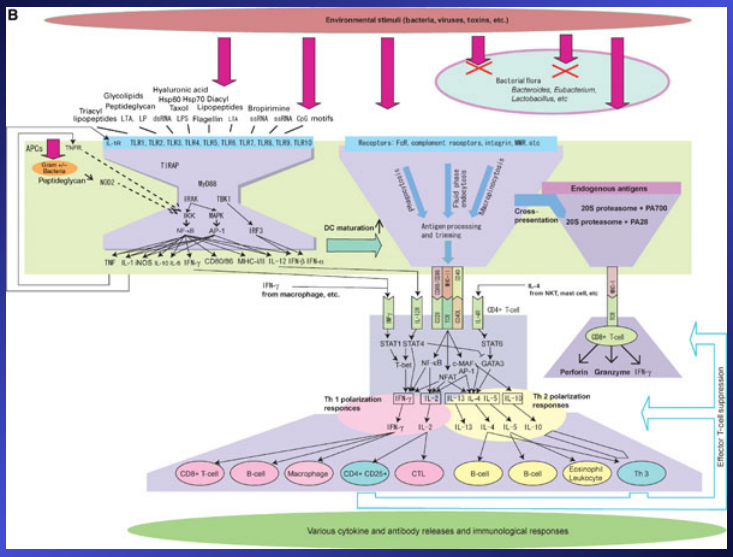
Armstrong, 2009

Cluster analysis on the network



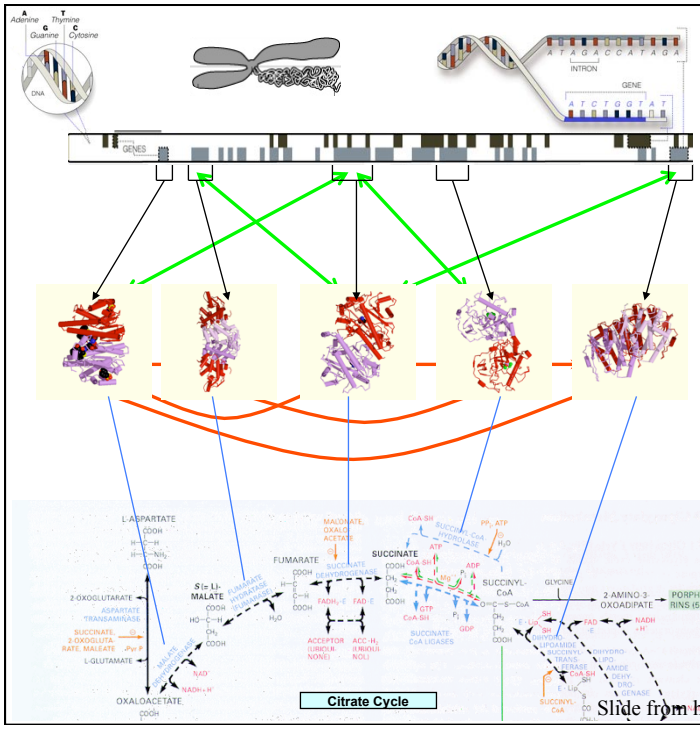
Armstrong, 2009

Bow-tie and nested bow-tie architectures



Armstrong, 2009

http://www.nature.com/msb/journal/v2/n1/fig_tab/msb4100039_F2.html



GENOME

protein-gene interactions

PROTEOME

protein-protein interactions

METABOLISM

Bio-chemical reactions

Slide from http://www.nd.edu/~networks/

Biological Profiling

- Microarrays
  - cDNA arrays
  - oligonucleotide arrays
  - whole genome arrays
- Proteomics
  - yeast two hybrid
  - PAGE techniques
  - Mass Spectrometry (Lecture 2)

Protein Interactions

- Individual Proteins form functional complexes
- These complexes are semi-redundant
- The individual proteins are sparsely connected
- The networks can be represented and analysed as an undirected graph

How to build a protein network

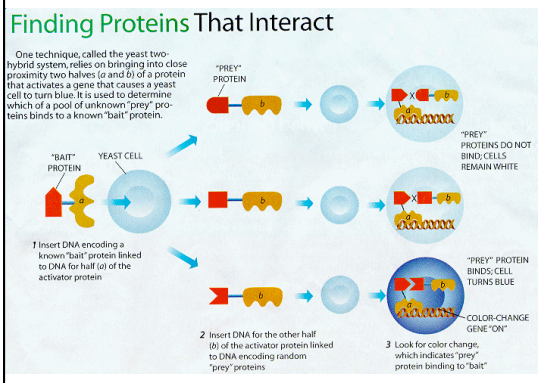
- Biological sample – how to you isolate your complex?
- What is in your complex?
- How is it connected?
  - Databases and Literature Mining
  - Yeast two hybrid screening & other cellular interaction assays
  - Mass-spec analysis
- Building and analysing the network
- An example

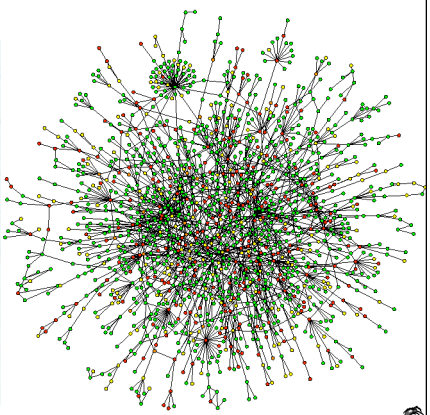
Armstrong, 2009

Yeast protein network

Nodes: proteins

Links: physical interactions (binding)





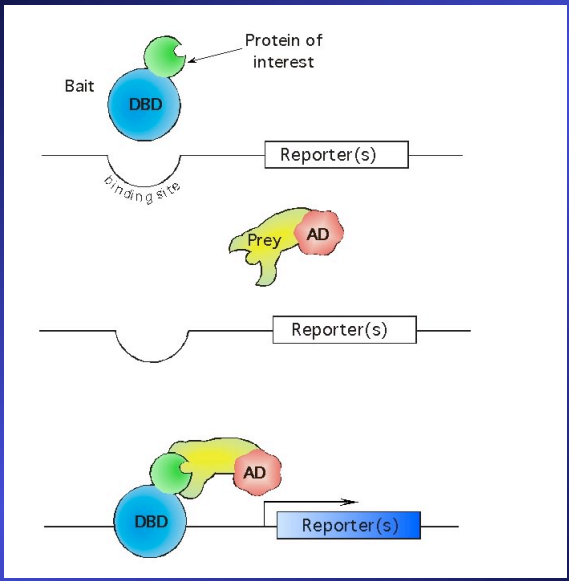
P. Uetz, *et al.* *Nature* **403**, 623-7 (2000).

Slide from http://www.nd.edu/~networks/

Yeast two hybrid

- Use two mating strains of yeast
- In one strain fuse one set of genes to a transcription factor DNA binding site
- In the other strain fuse the other set of genes to a transcriptional activating domain
- Where the two proteins bind, you get a functional transcription factor.

Armstrong, 2009

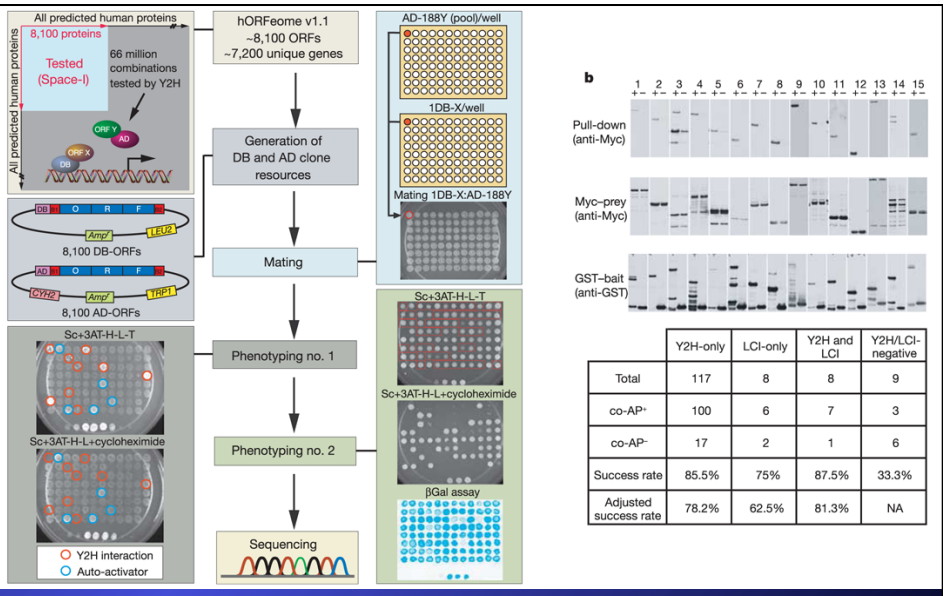


Armstrong, 2009

Data obtained

- Depending on sample, you get a profile of potential protein-protein interactions that can be used to predict functional protein complexes.
- False positives are frequent.
- Can be confirmed by affinity purification etc.

Armstrong, 2009



	Y2H-only	LCI-only	Y2H and LCI	Y2H/LCI-negative
Total	117	8	8	9
co-AP$^+$	100	6	7	3
co-AP$^-$	17	2	1	6
Success rate	85.5%	75%	87.5%	33.3%
Adjusted success rate	78.2%	62.5%	81.3%	NA

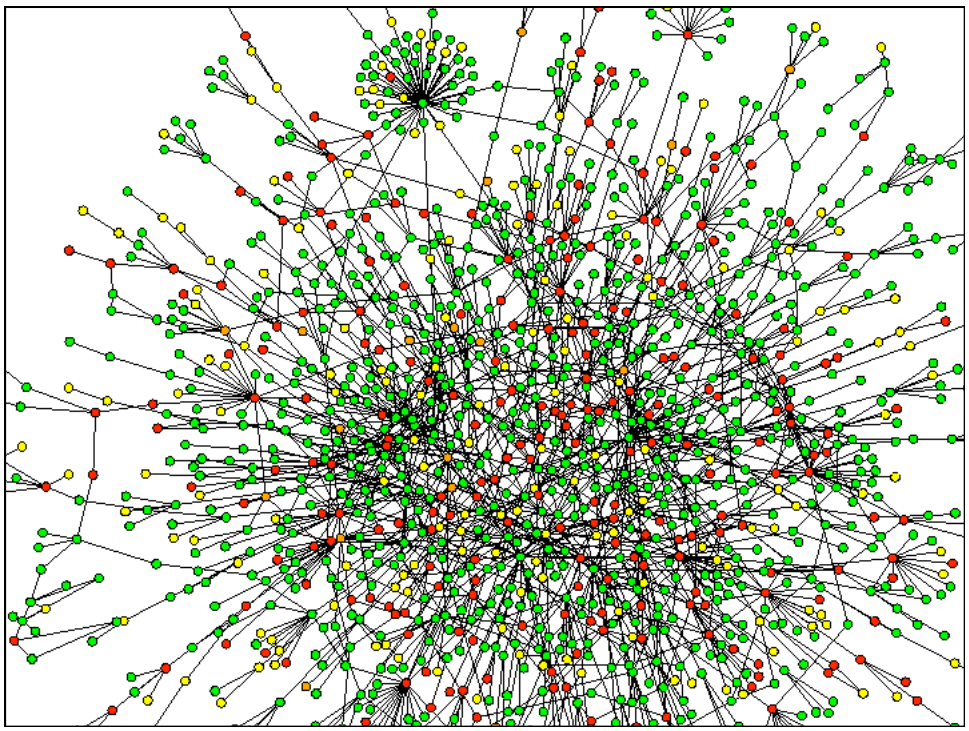
Interaction mapping schema from Rual et al 2005

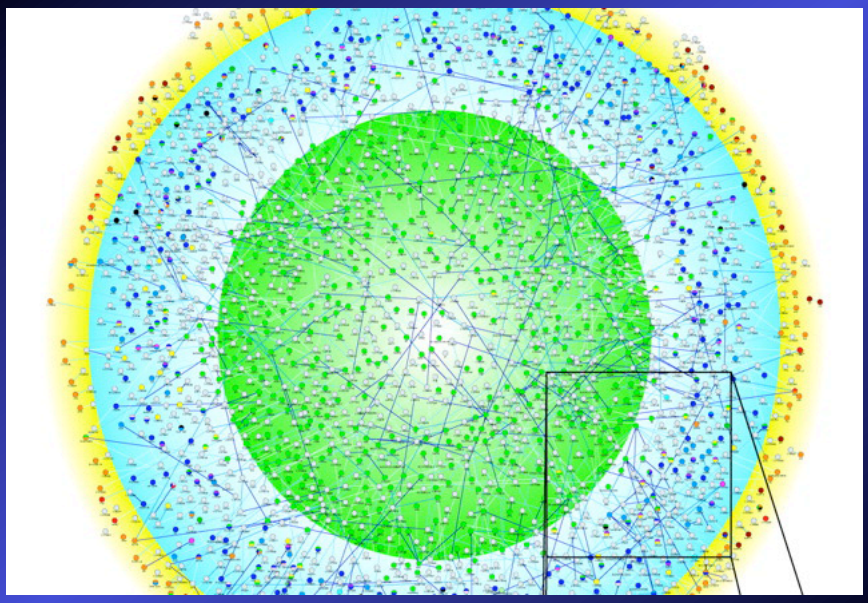
Armstrong, 2009

Protein Networks

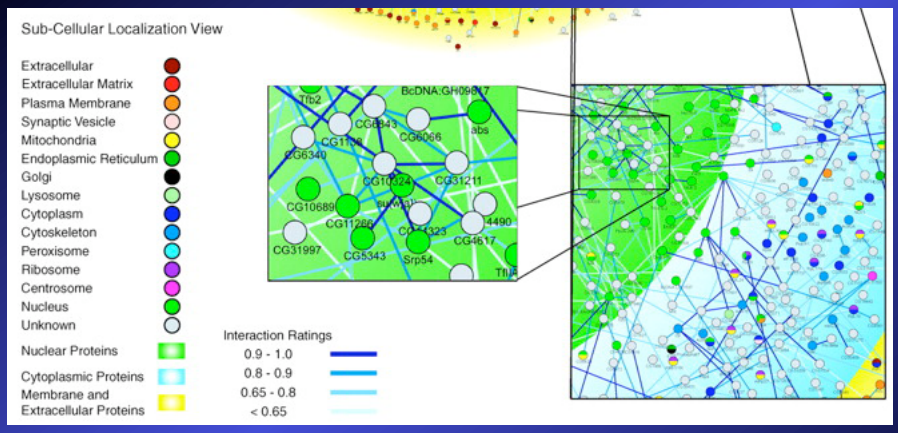
- Networks derived from high throughput yeast 2 hybrid techniques
  - yeast
  - *Drosophila melanogaster*
  - *C.elegans*
- Predictive value of reconstructed networks

Armstrong, 2009



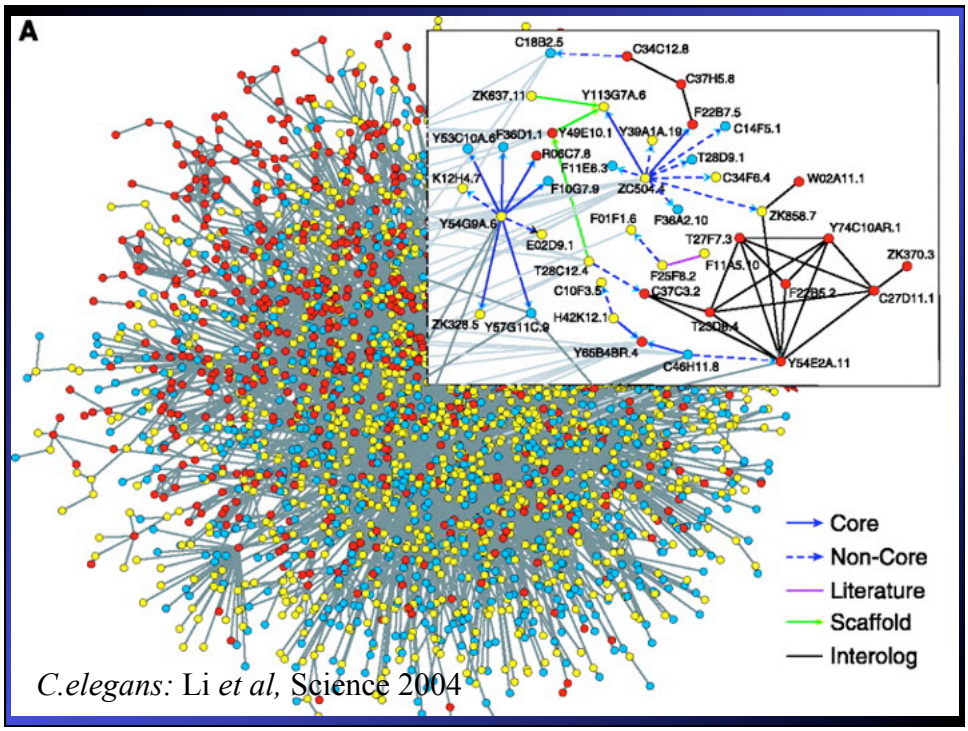


Armstrong, 2009

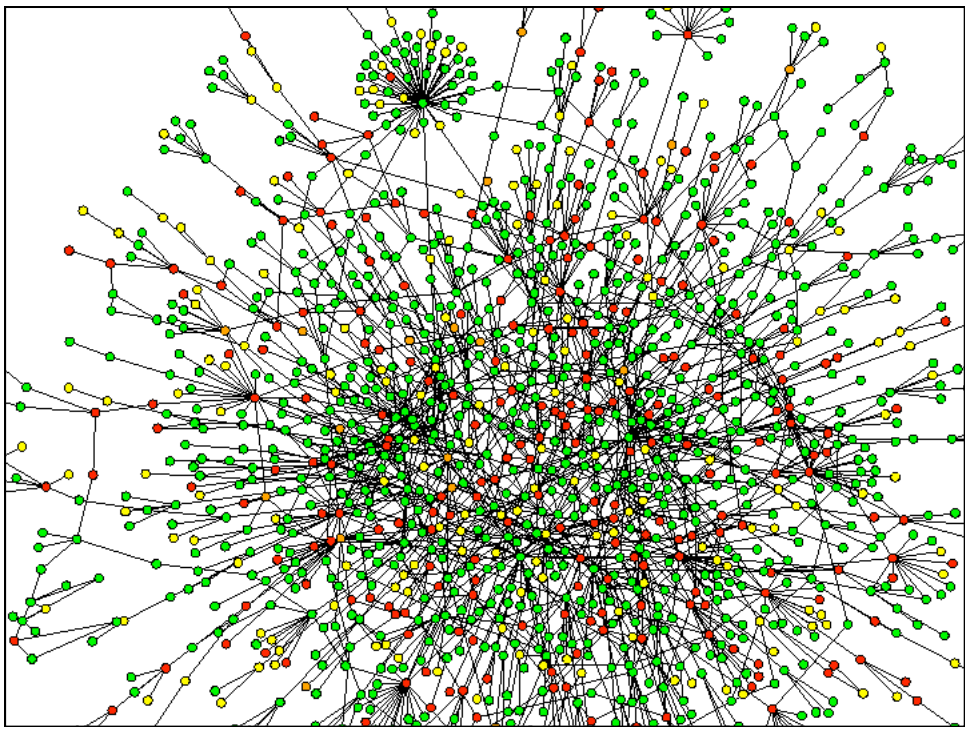


Giot *et al*, Science 2003

Armstrong, 2009



C.elegans: Li *et al,* Science 2004



Predictive value of networks

Jeong et al., (2001) Lethality and Centrality in protein networks. Nature 411 p41

- In the yeast genome, the essential vs. unessential genes are known.
- Rank the most connected genes
- Compare known lethal genes with rank order

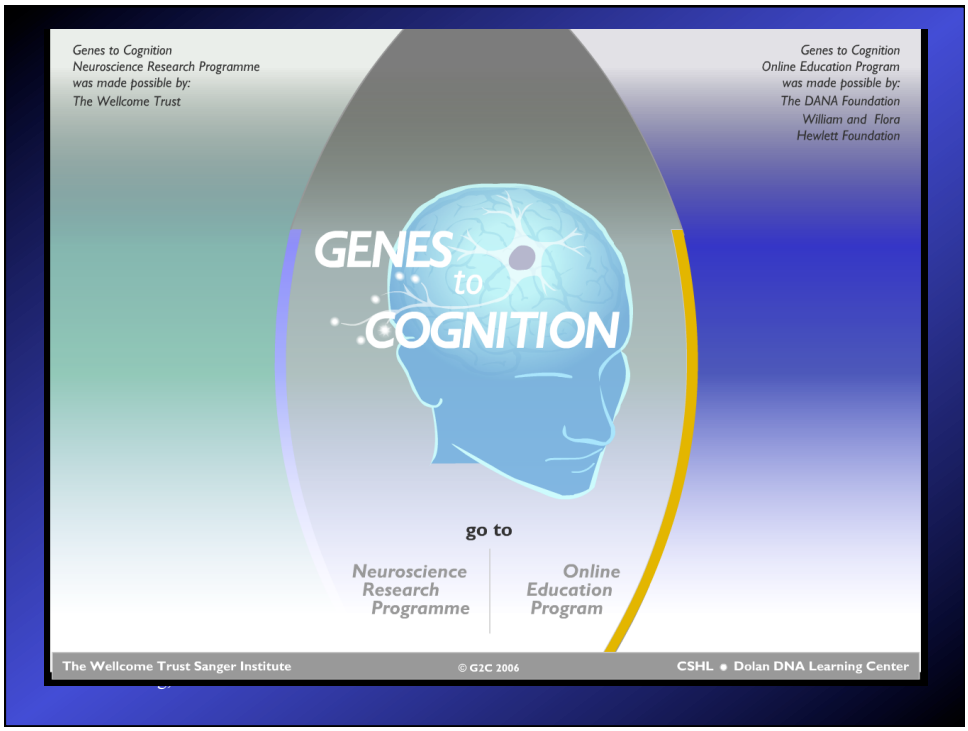
k	fraction	%lethal
<6	93%	21%
>15	0.7%	62%

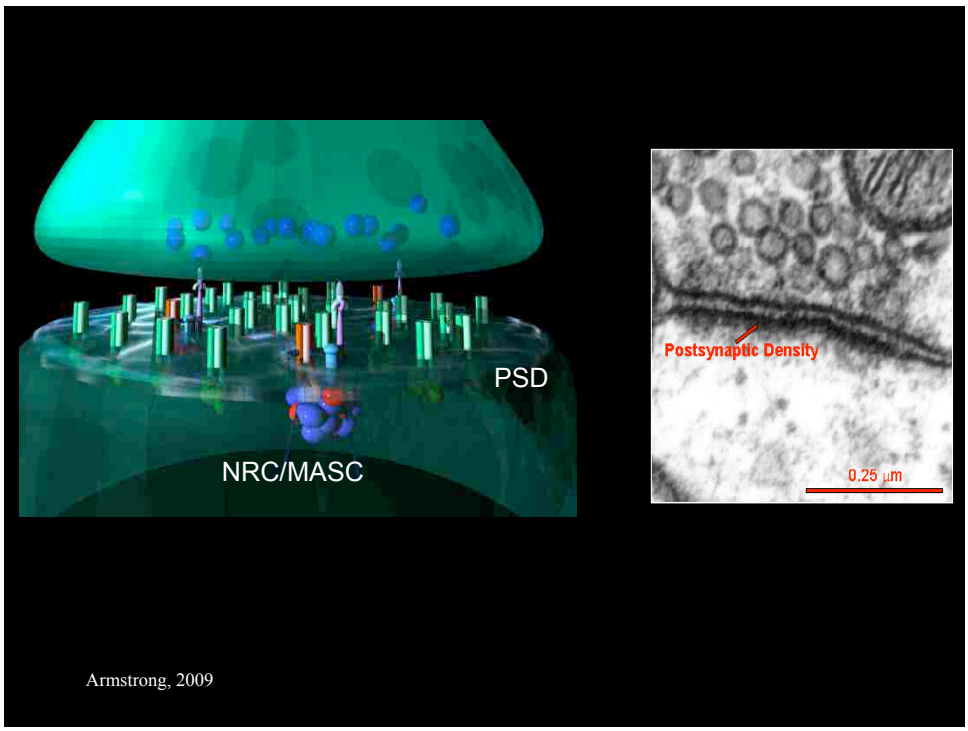
Armstrong, 2009

A walk-through example…

See linked papers on for further methodological details

Armstrong, 2009

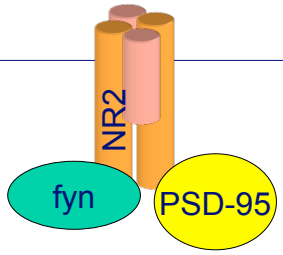




Armstrong, 2009

Genetic evidence for postsynaptic complexes

NMDA receptor



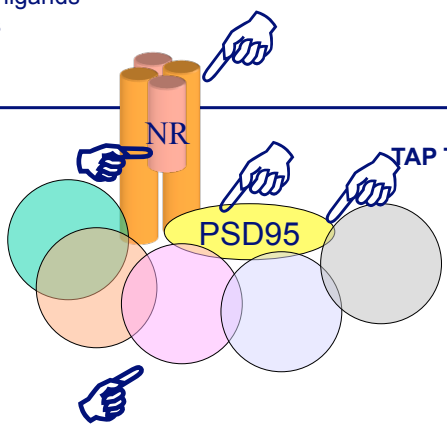
Learning impairments
Plasticity impairments

Grant, et al. Science, 258, 1903-10. 1992
Migaud et al, Nature , 396; 433-439. 1998
Sprengel et al. Cell 92, 279-89. 1998

Proteomic characterisation of NRC / MASC

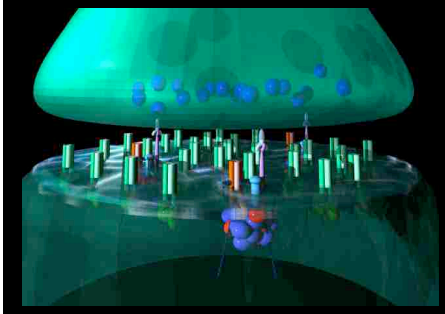
(MAGUK Associated Signaling Complex)

- glutamate ligands
- antibodies
- peptides
- TAP Tag



- ~2 MDa
- 77 proteins (2000)
- 186 (2005)

Husi et al. Nature Neuroscience, 3, 661-669. 2000.
Husi & Grant. J. Neurochem, 77, 281-291. 2001
Collins et al, J. Neurochem. 2005



Post Synaptic Density	1124
ER:microsomes	491
Splicesome	311
NRC/MASC	186
Nucleolus	147
Peroxisomes	181
Mitochondria	179
Phagosomes	140
Golgi	81
Choroplasts	81
Lysosomes	27
Exosomes	21

Armstrong, 2009

Grant. (2006) Biochemical Society Transactions. 34, 59-63. 2006

Literature Mining

- 680 proteins identified from protein preps
- Many already known to interact with each other
- Also interact with other known proteins
  - Immunoprecipitation is not sensitive (only finds abundant proteins)
- Literature searching has identified a group of around 4200 proteins
  - Currently we have extensive interaction data on 1700

Armstrong, 2009

Annotating the DB

- How do we find existing interactions?
  - **Search PubMed with keyword and synonym combinations**
  - Download abstracts
  - Sub-select and rank-order using regex's
  - Fast web interface displays the most 'productive' abstracts for each potential interaction

Armstrong, 2009

Keyword and synonym problem

- PSD-95:
  - DLG4,PSD-95,PSD95,Sap90,Tip-15,Tip15, Post Synatpic Density Protein - 95kD, PSD 95, Discs, large homolog 4, Presynaptic density protein 95
- NR2a:
  - Glutamate [NMDA] receptor subunit epsilon 1 precursor (N-methyl D-aspartate receptor subtype 2A) (NR2A) (NMDAR2A) (hNR2A) NR2a
- Protein interactions:
  - interacts with, binds to, does not bind to….

Armstrong, 2009

.+\sand\s.+\sinteract

(1..N characters) (space) and (1..N characters) interact

.+\s((is)|(was))\sbound\sto\s.+\s

(1..N characters) (space) (is or was) (space) bound (space) to (1..N characters) (space)

.+\sbinding\sof\s.+\s((and)|(to))\s.+

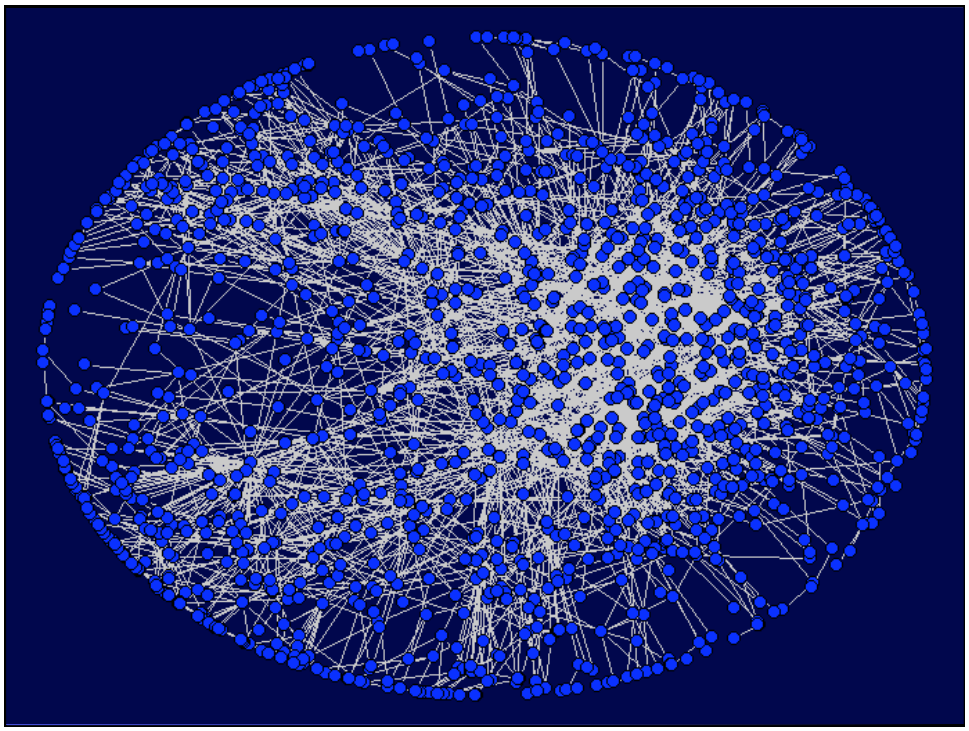
(1..N characters) (space) binding (space) of (and or to) (space) (1..N characters)

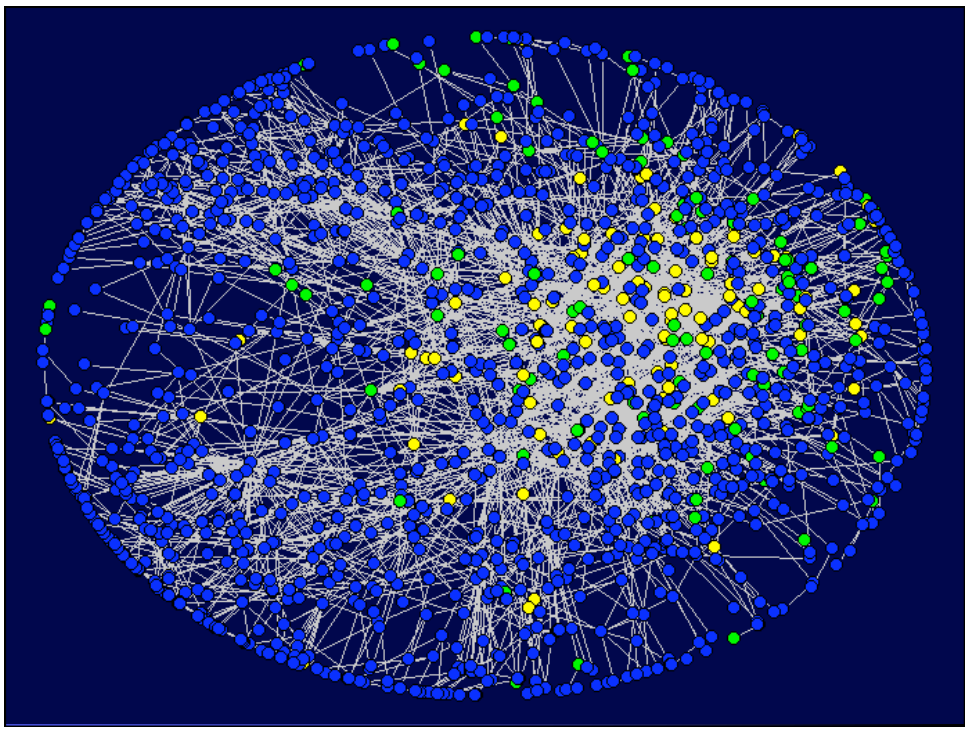
Armstrong, 2009

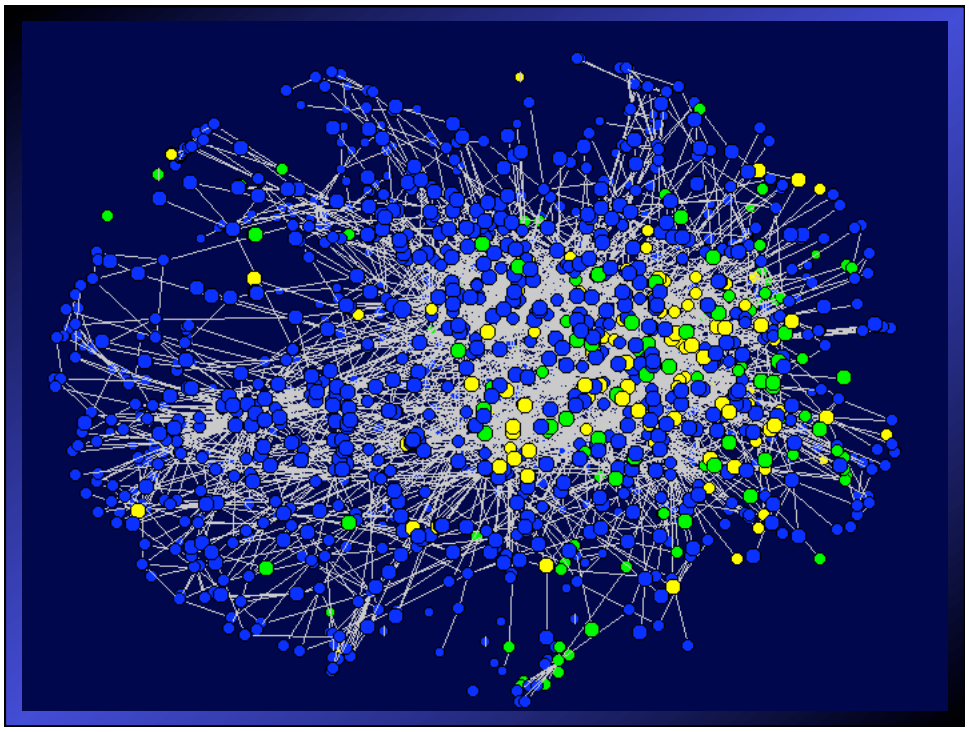
Annotating the DB

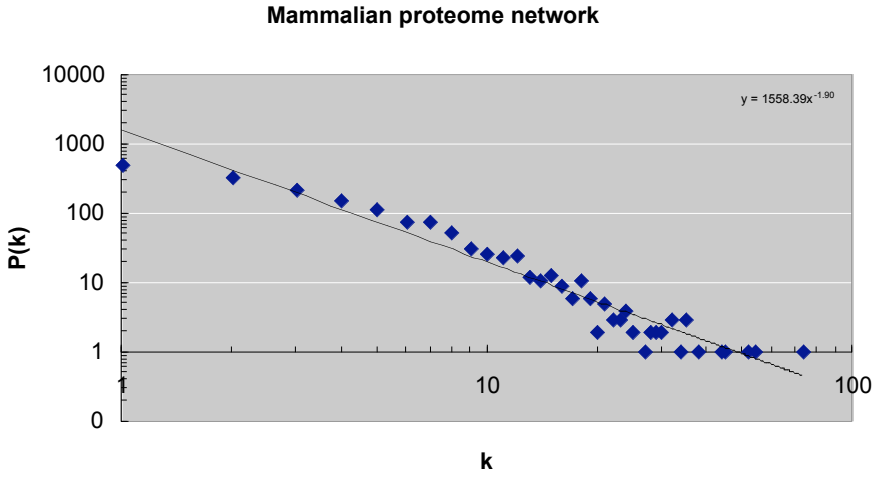
- How do we find existing interactions?
  - Search PubMed with keyword and synonym combinations
  - Download abstracts
  - Sub-select and rank-order using regex's
  - Fast web interface displays the most 'productive' abstracts for each potential interaction
  - *Learn from good vs. bad abstracts*

Armstrong, 2009

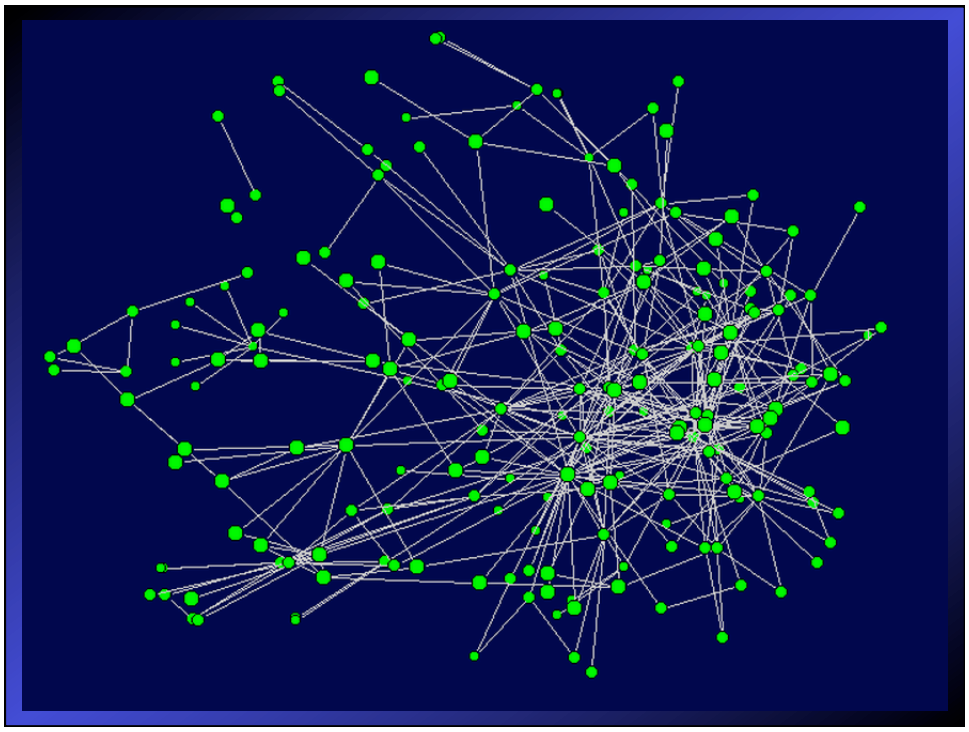


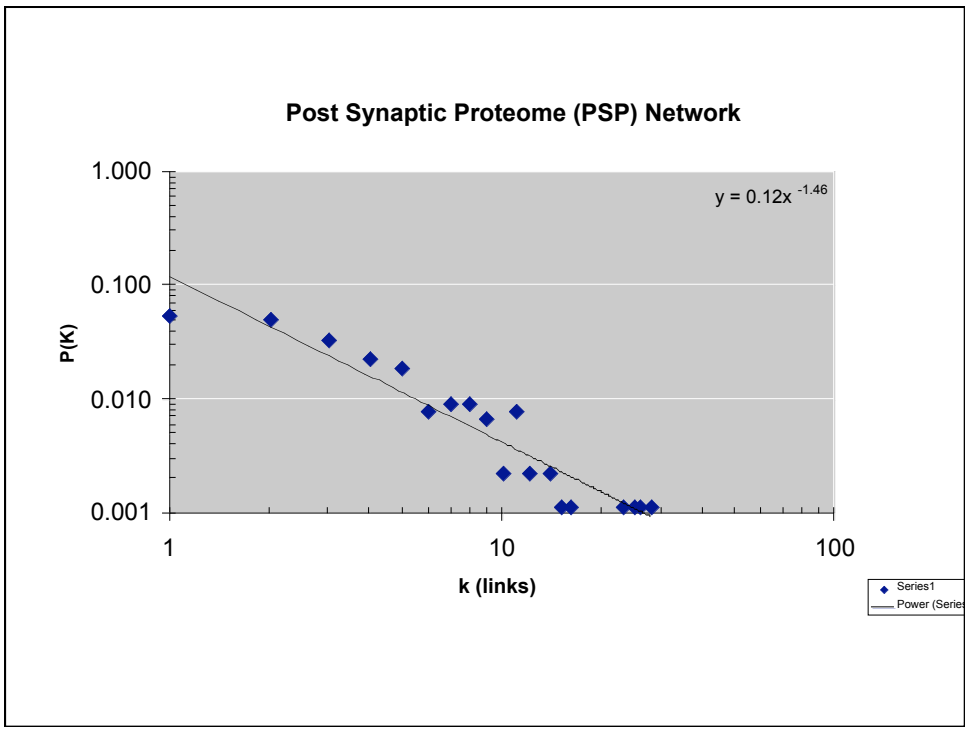


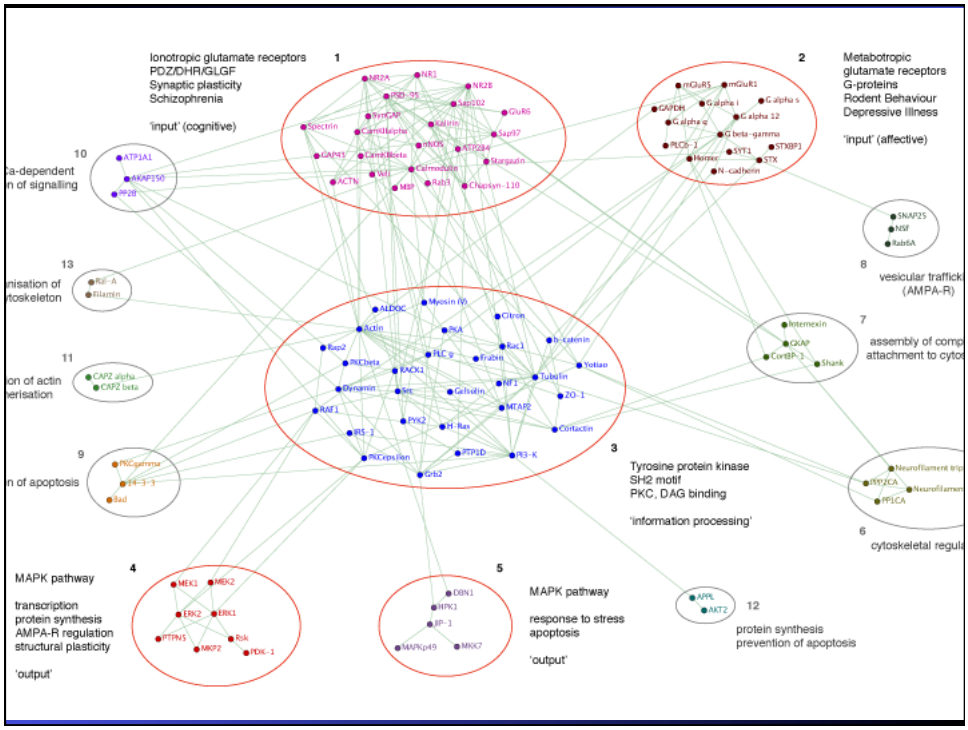




1729 proteins, 7739 links




Post Synaptic Proteome (PSP) Network
y = 0.12x -1.46
1.000
0.100
0.010
0.001
P(K)
1
10
100
k (links)
Series1
Power (Series

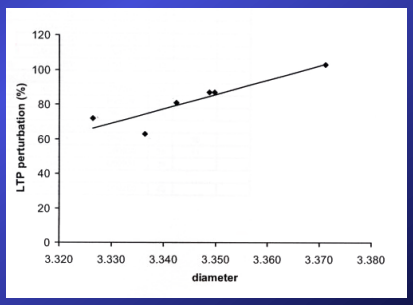


Simulated disruption vs. mutations

Linear correlation between simulation and *in vivo* assay

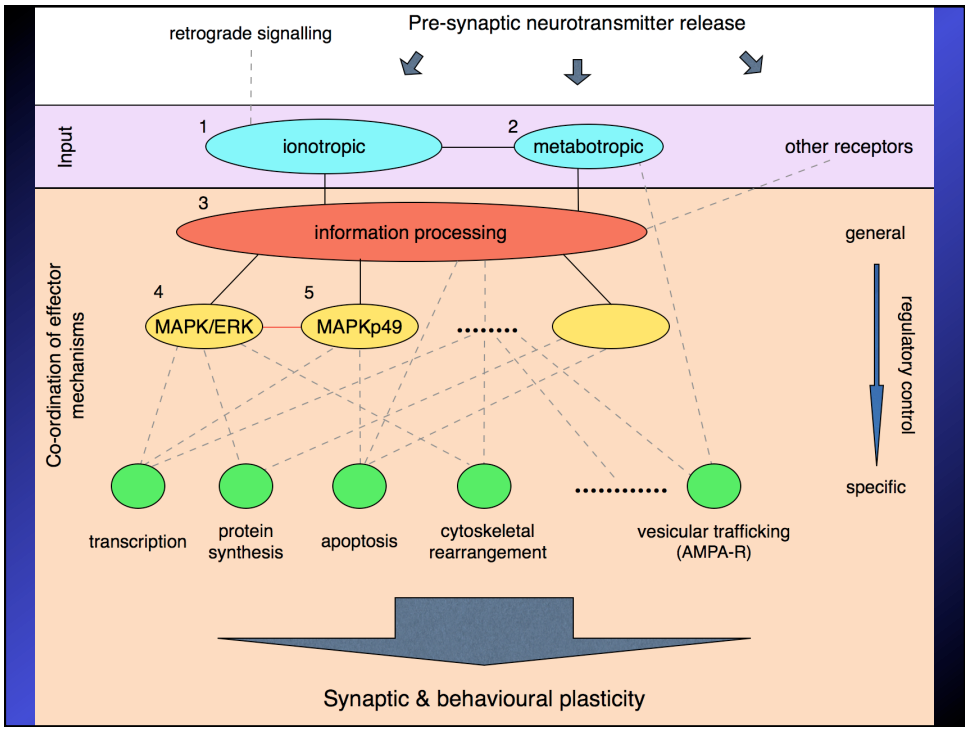
Details: Mutations in MEK1, SynGAP, NR2AC, PKA, PI3-kinase, PSD-95 were all analysed in a single laboratory (TJ O'Dell, UCSD) under controlled conditions and LTP disruption measured. ($p<0.05$)

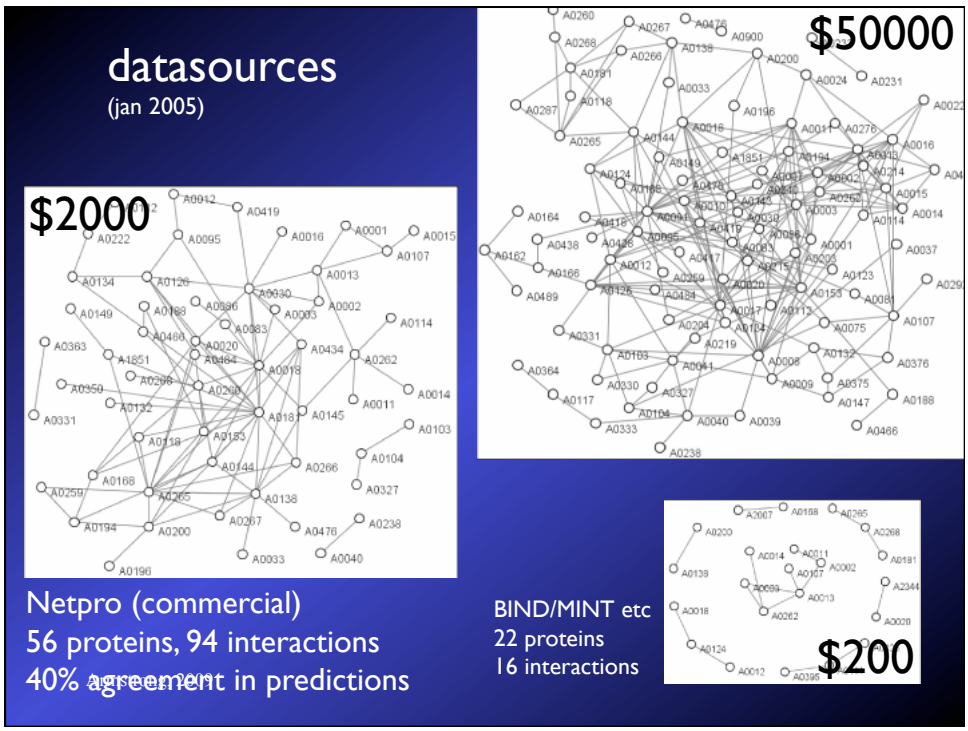
5 HZ



H. Husi J. Choudhary L.Yu M. Cumiskey W. Blackstock T.J. O'Dell P.M.Visscher J.D.Armstrong S.G.N.Grant, unpublished

Armstrong, 2009


retrograde signalling
Pre-synaptic neurotransmitter release
Input
1
ionotropic
2
metabotropic
other receptors
3
information processing
general
Co-ordination of effector mechanisms
4
MAPK/ERK
5
MAPKp49
regulatory control
specific
transcription
protein synthesis
apoptosis
cytoskeletal rearrangement
vesicular trafficking (AMPA-R)
Synaptic & behavioural plasticity


datasources
(jan 2005)
$2000
$50000
$200
Netpro (commercial)
56 proteins, 94 interactions
40% agreement in predictions
BIND/MINT etc
22 proteins
16 interactions

Synapse proteome summary

- Protein parts list from proteomics
- Literature searching produced a network
- Network is essentially scale free
- Hubs more important in cognitive processes
- Network clusters show functional subdivision
- Overall architecture resembles bow-tie model
- Expensive…

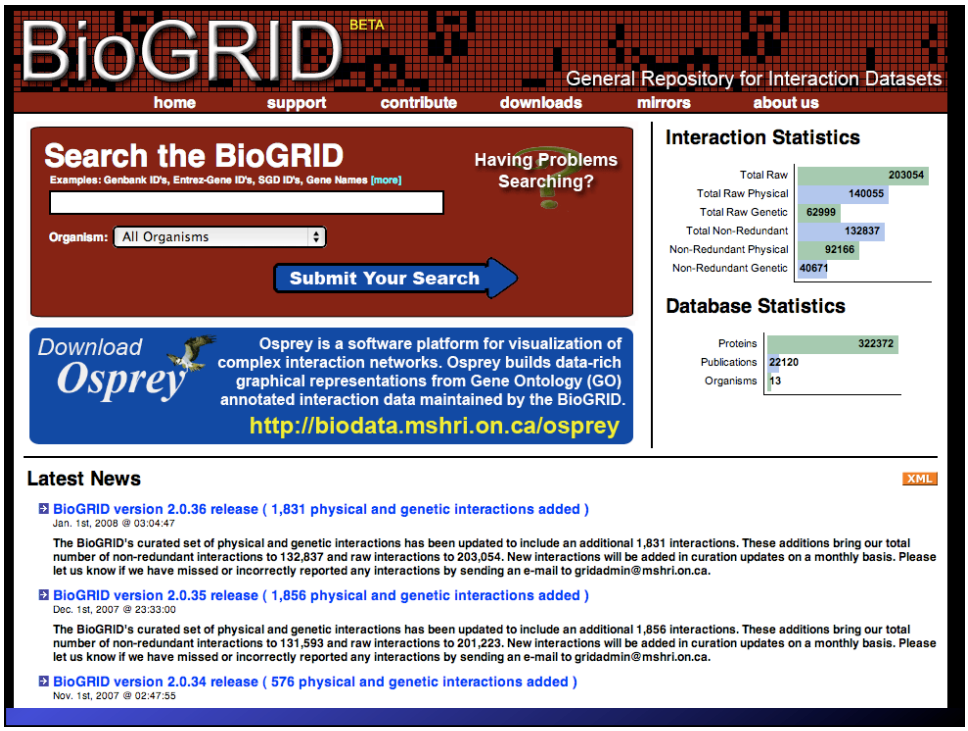
Armstrong, 2009

Protein (and gene) interaction databases

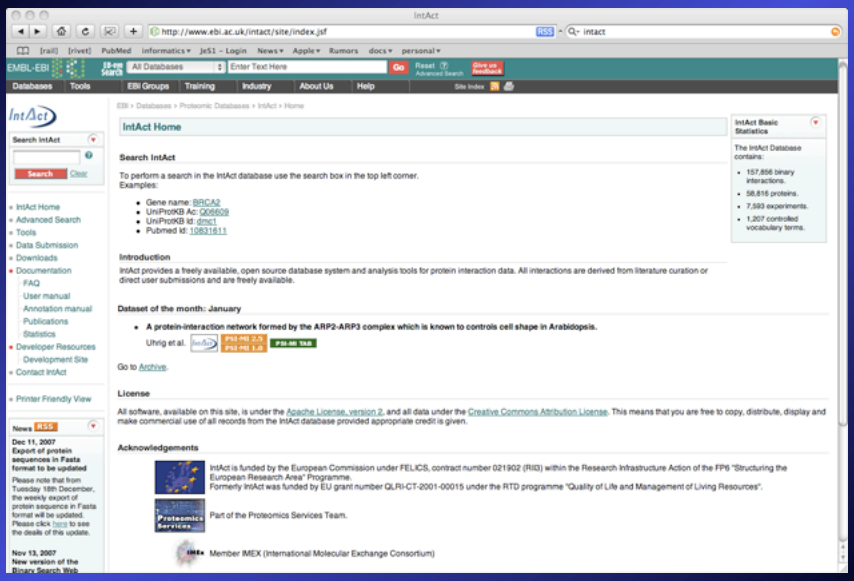
BioGRID- A Database of Genetic and Physical Interactions
DIP - Database of Interacting Proteins
MINT - A Molecular Interactions Database
IntAct - EMBL-EBI Protein Interaction
MIPS - Comprehensive Yeast Protein-Protein interactions
Yeast Protein Interactions - Yeast two-hybrid results from Fields' group
PathCalling- A yeast protein interaction database by Curagen
SPiD - Bacillus subtilis Protein Interaction Database
AllFuse - Functional Associations of Proteins in Complete Genomes
BRITE - Biomolecular Relations in Information Transmission and Expression
ProMesh - A Protein-Protein Interaction Database
The PIM Database - by Hybrigenics
Mouse Protein-Protein interactions
Human herpesvirus 1 Protein-Protein interactions
Human Protein Reference Database
BOND - The Biomolecular Object Network Databank. Former BIND
MDSP - Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometr
Protcom - Database of protein-protein complexes enriched with the domain-domain structures
Proteins that interact with GroEL and factors that affect their release
DPIDB - DNA-Protein Interaction Database
YPD™ - Yeast Proteome Database by Incyte

Source with links: http://proteome.wayne.edu/PIDBL.html

Armstrong, 2009

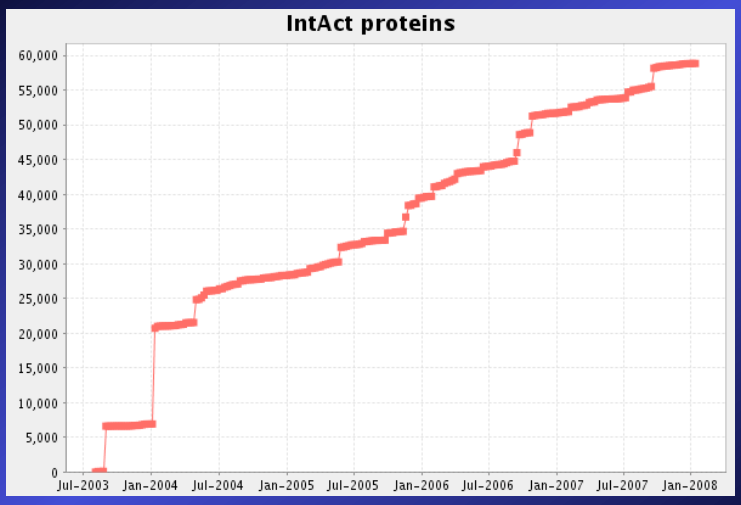


IntAct : www.ebi.ac.uk/intact



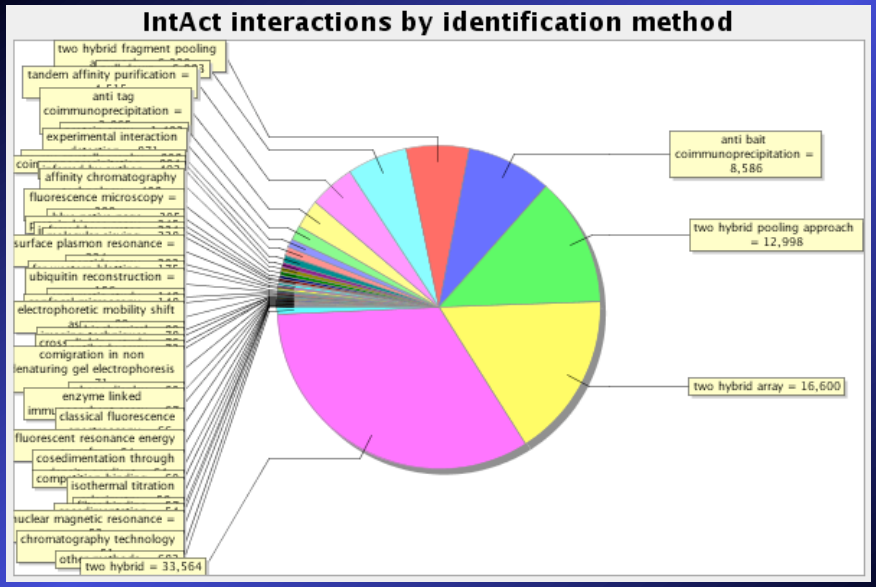
Armstrong, 2009

IntAct : www.ebi.ac.uk/intact



Armstrong, 2009

IntAct : www.ebi.ac.uk/intact

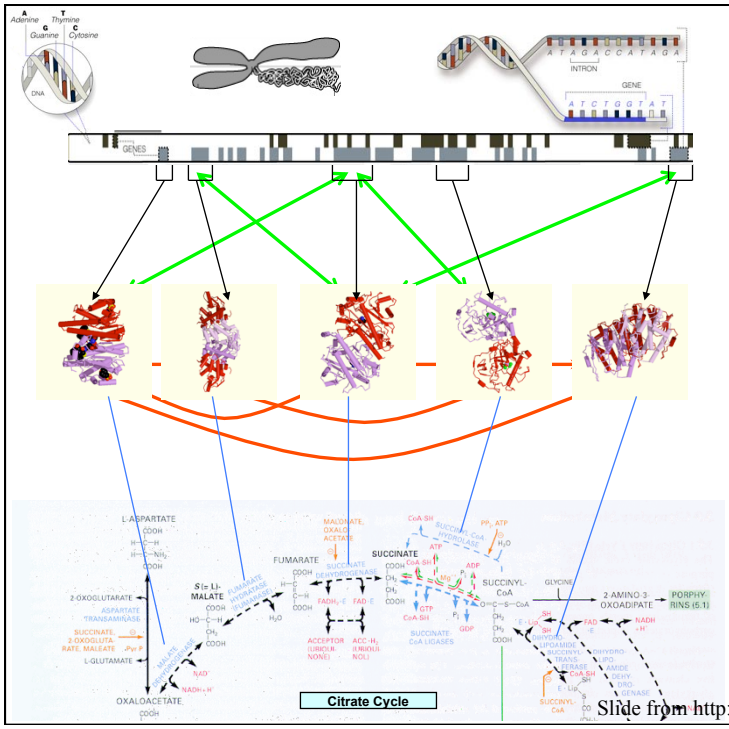


Armstrong, 2009

comparing two approaches

- Pocklington et al 2006
  - Emphasis on QC and literature mining
  - Focussed on subset of molecules
- Rual et al 2005
  - Emphasis on un-biased measurements
  - Focussed on proteome wide models
- Both then look at disease/network correlations

Armstrong, 2009



Slide from http://www.nd.edu/~networks/