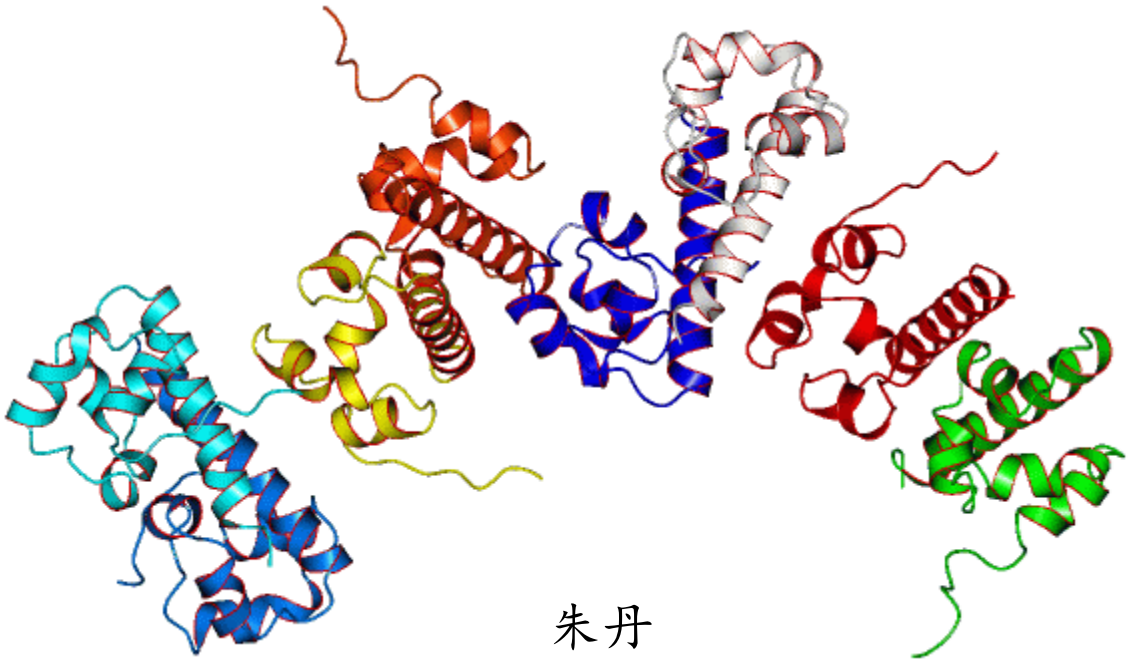
The interactome



朱丹

2010.12.10

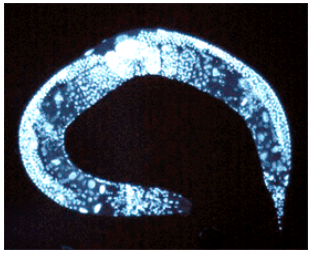
Outline

- Background
- Definition
- Research institutions and database
- The methods
- Classical papers
- Perspective

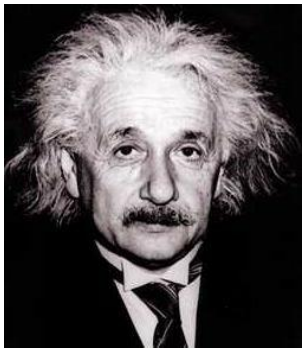
Why protein network?

- Proteins don't work in isolation !
- Assemblies represent more than the sum of their parts.

14,000 genes



13,500 genes



24,000 genes

25,000 genes

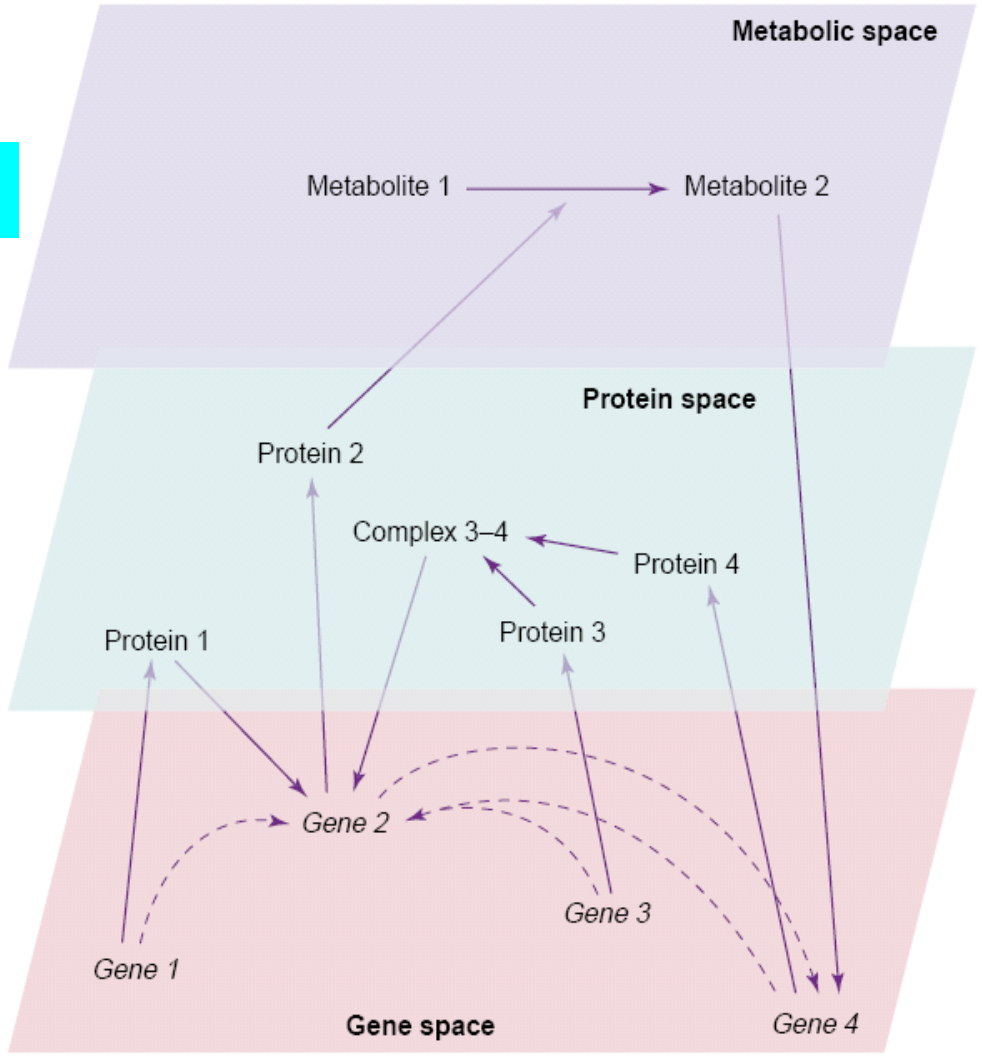
50,000 genes

What's the interactome?

Metabonomics

Proteomics

Genomics



TRENDS in Biotechnology

The definition of interactome

- The narrow sense

  A map of all the protein interactions within a given cell

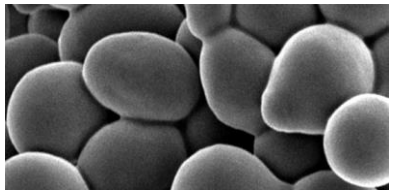
- It requires massive computational power
- It requires a high number of physical interaction data

- The broad sense

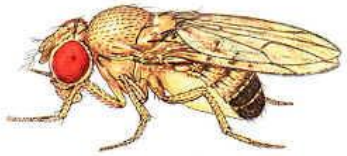
  The proteins interact with DNA、RNA 、sugar、lipid and small molecular in the whole cell

- It's Almost impossible to estimated

Classical papers



- *S. cerevisiae*
  - Schwikowski et al. (2000), Nature Biotechnol. 18, 1257
  - 1,548 proteins and 2,358 interactions



- *D. melanogaster*
  - Giot et al. (2003) Science 302:1727
  - 7048 proteins and 20405 interactions

- *C. elegans*
  - Li et al.(2004) Science 303:540
  - 1000 genes and 4000 interactions

- *H. sapiens*
  - Rual et al. (2005) Nature 437:1173
  - 8100 proteins and 2800 interactions

The institutions and database

- http://ppi.fli-leibniz.de/jcb_ppi_databases.html

Databases & Data Collections / Webtools

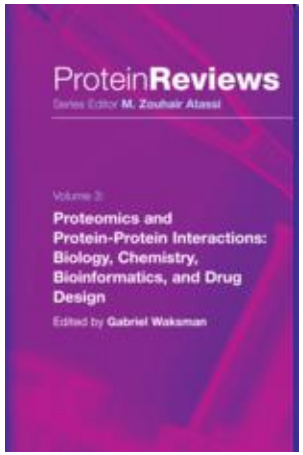
Last update: September 6, 2010

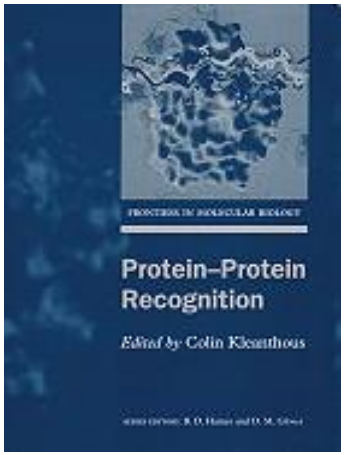
Databases & Data Collections

Experimental Data *Predictions* *Related Domain, Pathway and Network Databases*

Webtools **Books**

- Experimental Data
- Predictions
- Related Domain, Pathway and Network Databases
- Webtools
- Books





How to studying the interactome?

In vivo

- Protein fragment complementation assays
  - Yeast two hybrid
- Correlated mRNA expression
- Synthetic lethality

In vitro

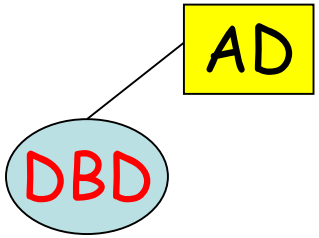
- Protein microarrays
- AP-MS
  - Tandem affinity purification (TAP)
  - HMS-PCI

In silico

- Gene co-expression
- Gene co-localization
- Functional category
- Gene ontology

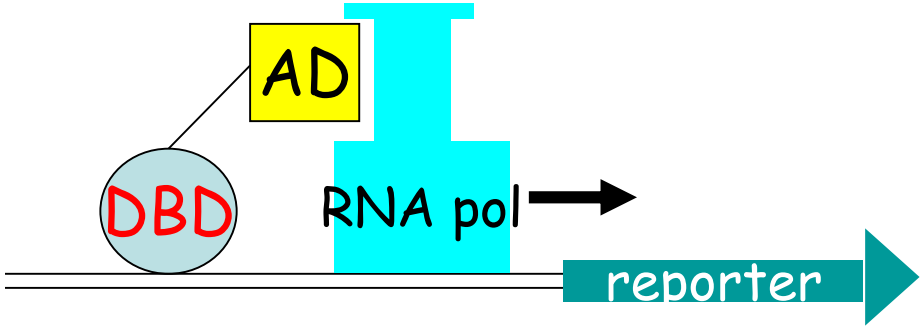
Yeast 2-Hybrid : *in vivo*

(review: Uetz 2002, Curr Op Chem Biol, 6:57)



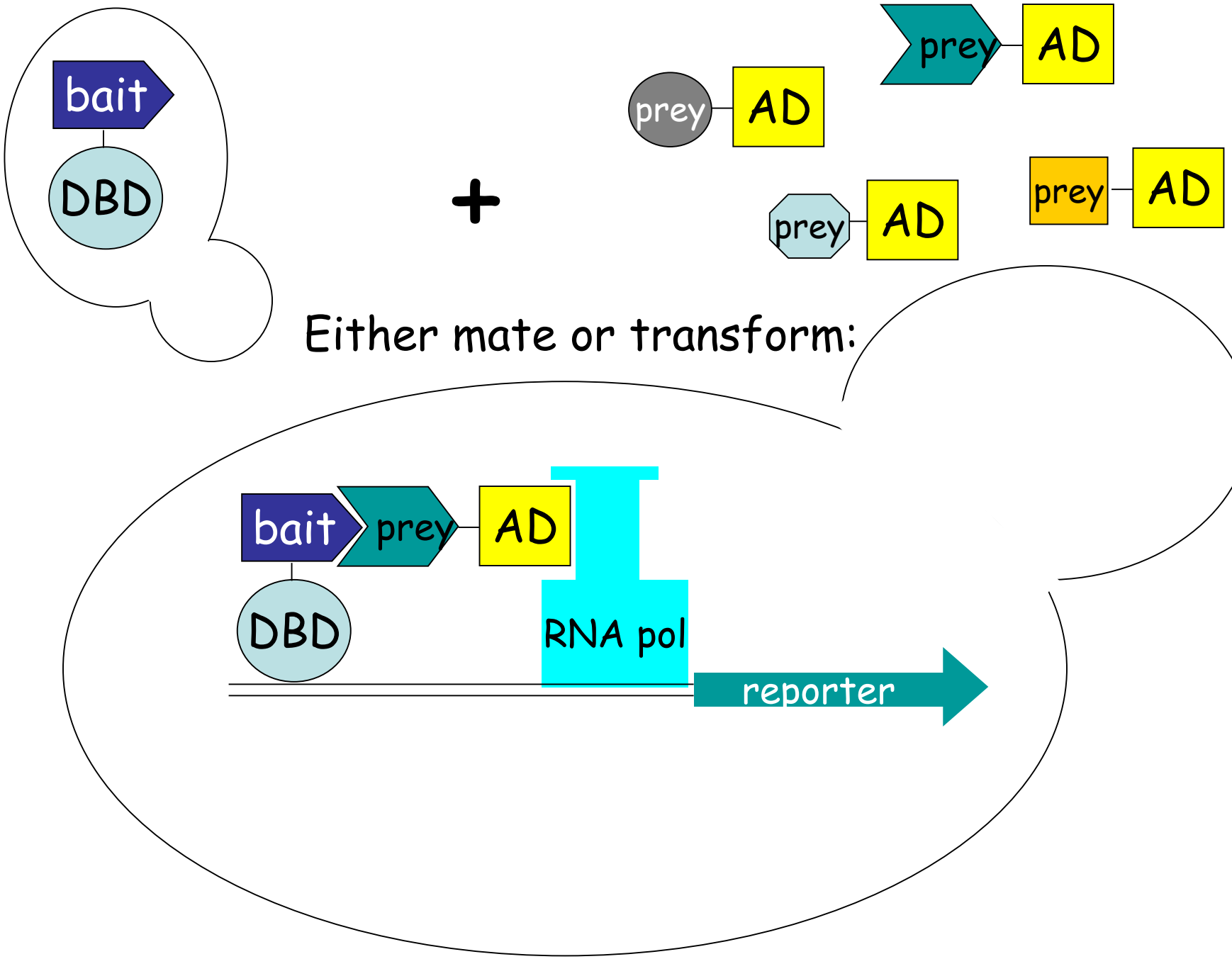
Transcription factor
DNA Binding Domain: binds promoter
Activation Domain: binds RNA polymerase

In the nucleus:



Reporter is expressed

If lac Z: blue colonies


bait
DBD
+
prey
AD
prey
AD
prey
AD
prey
AD
Either mate or transform:
bait
prey
AD
DBD
RNA pol
reporter

Pros and cons: Y2H

- In vivo: real proteins in a real cell
- Close range of interaction
- High throughput possible but not great (96-well plates)
- High throughput can eliminate false positives

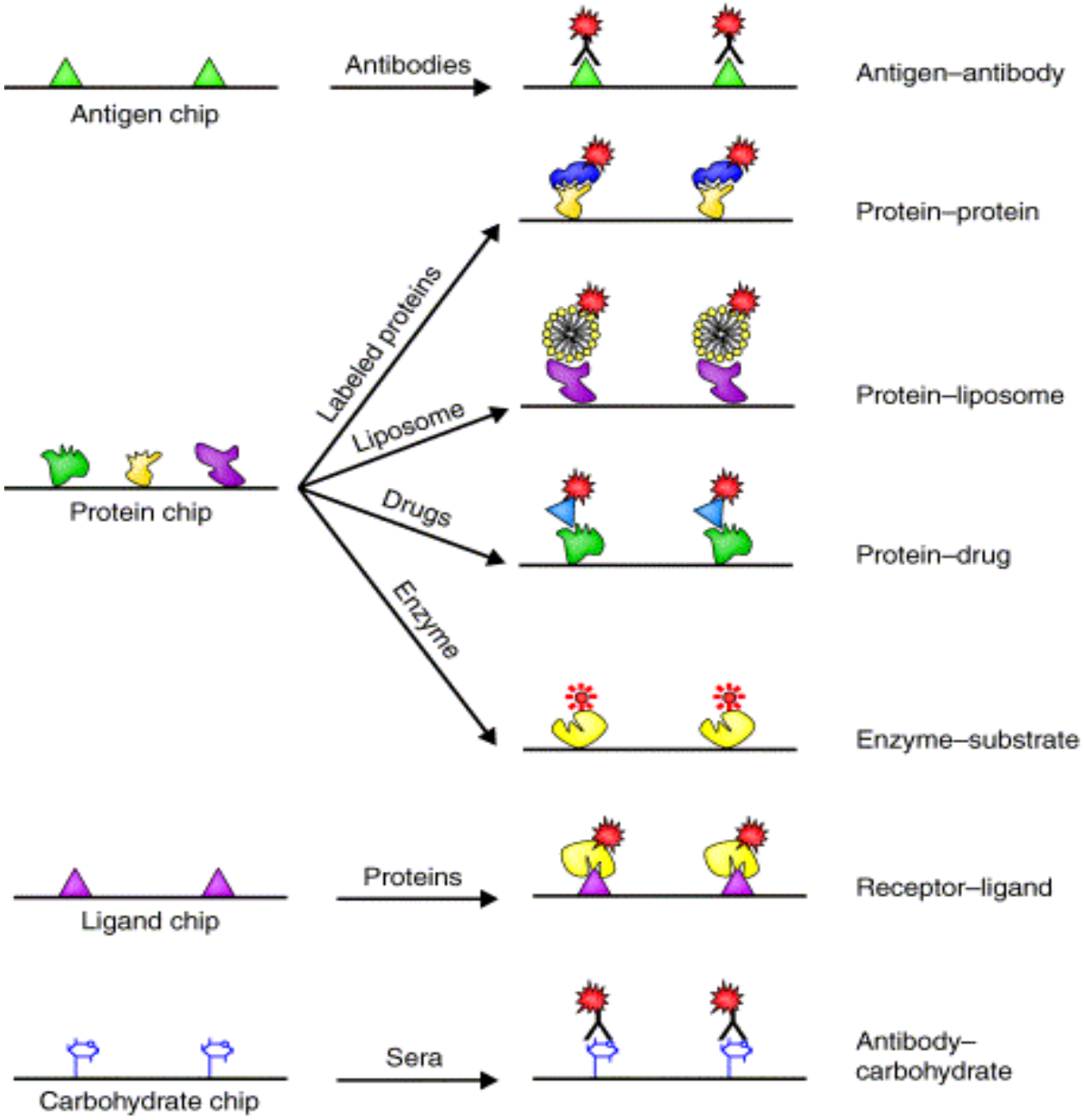
- Provides only binary information
  - What if the proteins are in a complex?
- Requires folding and transport to the nucleus
  - The hybrid protein may misfold/mistarget
- Limited to protein-protein interactions
- High number of false positives

(review: Uetz 2002, Curr Op Chem Biol, 6:57)

Protein microarrays : *in vitro*

- Grids that contain small amounts of purified proteins in a high density
- They can be screened for:
  - biochemical activity
  - Protein/protein interactions
  - Protein/DNA or protein/RNA interactions
  - Protein/ligand interactions

Zhu and Snider, Curr Op Chem Biol, 2003



Zhu and Snider, Curr Op Chem Biol, 2003

Pros and cons: Protein chips

- Very high throughput
- Can look for protein interactions with the widest range of ligands
- Microarrays of an entire eukaryotic proteome can be prepared and screened for diverse biochemical activities

- Need for high-throughput protein purification
- Chips are still difficult and expensive to make
- Low reproducibility at present, but improving fast

Zhu and Snider, Curr Op Chem Biol, 2003

Gene Ontology: *in silico*

- all : all (171472)
  - _ GO:0008150 : biological process (109503)
    - _ GO:0007582 : physiological process (70981)
      - _ GO:0008152 : metabolism (41395)
        - » _ GO:0009058 : biosynthesis (10256)
          - _ GO:0009059 : macromolecule biosynthesis (6876)
            - _ GO:0006412 : protein biosynthesis (4611)
        - » _ GO:0043170 : macromolecule metabolism (17198)
          - _ GO:0009059 : macromolecule biosynthesis (6876)
            - _ GO:0006412 : protein biosynthesis (4611)
          - _ GO:0019538 : protein metabolism (12856)
            - _ GO:0006412 : protein biosynthesis (4611)
  - _ GO:0005575 : cellular component (98453)
  - _ GO:0003674 : molecular function (108120)

Validation with annotated complexes

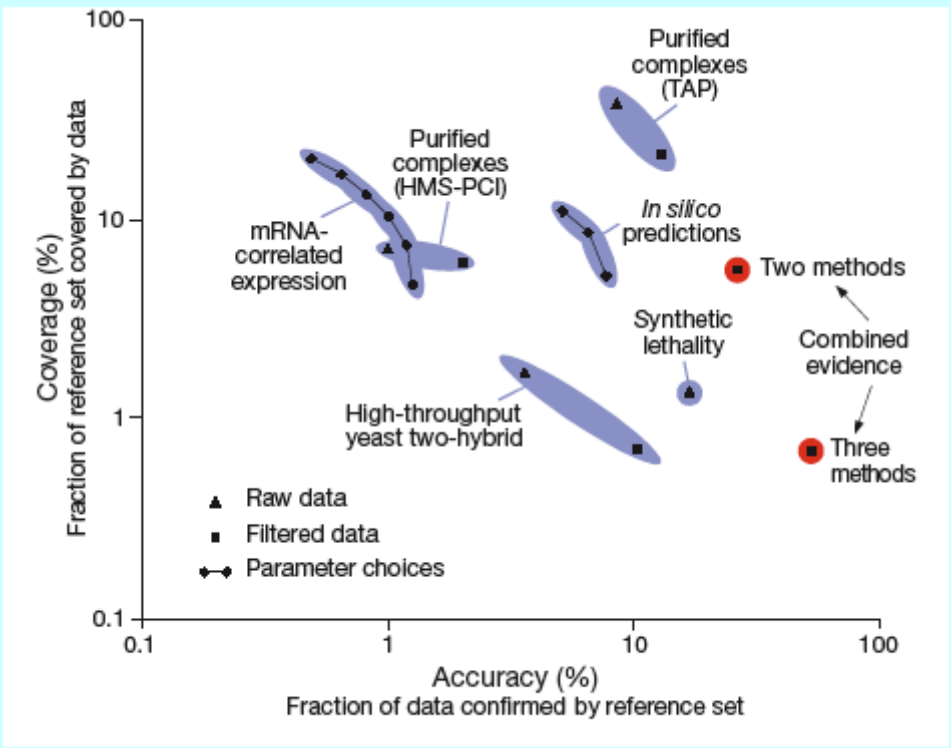


Figure 2 Quantitative comparison of interaction data sets. The various data sets are benchmarked against a reference set of 10,907 trusted interactions, which are derived from protein complexes annotated manually at MIPS[17] and YPD[24]. Coverage and accuracy are lower limits owing to incompleteness of the reference set. Each dot in the graph represents an entire interaction data set, and its position specifies coverage and accuracy (on a log–log scale). For the combined evidence, we considered only interactions supported by an agreement of two (or three) of any of the methods shown. For most data sets, raw and filtered data are shown, demonstrating the trade-off between coverage and accuracy achieved by filtering (see Supplementary Information for details on the filtering).

- the **coverage** and **positive predictive value** of each predictive method
  - Coverage
    - fraction of reference set covered by the data.
  - Positive predictive value
    - Fraction of data confirmed by reference set.

Von Mering et al (2002). Nature.

A network of protein–protein interactions in yeast

Benno Schwikowski[1,2*], Peter Uetz[3], and Stanley Fields[3,4]

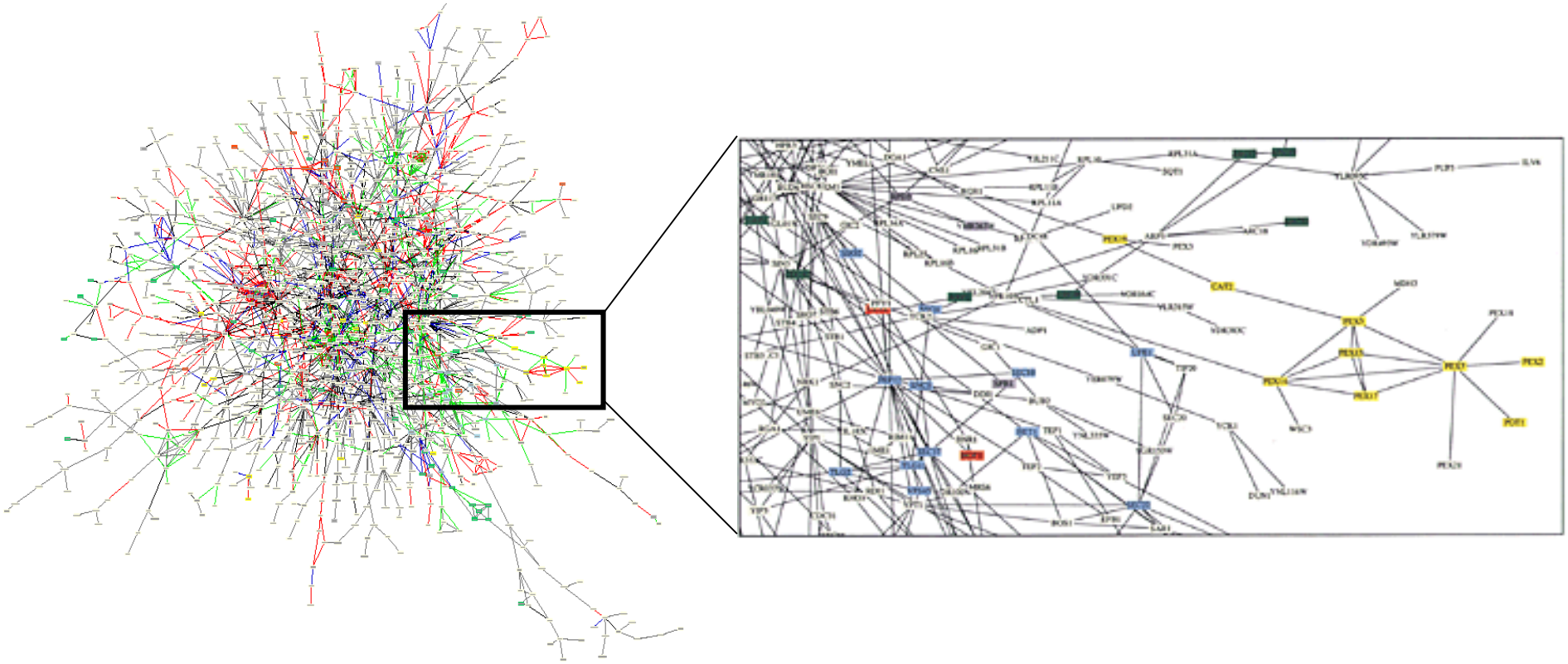
[1]The Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105. [2]Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195. [3]Departments of Genetics and Medicine, [4]Howard Hughes Medical Institute, University of Washington, Box 357360, Seattle, WA 98195. *Corresponding authors (benno@systemsbiology.org, uetz@u.washington.edu, fields@u.washington.edu).

Received 27 June 2000; accepted 13 October 2000

- → a single large network of 2,358 interactions among 1,548 proteins
- Source of data:
  - Y2H interactions databases
  - All previously published interactions from the entire community of yeast researchers (gathered in **Y**east **P**roteome **D**atabase)

Schwikowski et al., (2000), Nature Biotechnol

The interactome: *in silico*



- The map contains 1,548 proteins and 2,358 interactions
- Membrane fusion (blue), chromatin structure (gray), cell structure (green), lipid metabolism (yellow), and cytokinesis (red).

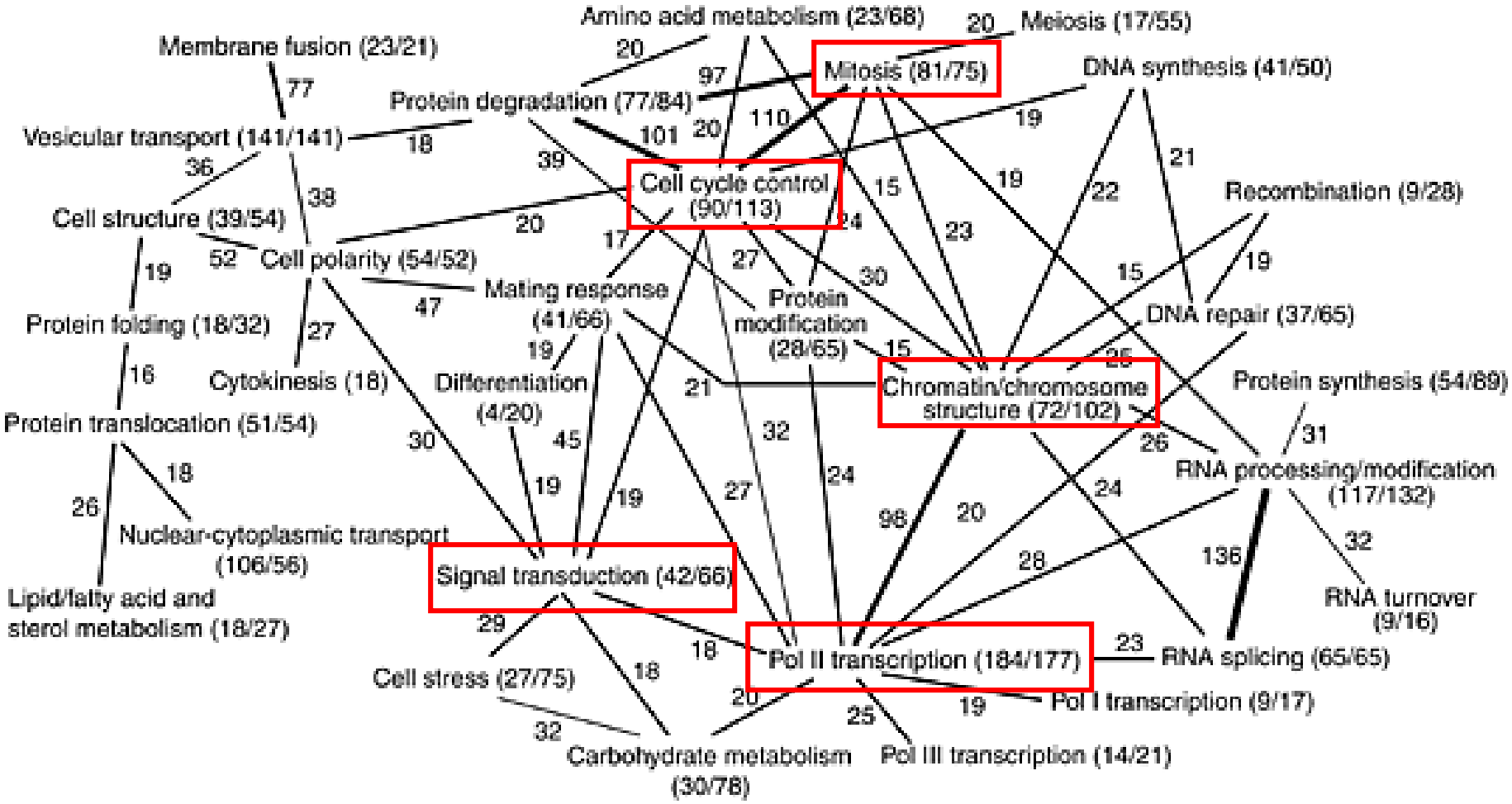
Schwikowski et al. (2000), Nature Biotechnol

The interactome: *in silico*

- Proteins of known function and cellular location tend to cluster together
- 63% of the interactions occur between proteins with a common functional assignment
- 76% occur between proteins found in the same subcellular compartment

Schwikowski et al. (2000), Nature Biotechnol

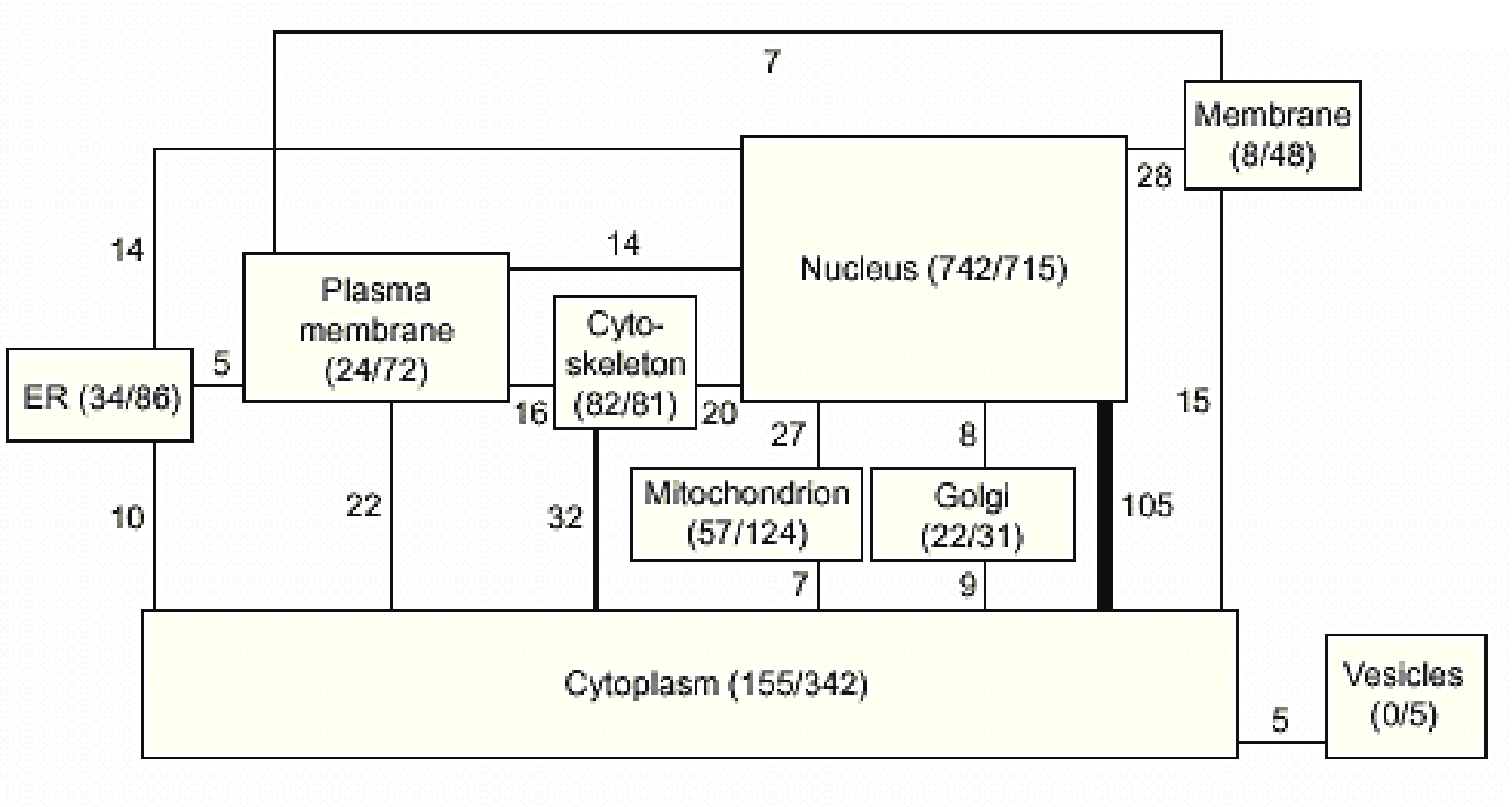
The interactome: *in silico*



Interactions between proteins of different function

Schwikowski et al. (2000), Nature Biotechnol

The interactome: *in silico*



Interactions between proteins of different compartments

Schwikowski et al. (2000), Nature Biotechnol

The interactome: *in silico*

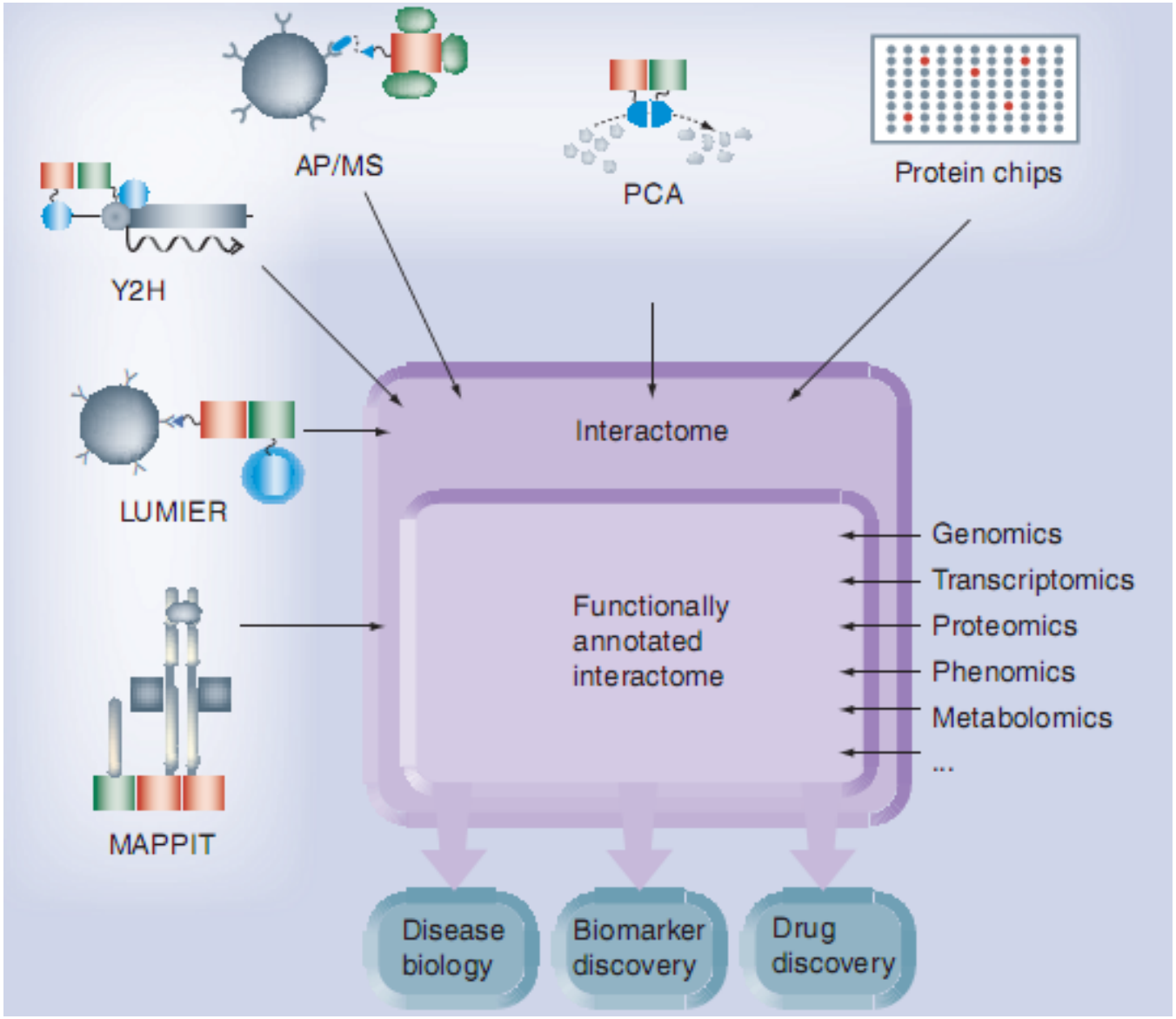
- Problems:
- quality of both the data and the database annotations
  - Interactions derived from genomic two-hybrid approaches are not normally backed up by additional experiments, and false positives are commonplace with this technique
  - Proteins which interact in a Y2H screen may not be present in the cell at the same time
- Membrane proteins are underrepresented
- Three quarters of the proteome are missing

Schwikowski et al. (2000), Nature Biotechnol

Conclusions

- Protein-protein interactions can be elucidated with a wide array of techniques
  - From one-to-one to whole proteome interactions
- Protein interaction data from different sources can be combined into networks of interactions – the interactome
- Interaction networks can be used to predict function of uncharacterised proteins
- The ultimate goal is to construct comprehensive, comparable and reliable networks

Perspective



Lievens, S., et. al 2010

Thanks for your attention !