

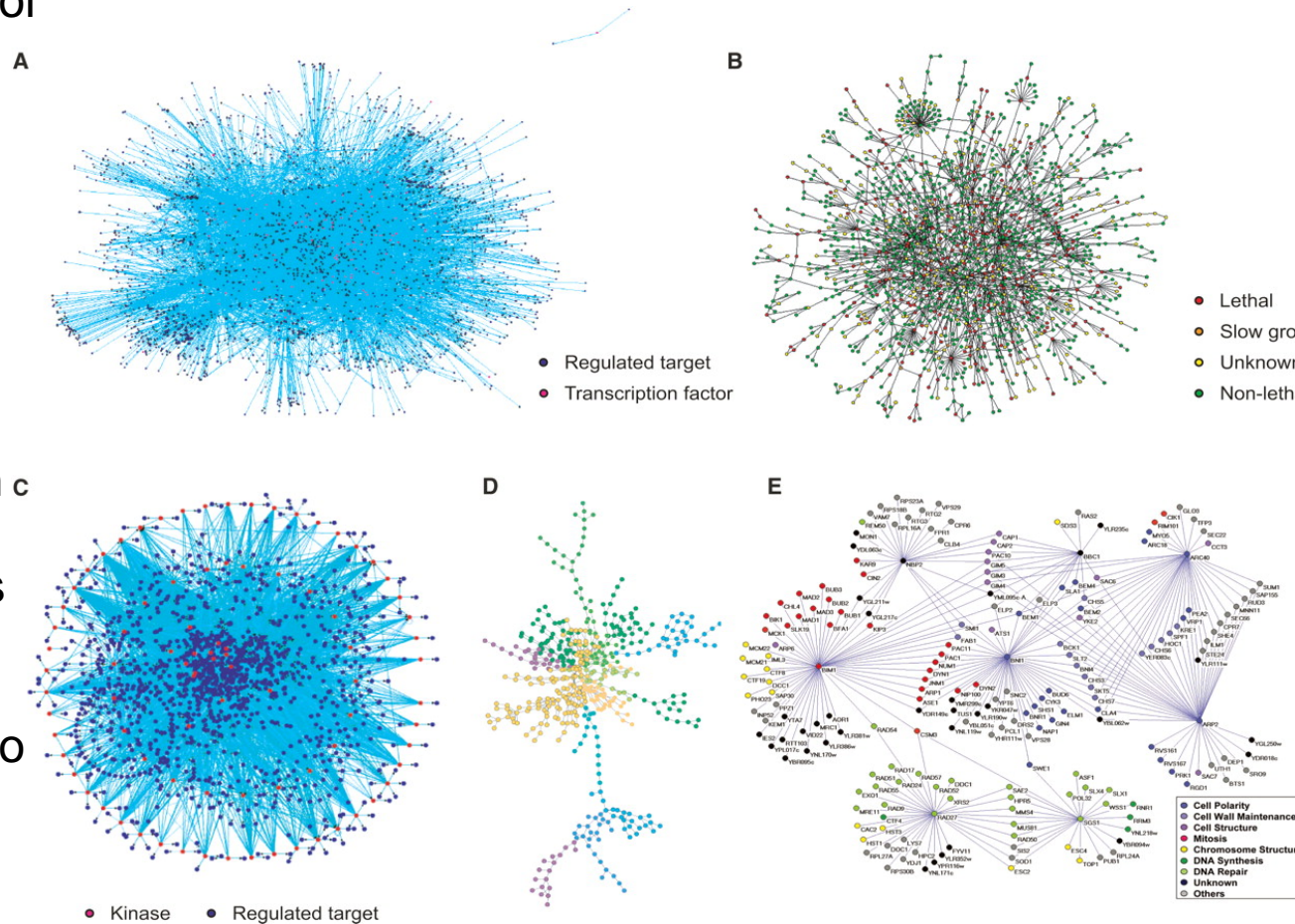
Netzwerke: Graphen zur Darstellung biologischer Information

- Graph besteht aus Knoten (nodes) und Kanten (edges)
- Kanten können Gewichte haben
- Interpretation:
 - Knoten = Gen oder Protein
 - Kante z.B. Regulation oder physikalische Interaktion

Examples of the five major biological networks.

A yeast transcription factor-binding network, composed of known transcription factor-binding data collected with large-scale ChIP–chip and small-scale experiments.

(B) A yeast protein–protein interaction network, containing protein–protein interactions identified by yeast two-hybrid and protein c complexes identified by affinity purification and mass spectrometry (C) A yeast phosphorylation network comprised primarily of in vitro phosphorylation events identified using protein microarrays (D) An E. coli metabolic network with 574 reactions and 473 metabolites colored according to their modules (E) A yeast genetic network constructed with synthetic lethal interactions using SGA analysis on eight yeast genes



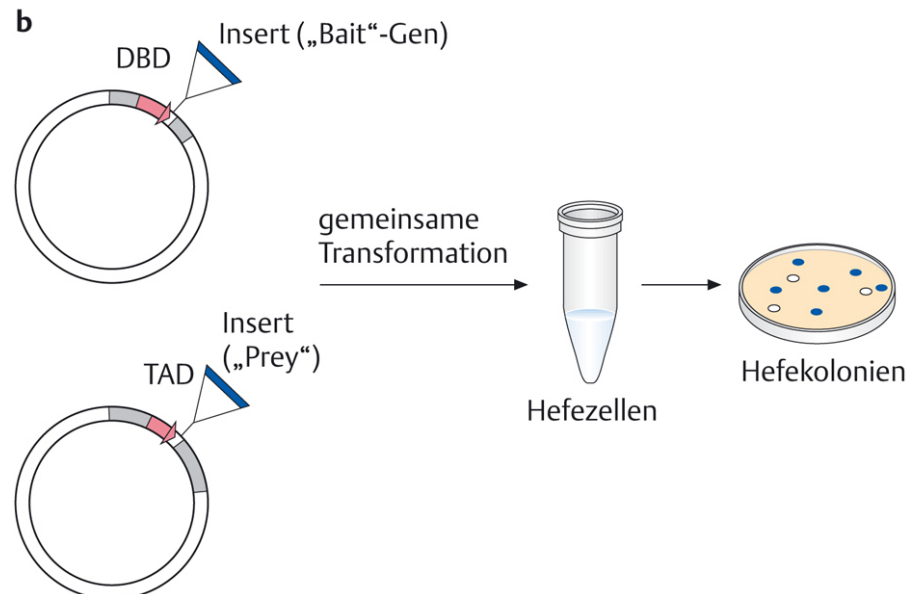
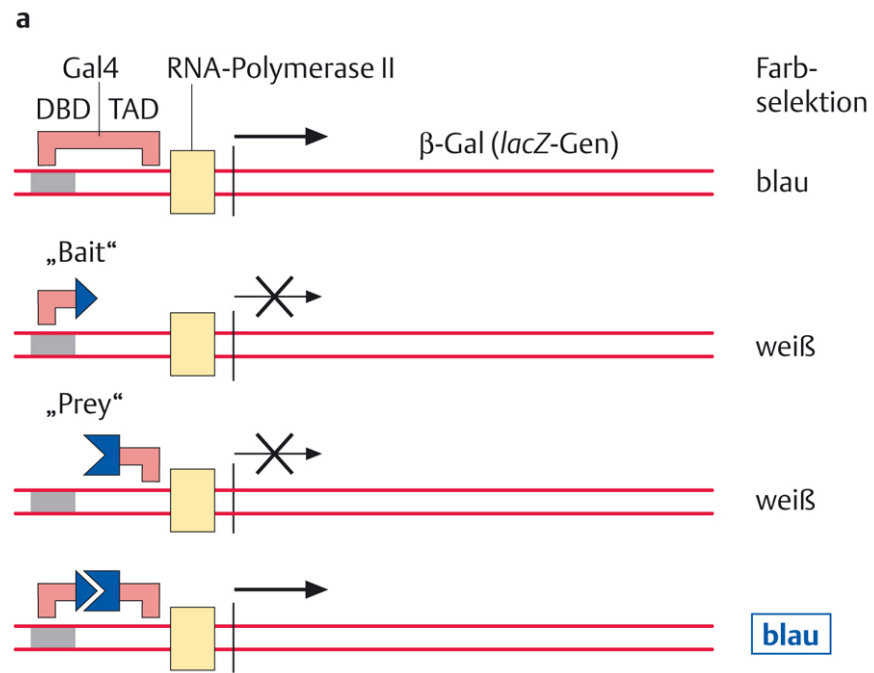
Zhu X et al. *Genes Dev.* 2007;21:1010-1024

Copyright © 2007, Cold Spring Harbor Laboratory Press



Yeast two-hybrid Experiment

Hefe 2-Hybrid System



STRING database

- STRING - Known and Predicted Protein-Protein Interactions
- STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:
 - Genomic Context
 - High-throughput Experiments
 - (Conserved) Coexpression
 - Previous Knowledge (literature)
- <http://string-db.org/>

Reverse Engineering von Gennetzwerken aus Expressionsdaten

- Gegeben: Genexpressionmatrix über viele Bedingungen
- Gesucht: Netzwerk zwischen den Genen
 - Koexpressionnetzwerk (Korrelation)
 - Netzwerk der Assoziationen (Mutual Information, u.a.)

Mutual Information

- Eine Methode zu lokalisieren kovariante Positionen in einem MSA ist **Mutual Information** von zwei Spalten.
- Formel zur Berechnung des mutual Information Content:

$$MI_{i,j} = \sum_{x,y} f(x_i, y_j) \log_2 \frac{f(x_i, y_j)}{f(x_i) f(y_j)}$$

- $f(x_i)$: Frequenz jedes Base x in $\{A,C,G,U\}$ in Spalte i des Alignments
- $f(x_i, y_j)$: zusammengeführte Frequenz zweier nukleotide x in Spalte i und y in Spalte j
- Ratio

Modules

- Modul in einem Netzwerk ist ein in sich eng verbundener und mit dem Rest schwach verbundener Teilgraph
- Vgl. Clustering
- Beobachtung: Gene in einem Modul haben oft gleiche Funktion

Many methods have been developed to identify possible network modules. A traditional method, hierarchical clustering, assigns a weight value to the distance between any two nodes in a network, and then gathers nodes with similar weight vectors together into strongly connected cores ([Rives and Galitski 2003](#)). Instead of detecting cores of modules in hierarchical clustering, the Girvan-Newman algorithm focuses on defining the boundaries of modules by searching for edges with high betweenness and therefore those that are more likely to link different modules ([Girvan and Newman 2002](#)). Other algorithms etc

Function prediction and Guilt-by-association

- Wenn ein Gen unbekannter Funktion mit Genen der gleichen bekannten Funktion verbunden ist
- Anwendung z.B. in der medizinischen Genetik: Mehrere mutierte Gene – welches könnte funktional wichtig sein?