

# V9 funktionelle Annotation

- **Analyse von Gen-Expression**
- **Funktionelle Annotation: Gene Ontology (GO)**
- **Signifikanz der Annotation: Hypergeometrischer Test**
- **Annotationsanalysen z.B. mit NIH-Tool DAVID**
- **Ähnlichkeit von GO-Termen automatisch bestimmen**
- **OMIM-Datenbank**

# Beispiel: differentielle Gen-Expression für ALL-Patienten

## Input:

Genexpressionsdaten für 128 Patienten mit akuter lymphatischer Leukämie (ALL).

Alle ALL-Patienten haben chromosomale Veränderungen.

Der Therapieerfolg ist sehr unterschiedlich.

Eine Gruppe von Patienten (ALL1/AF4) hat eine genetische Translokation zwischen den Chromosomen 4 und 11.

Eine zweite Gruppe von Patienten (BCR/ABL) hat eine genetische Translokation zwischen den Chromosomen 9 und 22.

Die Krankheitsursachen + optimale Therapie könnten für die beiden Gruppen verschieden sein.

## Ziel:

Identifiziere Gene, die zwischen den beiden Gruppen differentiell exprimiert werden.

Beispiel für die Anwendung der Bioconductor-Software (siehe Ref unten, bisher 6700 mal zitiert).

Gentleman et al. Genome Biology 5, R80 (2004)

# Auswahl der differentiell exprimierten Gene

```
> f <- factor(as.character(eset$mol))
> design <- model.matrix(~f)
> fit <- lmFit(eset, design)
> fit <- eBayes(fit)
> topTable(fit, coef = 2)
```

	ID	M	A	t	p-value	B	
	1016	1914_at	-3.1	4.6	-27	5.9e-27	56
	7884	37809_at	-4.0	4.9	-20	1.3e-20	44
	6939	36873_at	-3.4	4.3	-20	1.8e-20	44
	10865	40763_at	-3.1	3.5	-17	7.2e-18	39
	4250	34210_at	3.6	8.4	15	3.5e-16	35
	11556	41448_at	-2.5	3.7	-15	1.8e-15	34
	3389	33358_at	-2.3	5.2	-13	3.3e-13	29
	8054	37978_at	-1.0	6.9	-10	6.5e-10	22
	10579	40480_s_at	1.8	7.8	10	9.1e-10	21
	330	1307_at	1.6	4.6	10	1.4e-09	21

## Figure 1

Limma analysis of the ALL data. The leftmost numbers are row indices, ID is the Affymetrix HGU95av2 accession number, M is the log ratio of expression, A is the log average expression, and B is the log odds of differential expression.

Differential expression (D.E.) =  $\log(R) / \log(G)$

Log ratio M :  $2^M = \log(R) / \log(G)$ ; M = 1 -> zweifach D.E.

Wie signifikant ist dies? -> bewerte mit statistischem Test.

Vergleiche Gen-Expression in den beiden Gruppen.

Fokussiere auf Gene mit stark unterschiedlicher Expression.

Wähle alle Gene mit p-Wert < 0.05 aus.

Es bleiben 165 Gene übrig.

Gentleman et al. Genome Biology 5, R80 (2004)

# Differentielle Gen-Expression als Heatmap visualisieren

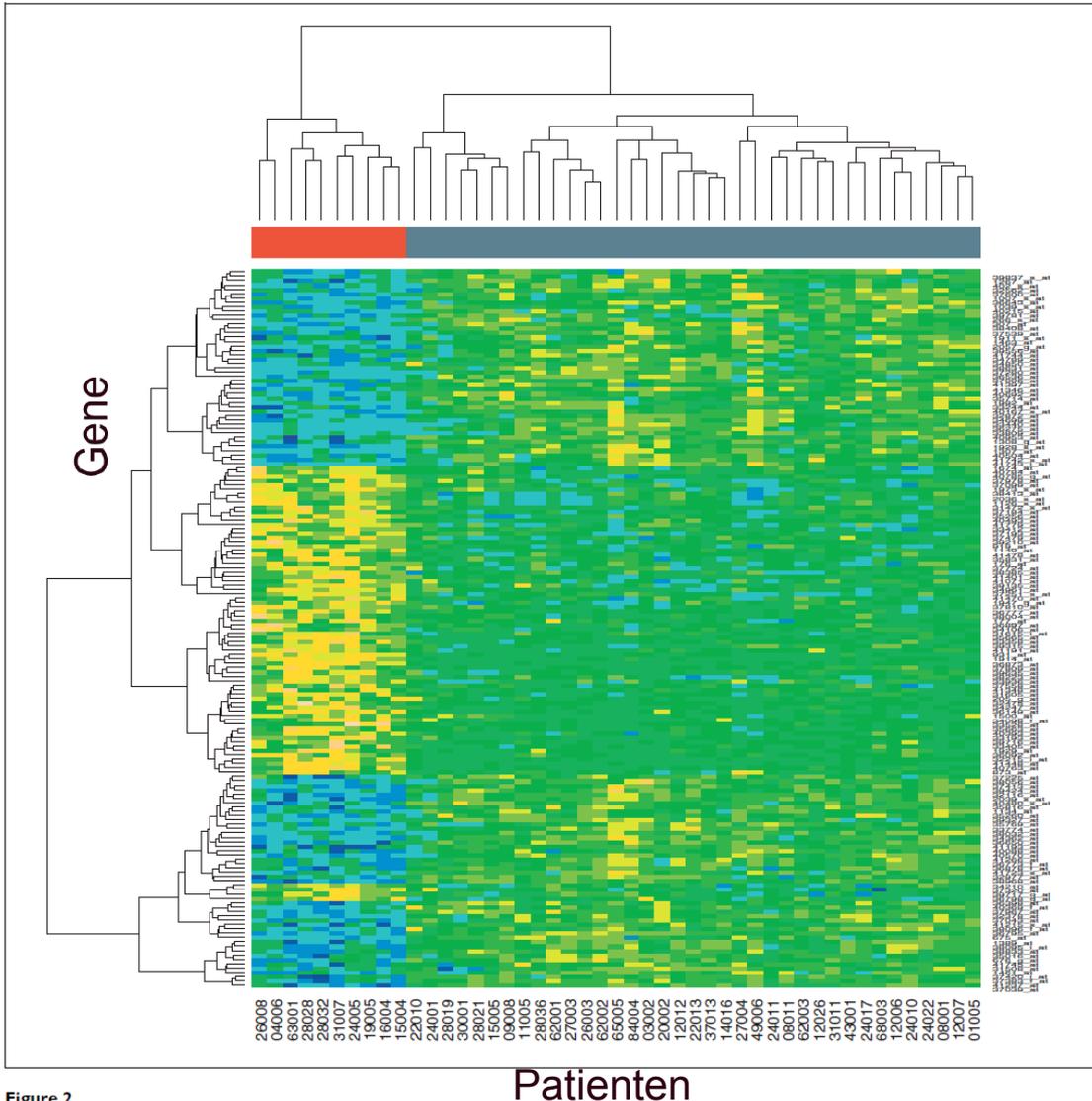


Figure 2  
Heat map (produced by the Bioconductor function `heatmap()`) of the ALL leukemia data.

Mit einem Abstandsmaß und einem Cluster-Algorithmus werden die Ähnlichkeiten zwischen den Patienten (x-Achse) und den einzelnen Genen (y-Achse) erfasst.

Die beiden Patienten-Gruppen lassen sich deutlich unterscheiden (rot/grau).

Gelb: stark hochreguliert  
Blau: stark runterreguliert

Gentleman et al. Genome Biology 5, R80 (2004)

# Zuordnung von Gen-Funktion

```
> f <- factor(as.character(eset$mol))
> design <- model.matrix(~f)
> fit <- lmFit(eset, design)
> fit <- eBayes(fit)
> topTable(fit, coef = 2)
```

Bioconductor  
Kommandos

	ID	M	A	t	<i>p</i> -value	B
1016	1914_at	-3.1	4.6	-27	5.9e-27	56
7884	37809_at	-4.0	4.9	-20	1.3e-20	44
6939	36873_at	-3.4	4.3	-20	1.8e-20	44
10865	40763_at	-3.1	3.5	-17	7.2e-18	39
4250	34210_at	3.6	8.4	15	3.5e-16	35
11556	41448_at	-2.5	3.7	-15	1.8e-15	34
3389	33358_at	-2.3	5.2	-13	3.3e-13	29
8054	37978_at	-1.0	6.9	-10	6.5e-10	22
10579	40480_s_at	1.8	7.8	10	9.1e-10	21
330	1307_at	1.6	4.6	10	1.4e-09	21

## Figure 1

Limma analysis of the ALL data. The leftmost numbers are row indices, ID is the Affymetrix HGU95av2 accession number, M is the log ratio of expression, A is the log average expression, and B is the log odds of differential expression.

Links gezeigt ist dieselbe Tabelle wie zwei Folien zuvor.

Neue Frage: welche Funktionen üben diese Gene in der Zelle aus?

Verwende dazu die Gene Ontology.

Gentleman et al. Genome Biology 5, R80 (2004)

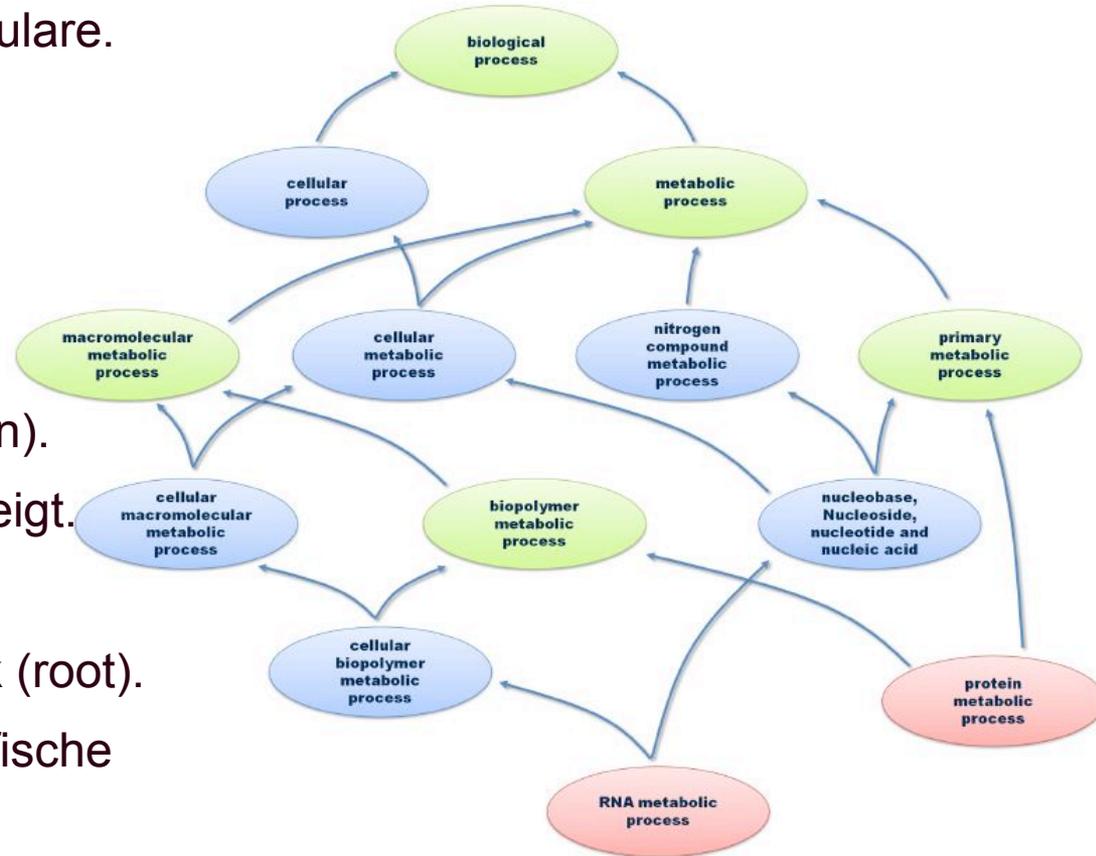
# Die Gene Ontology (GO)

Ontologien sind strukturierte Vokabulare.

Die Gene Ontology hat 3 Bereiche:

- biologischer Prozess (BP)
- molekulare Funktion (MF)
- zelluläre Komponente (Lokalisation).

Hier ist ein Teil des BP-Baums gezeigt.



Oben ist der allgemeinste Ausdruck (root).

Rot: Blätter des Baums (sehr spezifische GO-Terme)

Grün: gemeinsame Vorgänger.

Blau: andere Knoten.

Linien: „Y ist in X enthalten“-Beziehungen

Dissertation Andreas Schlicker (UdS, 2010)

# Signifkanz von GO-Annotationen

Sehr allgemeine GO-Terme wie z.B. “cellular metabolic process“ werden vielen Genen zugeordnet.

Sehr spezielle Terme gehören jeweils nur zu wenigen Genen.

Man muss also vergleichen, wie signifikant das Auftreten jedes GO-Terms in einer Testmenge an Genen im Vergleich zu einer zufällig ausgewählten Menge an Genen derselben Größe ist.

Dazu verwendet man meist den **hypergeometrischen Test**.

# Vorbemerkung

Zieht man aus einer Urne mit  $n$  Kugeln insgesamt  $k$  Kugeln **ohne Beachtung der Reihenfolge**, so gibt es hierfür genau

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{k!} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k} \text{ Möglichkeiten}$$

# Hypergeometrischer Test

$$\text{p-Wert} = \sum_{i=k_{\pi}}^{\min(n, K_{\pi})} \frac{\binom{K_{\pi}}{i} \binom{N-K_{\pi}}{n-i}}{\binom{N}{n}}$$

Der hypergeometrische Test ist ein statistischer Test, der z.B. überprüft, ob in einer vorgegebenen Testmenge an Genen eine biologische Annotation  $\pi$  gegenüber dem gesamten Genom statistisch signifikant angereichert ist.

- Sei  $N$  die Anzahl an Genen im Genom.
- Sei  $n$  die Anzahl an Genen in der Testmenge.
- Sei  $K_{\pi}$  die Anzahl an Genen im Genom mit der Annotation  $\pi$ .
- Sei  $k_{\pi}$  die Anzahl an Genen in der Testmenge mit der Annotation  $\pi$ .

Der hypergeometrische p-Wert drückt die Wahrscheinlichkeit aus, dass  $k_{\pi}$  oder mehr **zufällig** aus dem Genom ausgewählte Gene auch die Annotation  $\pi$  haben.

# Hypergeometrischer Test

Wähle  $i = k_\pi$  Gene mit  
Annotation  $\pi$  aus dem Genom.  
Davon gibt es genau  $K_\pi$ .

Die anderen  $n - i$  Gene in der  
Testmenge haben dann nicht die  
Annotation  $\pi$ . Davon gibt es im Genom  
genau  $N - K_\pi$ .

$$\text{p-Wert} = \sum_{i=k_\pi}^{\min(n, K_\pi)} \frac{\binom{K_\pi}{i} \binom{N - K_\pi}{n - i}}{\binom{N}{n}}$$

Die Summe läuft von mindestens  
 $k_\pi$  Elementen bis zur maximal  
möglichen Anzahl an Elementen.

Eine Obergrenze ist durch die  
Anzahl an Genen mit Annotation  $\pi$   
im Genom gegeben ( $K_\pi$ ).

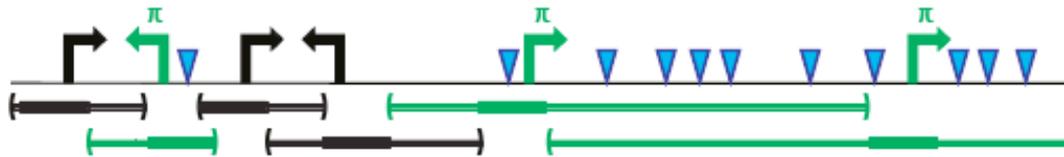
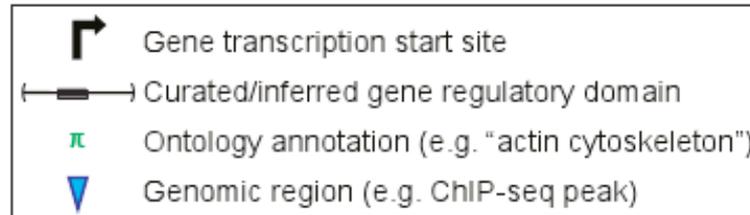
Die andere Obergrenze ist die Zahl  
der Gene in der Testmenge ( $n$ ).

Korrigiert für die kombinatorische  
Vielfalt an Möglichkeiten um  $n$   
Elemente aus einer Menge mit  $N$   
Elementen auszuwählen.

N.B. dies gilt für den Fall, dass  
die Reihenfolge der Elemente  
egal ist.

# Beispiel

$$\text{p-Wert} = \sum_{i=k_{\pi}}^{\min(n, K_{\pi})} \frac{\binom{K_{\pi}}{i} \binom{N-K_{\pi}}{n-i}}{\binom{N}{n}}$$



Hypergeometric test over genes

$N$  = 6 total genes

$K_{\pi}$  = 3 genes annotated with  $\pi$

$n$  = 3 genes with an associated genomic region

$k_{\pi}$  = 3 genes annotated and with a genomic region

P-value = 0.05

Frage: ist die Annotation  $\pi$  in der Testmenge von 3 Genen signifikant angereichert?

Ja!  $p = 0.05$  ist signifikant.

<http://great.stanford.edu/>

# Messe funktionelle Ähnlichkeit von GO-Termen

Die **Wahrscheinlichkeit eines Knoten**  $t$  kann man auf 2 Arten ausdrücken:

Wieviele Gene besitzen die  
Annotation  $t$  relativ zur Häufigkeit  
der Wurzel?

Anzahl an GO-Termen im bei  $t$   
startenden Unterbaum relativ zu der  
Anzahl an GO-Termen im Gesamtbaum.

$$P_{anno}(t) = \frac{occur(t)}{occur(root)}$$

$$P_{graph}(t) = \frac{D(t)}{D(root)}$$

Die Wahrscheinlichkeit hat Werte zwischen 0 und 1 und nimmt zwischen den Blättern bis zur Wurzel monoton zu.

Aus der Wahrscheinlichkeit  $p$  berechnet man den **Informationsgehalt** jedes Knotens:

$$IC(t) = -\log p(t)$$

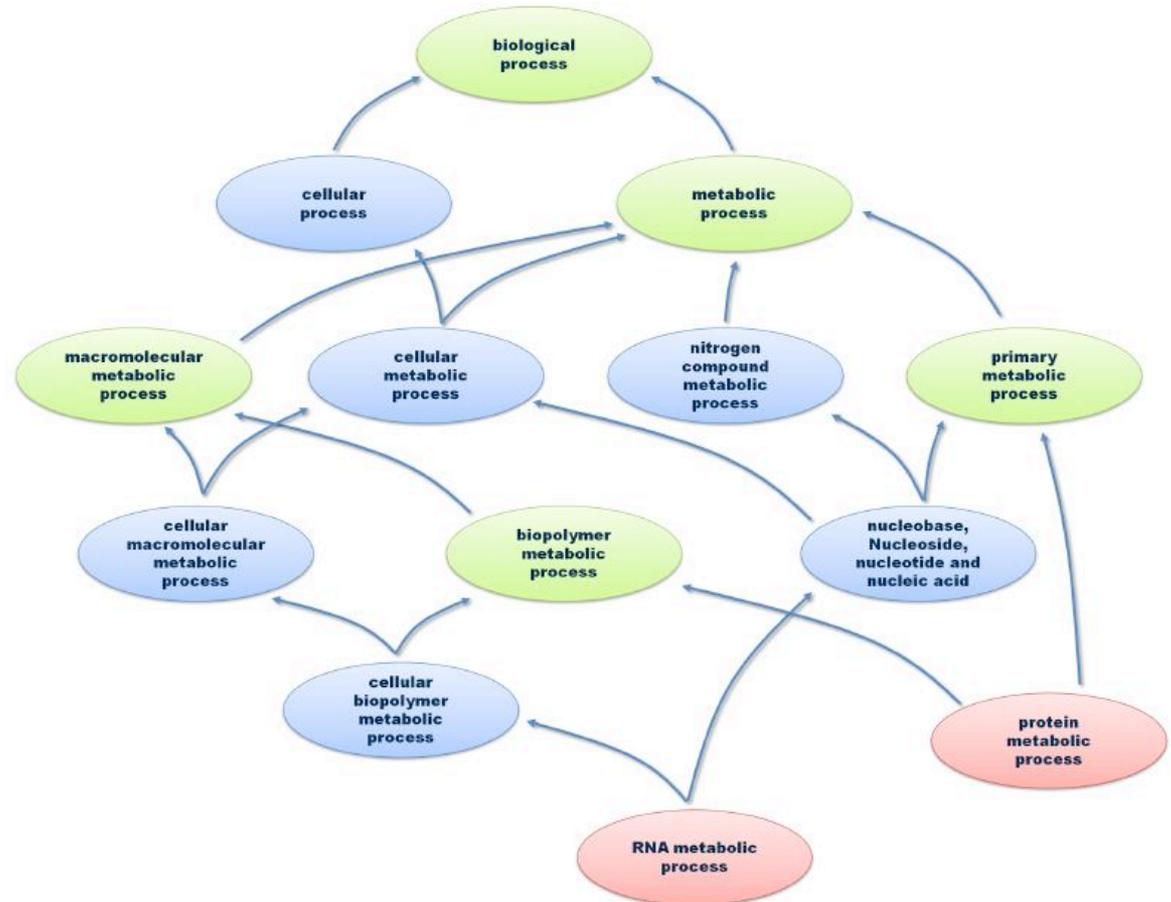
Je seltener ein Knoten ist, desto höher sein Informationsgehalt.

# Messe funktionelle Ähnlichkeit von GO-Termen

Die Menge an gemeinsamen Vorgängern (common ancestors (CA) ) zweier Knoten  $t_1$  und  $t_2$  enthält alle Knoten, die auf einem Pfad von  $t_1$  zum Wurzelknoten **UND** auf einem Pfad von  $t_2$  zum Wurzelknoten liegen.

Der **most informative common ancestor (MICA)** der Terme  $t_1$  und  $t_2$  ist der Term mit dem höchsten Informationsgehalt in CA.

Normalerweise ist das der gemäß dem Abstand nächste gemeinsame Vorgänger.



# Messe funktionelle Ähnlichkeit von GO-Termen

Schlicker et al. definierten aus dem Abstand zum **most informative common ancestor** (MICA) die Ähnlichkeit der Terme  $t_1$  und  $t_2$

$$sim_{Rel}(t_1, t_2) = \frac{2 \cdot IC(MICA)}{IC(t_1) + IC(t_2)} \cdot (1 - p(MICA))$$

Der hintere Faktor gewichtet die Ähnlichkeit mit der Häufigkeit  $p(MICA)$ . Dies ergab Vorteile in der Praxis.

# Zusammenfassung

Daten aus Microarray-Analyse wurden ursprünglich als sehr „verrauscht“ angesehen.

Mittlerweile wurden jedoch sowohl die experimentellen Schritte wie auch die Datenauswertung gründlich verfeinert.

Microarray-Analyse ist daher heute eine (zwar teure, aber zuverlässige) Routine-Methode, die in allen großen Firmen verwendet wird.

Die Datenaufbereitung kann folgende Schritte enthalten:

Normalisierung, Logarithmierung, Clustering, evtl. Ko-Expressionsanalyse, Annotation der Genfunktion

Sehr wichtig ist es, die Signifikanz der Ergebnisse zu bewerten.

Gentleman et al. Genome Biology 5, R80 (2004)

# V10 metabolische Netzwerke / Protein-Interaktionsnetzwerke

- Methoden der Systembiologie ermöglichen die integrierte, simultane Betrachtung von kompletten metabolischen Netzwerken.
- **wichtige Fragen:**
  - (1) wie soll man die Gesamtheit der metabolischen Flüsse im Netzwerk beschreiben?
  - (2) wie können mathematische Techniken die gezielte Manipulation von Mikroorganismen im Hinblick auf die Produktion bestimmter Substanzen unterstützen?
- **Inhalt V10:**
  - (1) stöchiometrische Matrix
  - (2) Flux Balance Analysis - Methode
  - (3) Elementary Flux Mode – Analyse (anschaulich + Anwendung)
  - (4) Aufbau und Visualisierung von Protein-Protein-Interaktionsnetzwerken

# Flux Balance Analyse

Jede chemische Reaktion erfüllt die **Erhaltung der Massen**.

Mit dieser grundlegenden Bedingung können metabolische Systeme untersucht werden.

Man muss lediglich die Stöchiometrie aller metabolischen Pfade und die metabolischen Anforderungen kennen. Für jeden Metabolit gilt:

$$V_i = \frac{dX_i}{dt} = V_{\text{synthesized}} - V_{\text{degraded}} - (V_{\text{used}} - V_{\text{transported}})$$

Im Gleichgewicht kann man die Massenerhaltung in einem metabolischen Netzwerk mathematisch durch eine Matrixgleichung darstellen:

$$S \cdot v = 0$$

Hierbei ist die Matrix  $S$  wiederum die  $m \times n$  **stöchiometrische Matrix**,  
 $m$  = Anzahl der Metabolite,  $n$  = Anzahl der Reaktionen im Netzwerk.

Der Vektor  $v$  beinhaltet alle Flüsse des metabolischen Netzwerks,  
nämlich die internen Flüsse, die Transportflüsse, und das Zellwachstum.

# Lösungsraum der metabolischen Flüsse

Da die Zahl an Metaboliten gewöhnlich kleiner ist als die Zahl an Reaktionen ( $m < n$ ) ist die Flussgleichung üblicherweise **unterbestimmt**.

Daher gibt es im Allgemeinen eine Vielzahl an erlaubten Flussverteilungen, die die Massenerhaltung erfüllen. Dieser Satz von Lösungen entspricht dem **Nullraum** der Matrix  $S$ .

Um den „wahren“ biologischen Fluss zu finden ( $\rightarrow$  e.g. Heinzle/UdS) benötigt man zusätzliche Informationen, z.B. aus Experimenten mit  $C^{13}$ -Markierung.

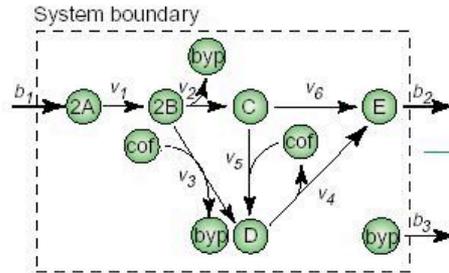
Alternativ kann man Grenzwerte  $\alpha_i \leq v_i \leq \beta_i$

für den Minimal- und Maximalbetrag jedes einzelnen metabolischen Flusses vorschreiben.

Der Durchschnitt des Nullraums und des durch diese linearen Ungleichungen eingeschränkten Raums definiert dann ein Volumen im hochdimensionalen Flussraum, das alle möglichen Flussverteilungen enthält.

# Berechnung von Elementaren Flussmoden aus der stöchiometrischen Matrix

Darstellung des Reaktionsnetzwerks mit stöchiometrischer Matrix S.

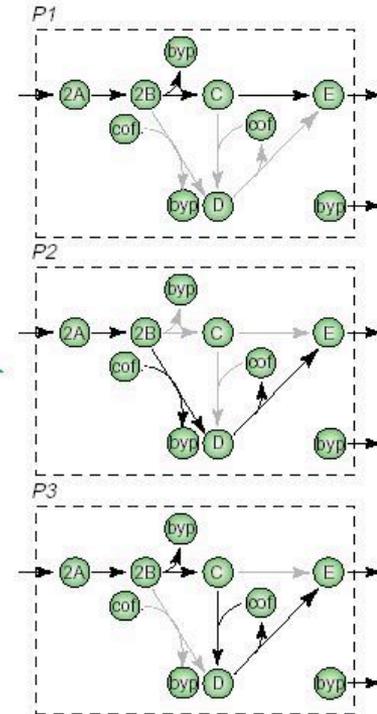


$$S = \begin{array}{cccccc|ccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 \\ \hline -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ +1 & -2 & +2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & +1 & -1 & 0 & 0 & 0 & 0 \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ \text{byp} \\ \text{cof} \end{array}$$

Metabolite

Analyse der Matrix S → Pathway-Darstellung P. Deren Zeilen enthalten den Reaktionen entsprechende Flüsse und die Spalten die sich ergebenden Pfade.

$$P = \begin{array}{ccc|cccc} P_1 & P_2 & P_3 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ \hline 2 & 2 & 2 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -1 & 1 & 0 & 2 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & b_1 & b_2 & b_2 \end{array}$$



Elementare Flussmoden für dies System.

Papin *et al.* TIBS 28, 250 (2003)

# Protein-Interaktionsnetzwerk in *S. cerevisiae*

Abb. zeigt das Protein–Protein Interaktionsnetzwerk in *Saccharomyces cerevisiae*, basierend auf **yeast two-hybrid** Experimenten.

Knoten: einzelne Proteine

Verbindungen/Kanten: physikalische Interaktionen.

Das Netzwerk wird von einigen Knoten mit sehr vielen Verbindungen zusammengehalten.

Diese nennt man **Hubs**.

Dieser größte Cluster enthält 78% aller Proteine.

Rot: Gendeletion ist tödlich

Grün: Gendeletion ist nicht tödlich

Gelb: Effekt der Gendeletion ist unbekannt.



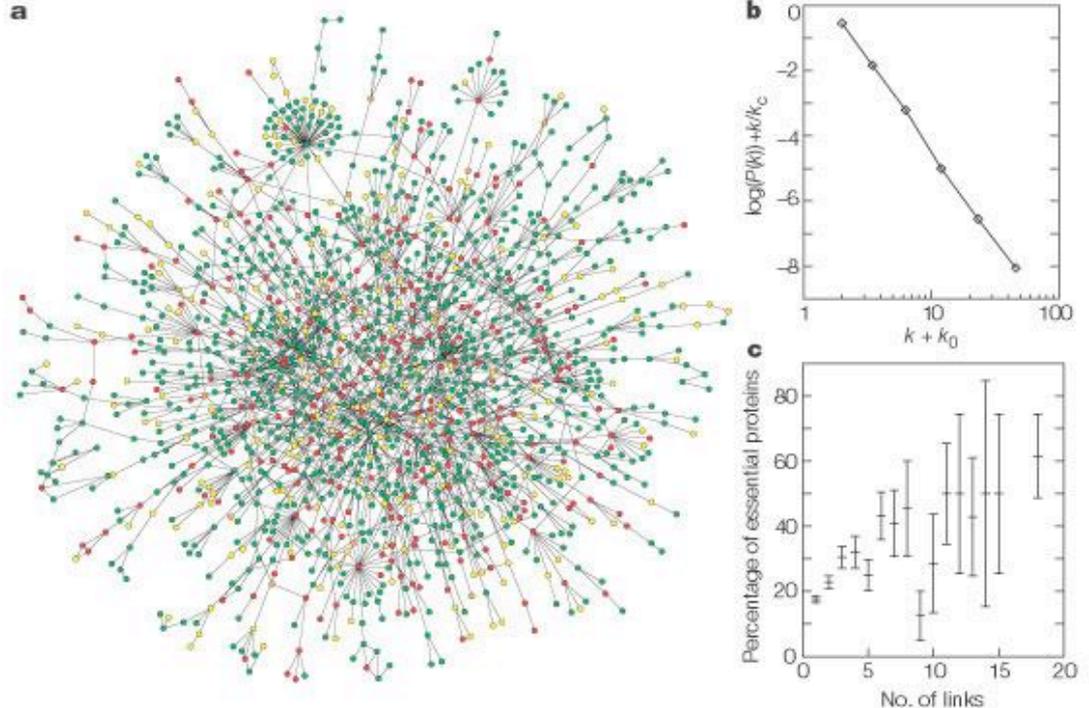
Nature Reviews | Genetics

Barabasi & Oltvai, Nature Rev Gen 5, 101 (2004)

# Welche Proteindeletionen sind tödlich?

(b) Die Häufigkeit von Hub-Proteinen mit  $k$  Interaktionen nicht exponentiell ab (wie in einem Zufallsnetzwerk mit Poisson-Verteilung der Verknüpfungen erwartet), sondern gemäss einem Potenzgesetz (*engl. power law*).

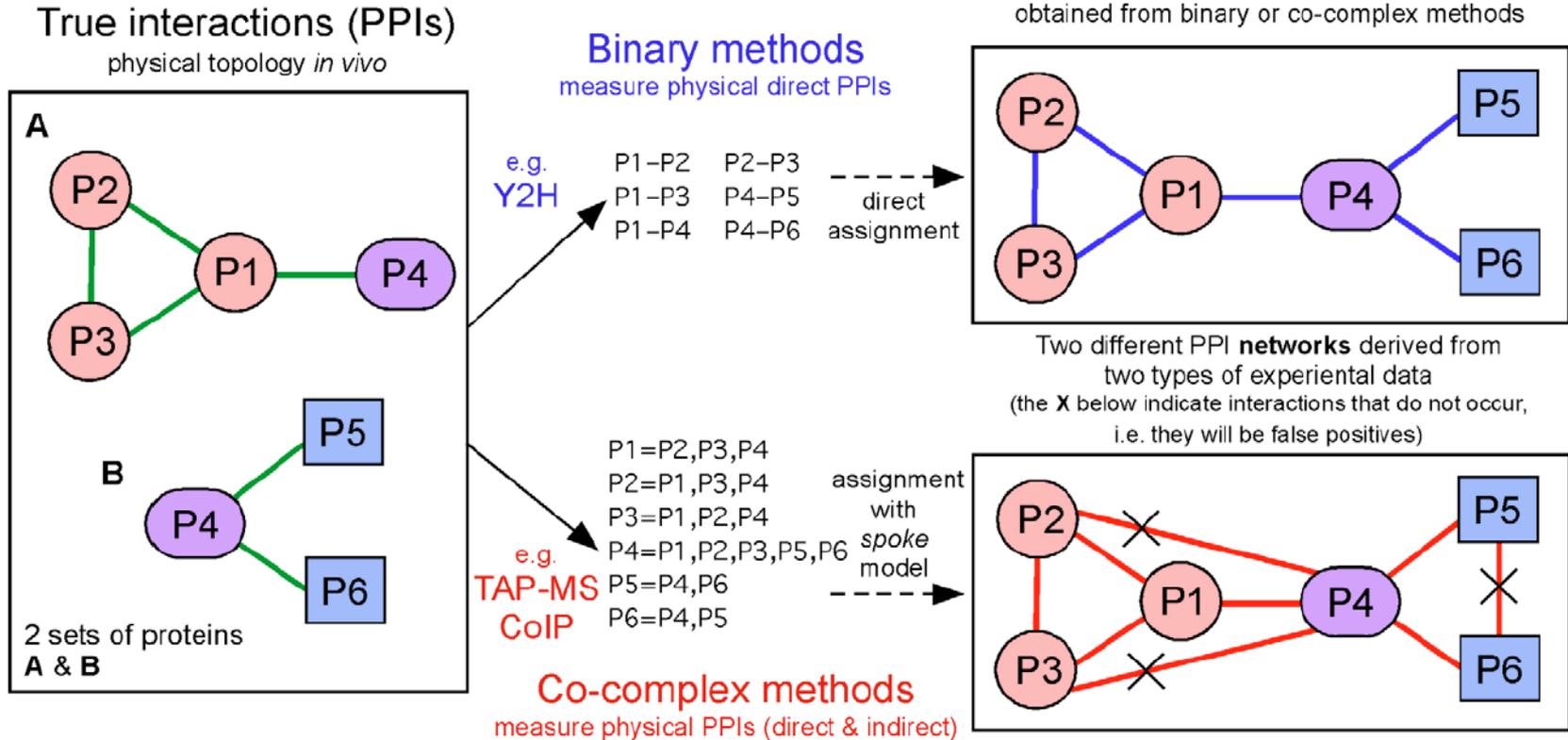
(c) Die Deletion von Genen, die für Hub-Proteine kodieren, die viele Links besitzen, ist eher tödlich als die Deletion von Proteinen, die mit wenigen anderen Proteinen interagieren.



H. Jeong, S. P. Mason, A.-L. Barabási and Z. N. Oltvai  
*Nature* **411**, 41-42

# Proteininteraktionsnetzwerke

Verschiedene experimentelle Techniken messen verschiedene Eigenschaften von Proteinkomplexen.

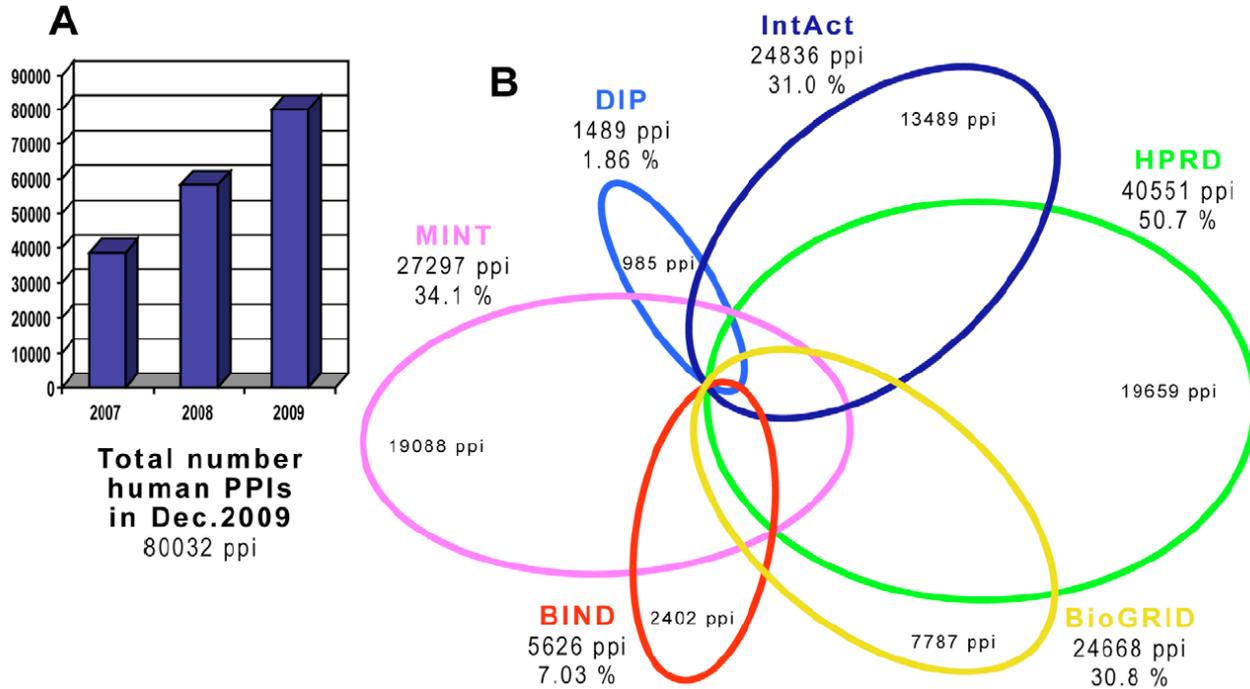


De Las Rivas, PLOS Comp Biol. 6, e1000807 (2010)

# Proteininteraktionsnetzwerke

## Human Interactome

Coverage of human PPIs on major public repositories



Der Überlapp zwischen verschiedenen Datenbanken ist leider gering.

Experimentell nachgewiesen wurden ca. 80 000 Interaktionen von menschlichen Proteinen.

Man schätzt, dass es in *S. cerevisiae* (6000 Gene) ca. 20 000 – 30 000 Proteininteraktionen gibt.

D.h. im Mensch wird eine erheblich höhere Anzahl (ca. 300.000) erwartet.

De Las Rivas, PLOS Comp Biol. 6, e1000807 (2010)

**V11B**  
**SBML und Modell-Erstellung**

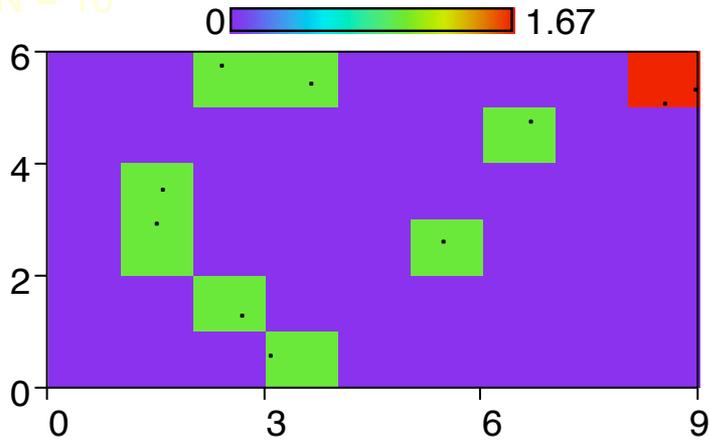
22. Januar 2015

**V12**  
**Stochastische Effekte**  
**+ Diffusion**

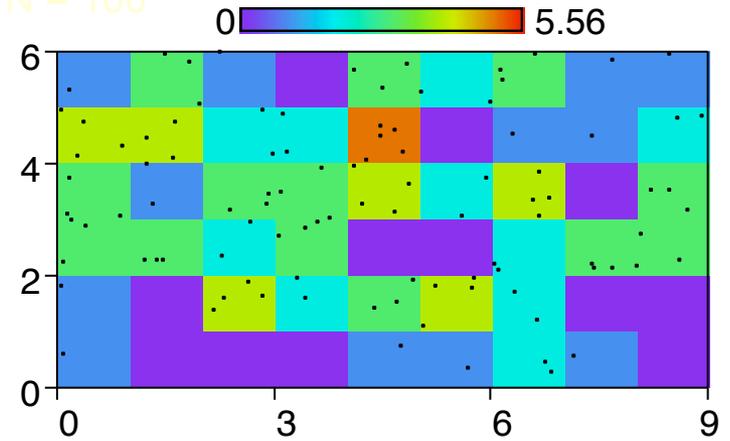
29. Januar 2015

# Dichtefluktuationen

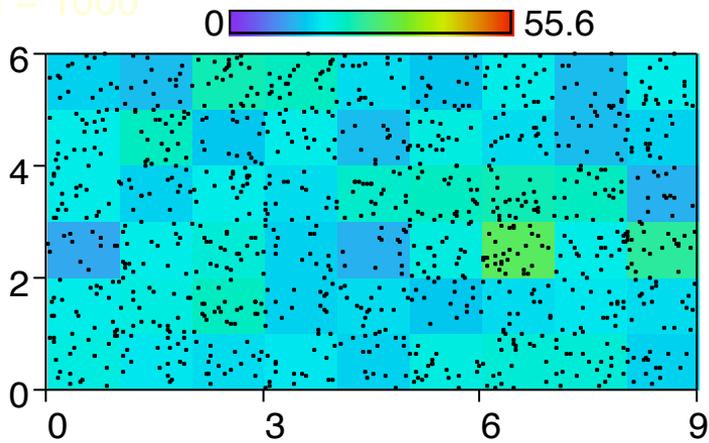
N = 10



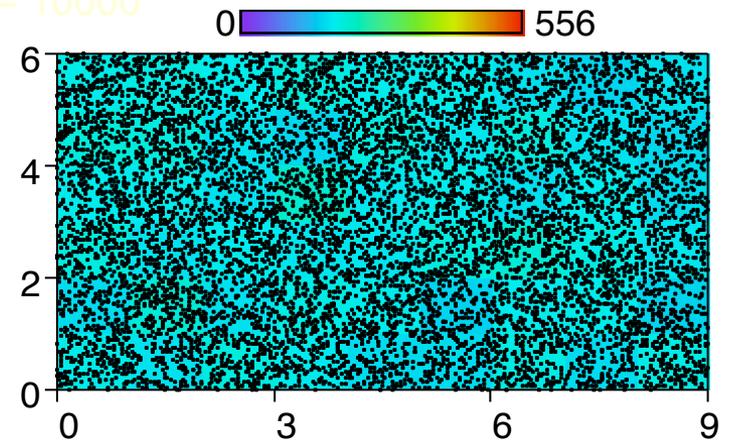
N = 100



N = 1000



N = 10000



# Poisson-Verteilung

Betrachte Kontinuum  $w$  mit im Mittel  $\lambda$  Ereignissen pro Einheitsintervall  $\Delta w$

Annahmen:

- i) Seltenheit:  $\ll 1$  Ereignisse in  $[w, w+\Delta w]$ , höchstens ein Ereignis
- ii) Proportionalität:  $\langle N \rangle = \lambda \Delta w$
- iii) Geschichtslosigkeit (Markov-Prozess)

Wahrscheinlichkeit, dass  $k$  Ereignisse pro Einheitszeitintervall auftreten:  $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$

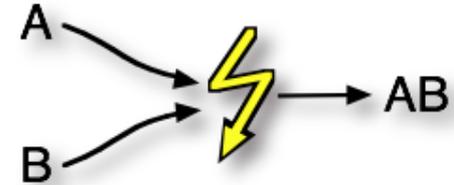
Mittelwert:  $\langle k \rangle = \sum k p_k = \lambda$       Varianz:  $\sigma^2 = \sum p_k (k - \langle k \rangle)^2 = \lambda$

Relative Streuung (Fehler):  $\frac{\Delta k}{k} = \frac{\sigma}{\langle k \rangle} = \frac{1}{\sqrt{\lambda}}$

Mittlere Teilchenzahl	100	1000	1 Mol
relative Unsicherheit	10%	3%	1e-12

# Reaktionen im Teilchenbild

Assoziation:



Kontinuierliche Ratengleichung:  $\frac{d[AB]}{dt} = k[A][B]$

Anzahl neuer AB in  $V$  während  $\Delta t$ :

$$\begin{aligned}\Delta N_{AB} &= \frac{d[AB]}{dt} V \Delta t \\ &= k_{AB} \frac{N_A}{V} \frac{N_B}{V} V \Delta t \\ &= \frac{k_{AB} \Delta t}{V} N_A N_B \\ &= P_{AB} N_A N_B\end{aligned}$$

Reaktionsrate  $k_{AB} \Rightarrow$  Reaktionswahrscheinlichkeit  $P_{AB}$

# Direkte Implementierung



```
Continuous_AB.py
# continuous association of A and B

# parameter
tEnd = 5.0
dt = 0.01
volume = 100.0

# rate and probability
kAB = 1.0
prob = kAB * dt / volume

# initial conditions: particle numbers
A = 1000
B = 1000
AB = 0

# convert to densities
A = A/volume
B = B/volume
AB = AB/volume

# main loop
t = 0.0
print t, "\t", A, "\t", B, "\t", AB

while(t<tEnd):
    dAB = dt * kAB * A * B

    AB += dAB
    A -= dAB
    B -= dAB

    # increment time and output
    t += dt
    print t, "\t", A, "\t", B, "\t", AB
```

```
Stochastic_AB.py
# Stochastic association of A + B => AB

import random

# parameter
tEnd = 5.0
dt = 0.01
volume = 100.0

# rate and probability
kAB = 1.0
prob = kAB * dt / volume

# initial conditions
A = 1000
B = 1000
AB = 0

# main loop
t = 0.0
print t, "\t", A/volume, "\t", B/volume, "\t", AB/volume

while(t<tEnd):
    dAB = 0
    # check for every pair A, B
    for ia in xrange(A):
        for ib in xrange(B):
            r = random.random()
            if (r < prob):
                dAB+=1

    AB += dAB
    A -= dAB
    B -= dAB

    # increment time and output
    t += dt
    print t, "\t", A/volume, "\t", B/volume, "\t", AB/volume
```

# Beispiel: Reaktions-Kette



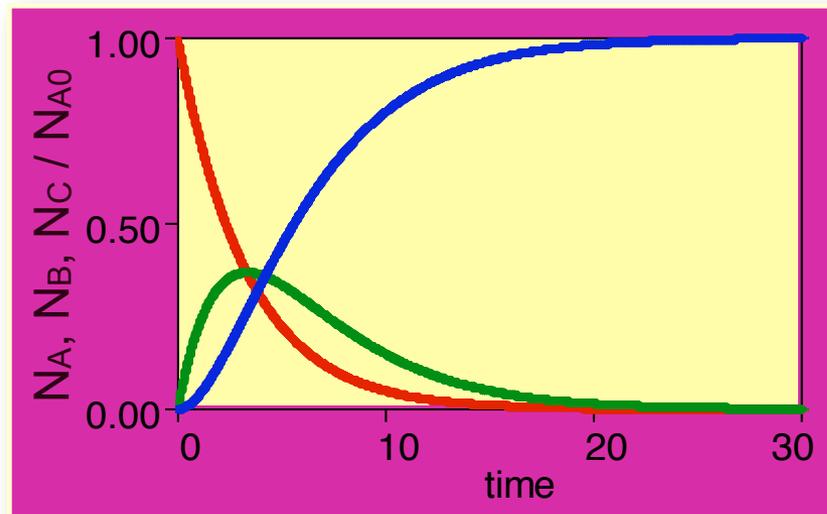
Raten:

$$\frac{dA}{dt} = -k_1 A$$

$$\frac{dB}{dt} = k_1 A - k_2 B$$

$$\frac{dC}{dt} = k_2 B$$

Zeitentwicklung aus den kontinuierlichen Raten:



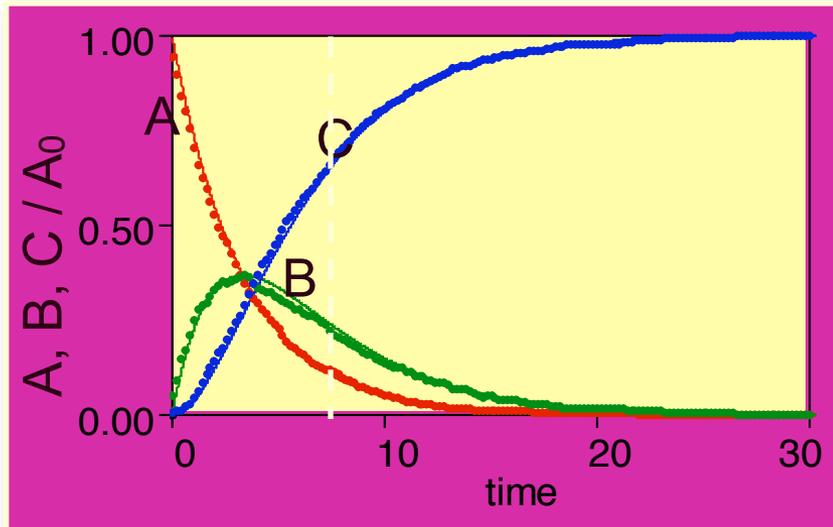
$$k_1 = k_2 = 0.3$$

# Stochastische Simulation

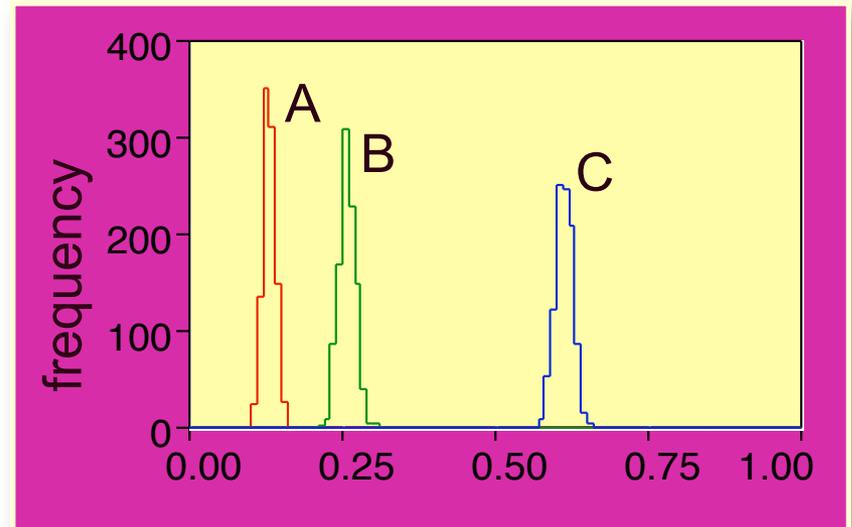


$A_0 = 1000$  Teilchen bei  $t = 0$

$t = 7$



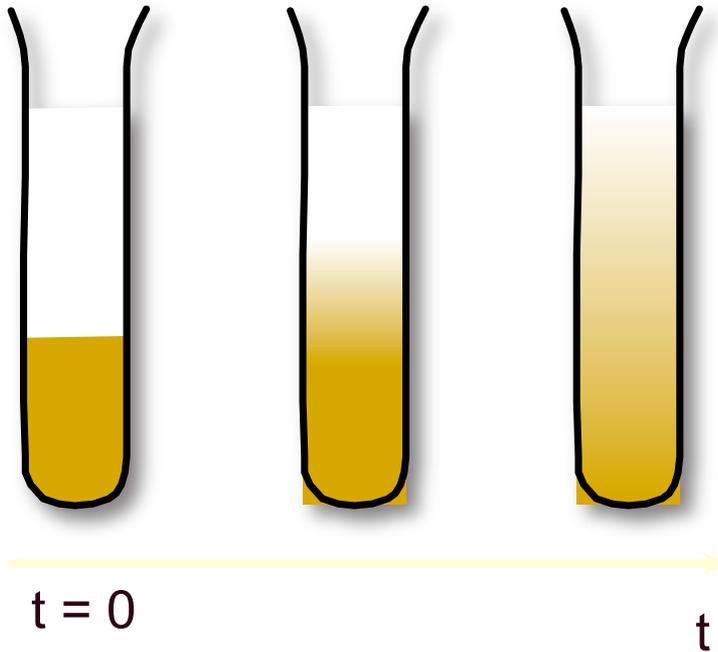
$k_1 = k_2 = 0.3$



Werte bei  $t = 7$  (1000 Läufe)

=> Fluktuationen

# Diffusion



Diffusion  
=> verschmiert Unterschiede

Entwicklung der ortsabh. Dichte  
<=> Diffusionsgleichung

$$\rho(\vec{r}, t) = \frac{\Delta N(\vec{r}, t)}{\Delta V}$$

+ ortsabhängige Quellen und Senken

# Kontinuitätsgleichung

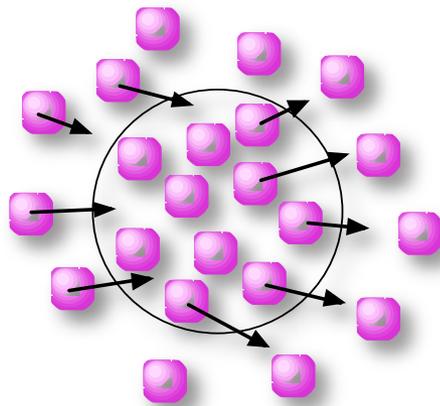
Zwei Beiträge zur Diffusionsgleichung:

1) Kontinuitätsgleichung: wo bleibt das Material?

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla \cdot \vec{j}(\vec{r}, t) = -\text{div } \vec{j}(\vec{r}, t)$$

Änderung der  
Dichte  $\rho$  bei  $(r, t)$

Divergenz  
des Stromes =  
Quellen und  
Senken für Teilchen



partielle Ableitung:  
=> betrachte nur Änderungen von  $\rho$  in der Zeit an  
einem festgehaltenen Ort  $r$  (nicht:  
Ortsverschiebungen  $r = r(t)$ )

$$\Delta N = N_{\text{in}} - N_{\text{out}} = 3 - 5 = -2$$

# Diffusionsstrom

2) Diffusionsstrom durch Dichteunterschiede (Gradienten) – Fick'sches Gesetz:

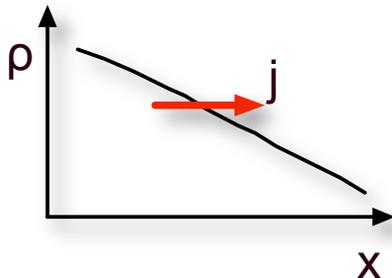
$$\vec{j}(\vec{r}, t) = -D \nabla \rho(\vec{r}, t) = -D \text{ grad } \rho(\vec{r}, t)$$

Diffusionsstrom  
bei  $(r, t)$

Strom fließt weg  
von hohen  
Dichten

Diffusions-  
koeffizient

Dichte-  
fluktuationen  
(=Gradienten)

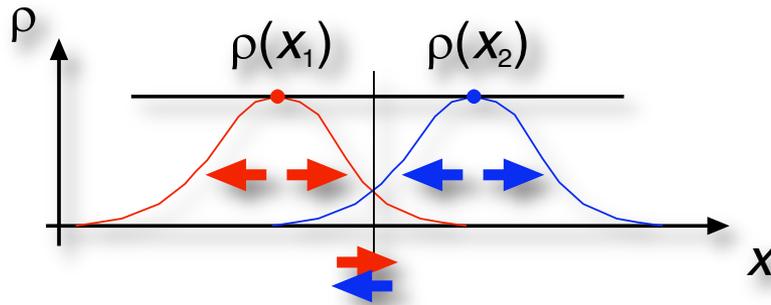


hier: phänomenologischer  
Umrechnungsfaktor von  
Dichteunterschieden in Teilchenströme

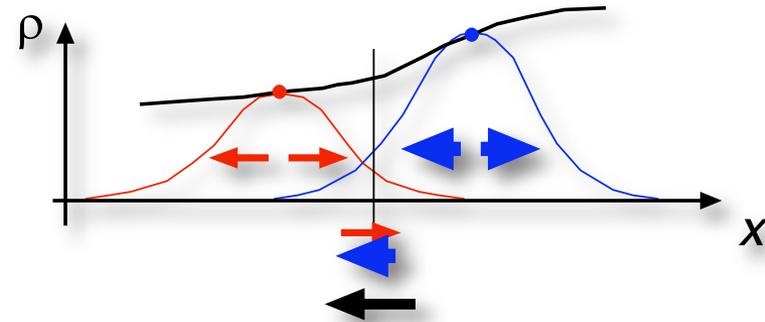
# Diffusion mikroskopisch

Ohne externe Kräfte

=> Teilchen bewegen sich in alle Richtungen gleich wahrscheinlich  
(Gauss'sche Wahrscheinlichkeit)



$$\rho(x_1) = \rho(x_2) \Rightarrow j_{diff} = 0$$



$$\rho(x_1) < \rho(x_2) \Rightarrow j_{diff} < 0$$

$$j_{diff} \propto -\frac{\rho(x_2) - \rho(x_1)}{x_2 - x_1} \Rightarrow \frac{d\rho}{dx}$$

Gleiche Dichten an  $x_1$  und  $x_2$ :  
=> gleiche Anzahl Teilchen springt  
von  $x_1 \Rightarrow x_2$  wie von  $x_2 \Rightarrow x_1$

Unterschiedliche Dichten:  
=> mehr Teilchen springen  
von  $x_2 \Rightarrow x_1$  als von  $x_1 \Rightarrow x_2$

# Diffusionsgleichung: partielle DGL

Diffusionsstrom

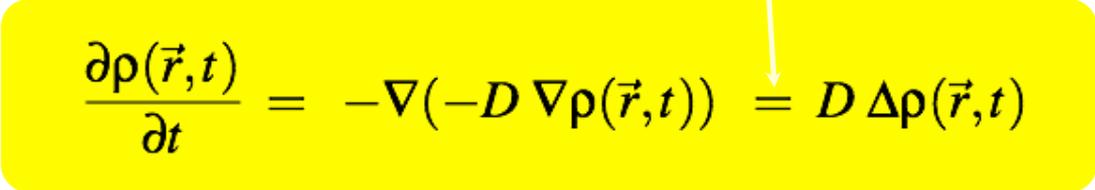
$$\vec{j}(\vec{r}, t) = -D \nabla \rho(\vec{r}, t) = -D \text{grad } \rho(\vec{r}, t)$$

in Kontinuitätsgleichung einsetzen

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla \cdot \vec{j}(\vec{r}, t) = -\text{div } \vec{j}(\vec{r}, t)$$

=> Diffusionsgleichung:

$$D(\vec{r}, t) = \text{const}$$


$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla \cdot (-D \nabla \rho(\vec{r}, t)) = D \Delta \rho(\vec{r}, t)$$

=> Vollständige Beschreibung der zeitabhängigen Dichteverteilung  
(ohne externe Kräfte)

# Integration

Bisher: (System von) ODEs

$$\frac{d}{dt}X_i = f_i(X_1, X_2, \dots)$$

- Zeitentwicklung abhängig von den **lokalen** Werten der Systemparameter
- alle Ableitungen nach der Zeit

Jetzt: Diffusionsgl. mit konstantem D:

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = D \Delta \rho(\vec{r}, t)$$

- Zeitentwicklung bestimmt durch **globale** Werte (Verteilungen) der Variablen (gesamte Dichte  $\rho(r)$  nötig für Gradient)
- Ableitungen nach **Zeit und Ort**

# FTCS-Integrator

Diffusionsgleichung mit konstantem  $D$  in 1D:

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = D \frac{\partial^2 \rho(\vec{x}, t)}{\partial x^2}$$

Direkte Implementierung auf einem Gitter  $\{\rho(x_i)\}$  mit Abstand  $\Delta x$ :

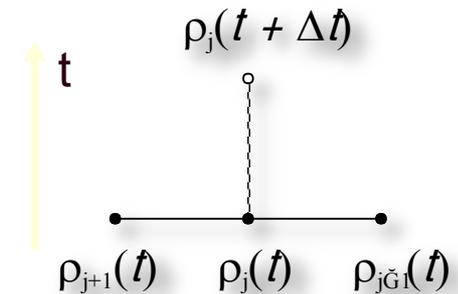
$$\frac{\rho_j(t + \Delta t) - \rho_j(t)}{\Delta t} = D \frac{\rho_{j+1}(t) - 2\rho_j(t) + \rho_{j-1}(t)}{\Delta x^2}$$

**Forward in Time**

**Centered in Space**

Propagationsschritt:

$$\rho_j(t + \Delta t) = \rho_j(t) + \Delta t D \frac{\rho_{j+1}(t) - 2\rho_j(t) + \rho_{j-1}(t)}{\Delta x^2}$$



Stabil für:

$$\Delta t \leq \frac{\Delta x^2}{2D}$$

( $\Delta t < \text{Diffusionszeit über Abstand } \Delta x$ )