



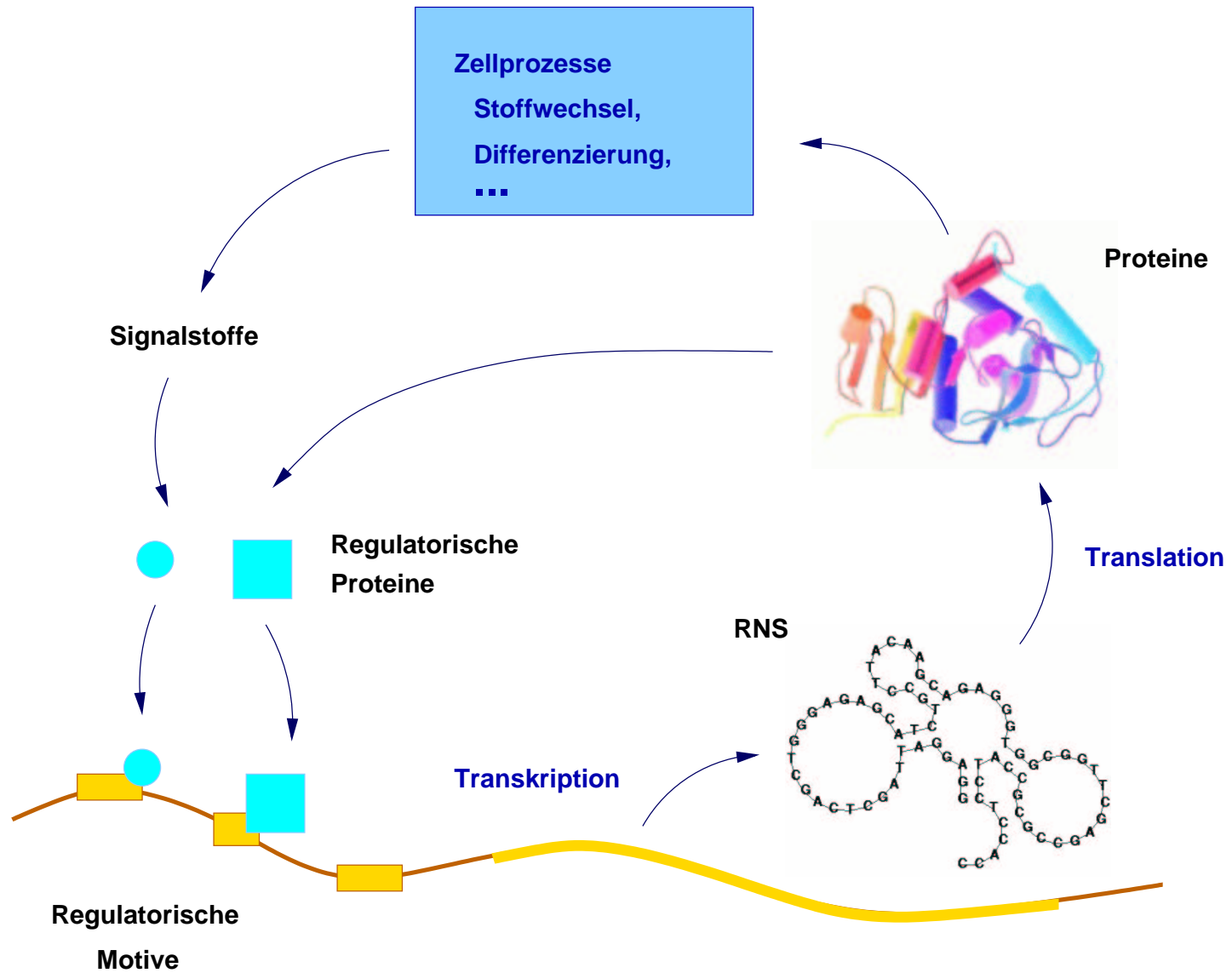
Analyse der optimalen differenziellen Genexpression

Wolfram Liebermeister

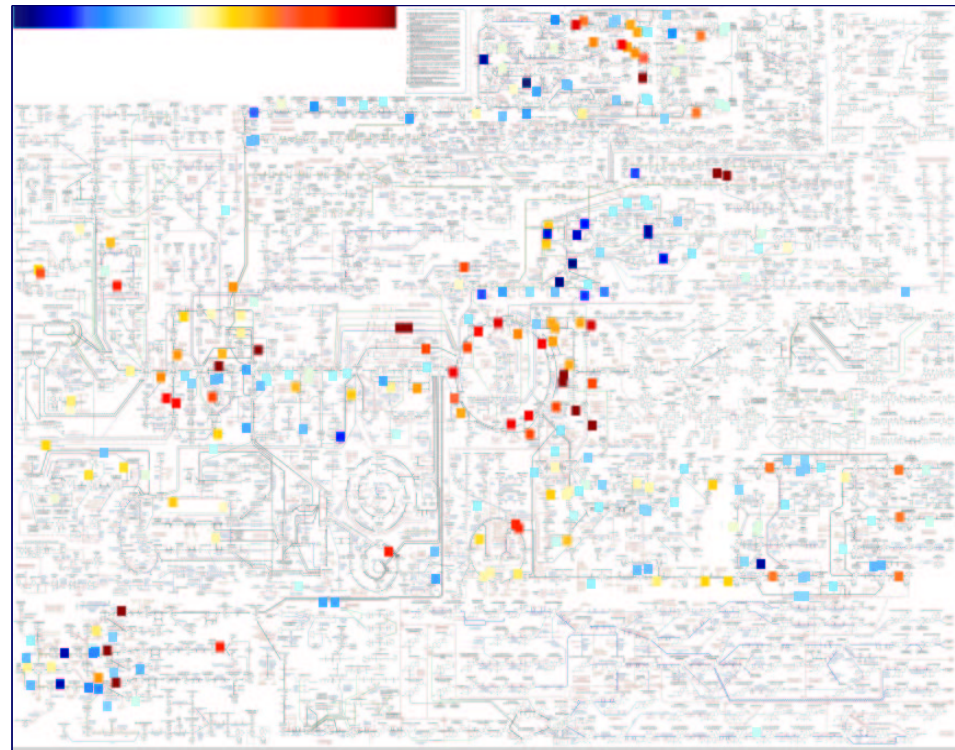
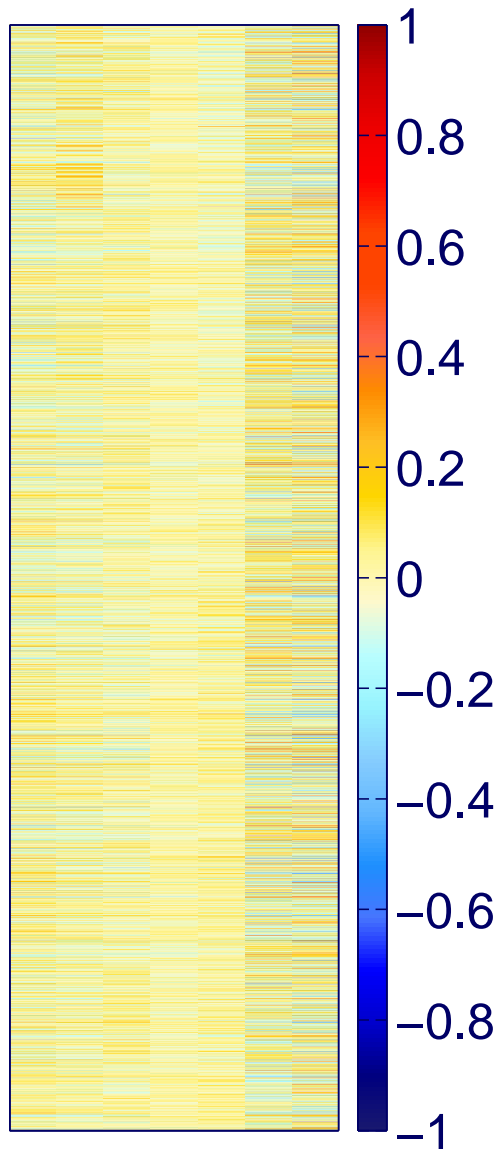
13. Januar 2004

Humboldt-Universität zu Berlin
Theoretische Biophysik

Genexpression



Microarraydaten



log₂ diff. Expression **-Glukose/+Glukose**

Daten aus

J.L. DeRisi et al. (1997), *Exploring the metabolic and genetic control of gene expression on a genomic scale*, Science 278

Zielstellung meiner Arbeit

- **Datenanalyse mit statistischen Modellen**

Globale Moden in Microarraydaten als additive Faktoren

Koexpression z.B. bei Proteinkomplexen und im Stoffwechsel

- **Problem:**

Theoretische Erklärung für die Beziehung zwischen Expression und Funktion?

- **Vorhersage von Expressionsprofilen**

Expression dient der Steuerung der Proteinmenge

Komplexe Steuerung: viele Gene, viele Zellprozesse

Welche Korrelationen sind in Expressionsdaten zu erwarten?

Ziel: Allgemeine Beziehungen für große Zellmodelle

Ein Modell für optimale Genexpressionsmuster

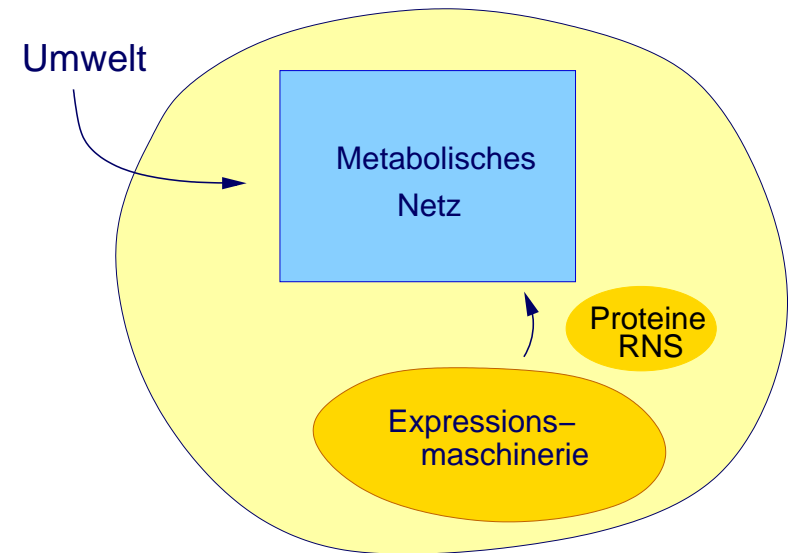
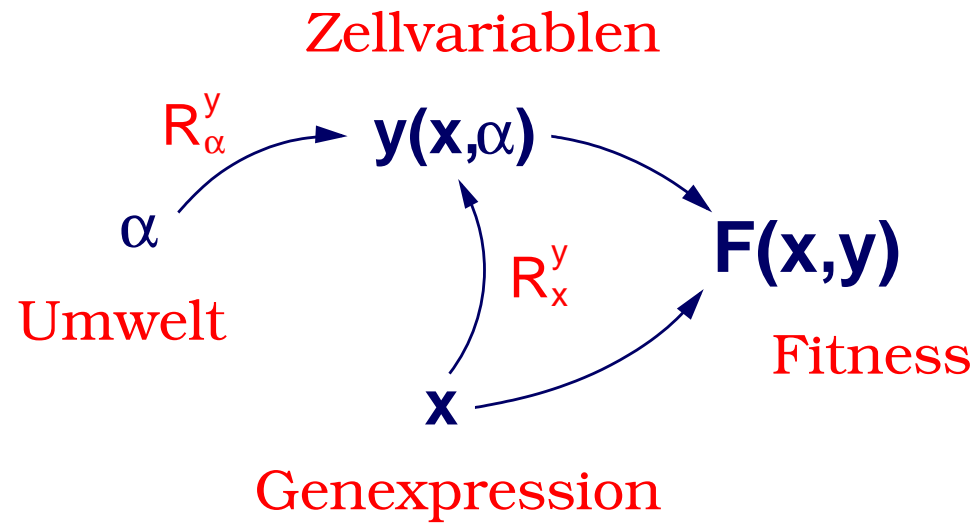
Das Modell soll

1. optimale differentielle Expression vorhersagen
2. die koordinierte Expression vieler Gene beschreiben
3. die korrelierte Expression im Stoffwechsel erklären

Modellannahmen

1. Differentielle Expression dient der optimalen Kontrolle von Zellprozessen
2. Modell der Zellprozesse mit stabilem stationären Zustand
3. Optimalität bedeutet: Maximierung einer Zielfunktion ("Zellfitness")
4. Gene mit komplexen Effekten:
Kompromiss zwischen Nutzen und Kosten für alle Gene
5. Differentielle Störungen \Rightarrow lokale Näherung

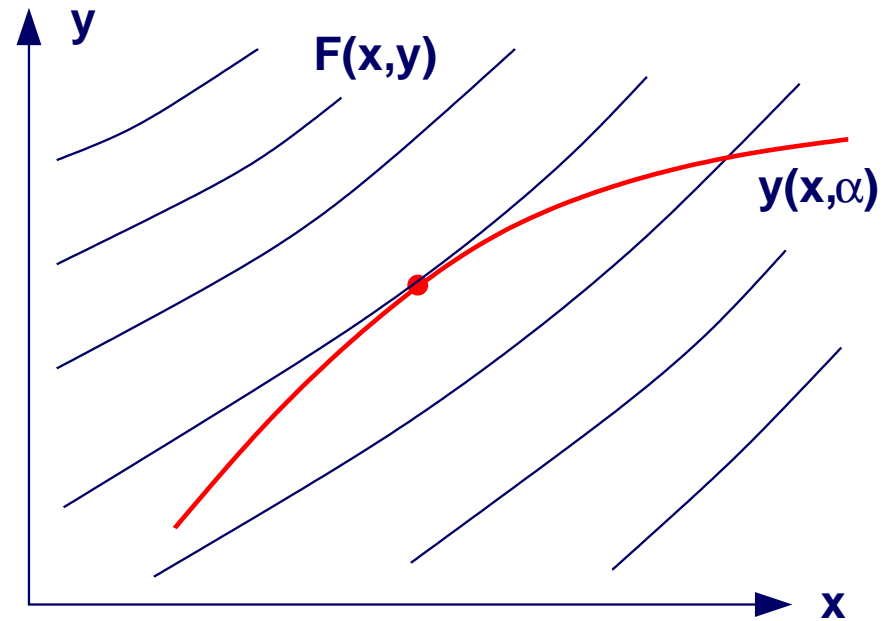
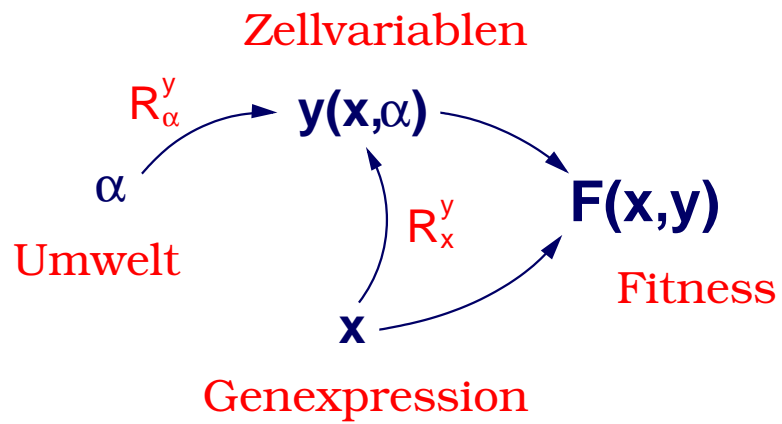
Das Modell



Das Optimalitätsprinzip:

Die Expressionswerte x_i maximieren die Fitness $F(x, y(x, \alpha))$

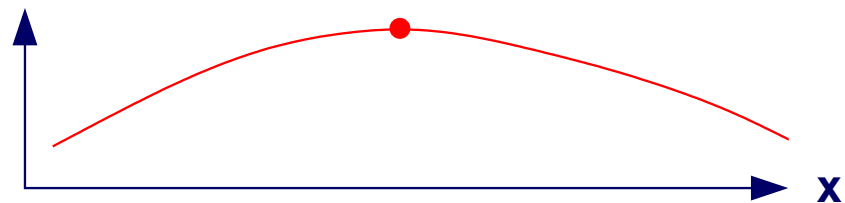
Das Modell



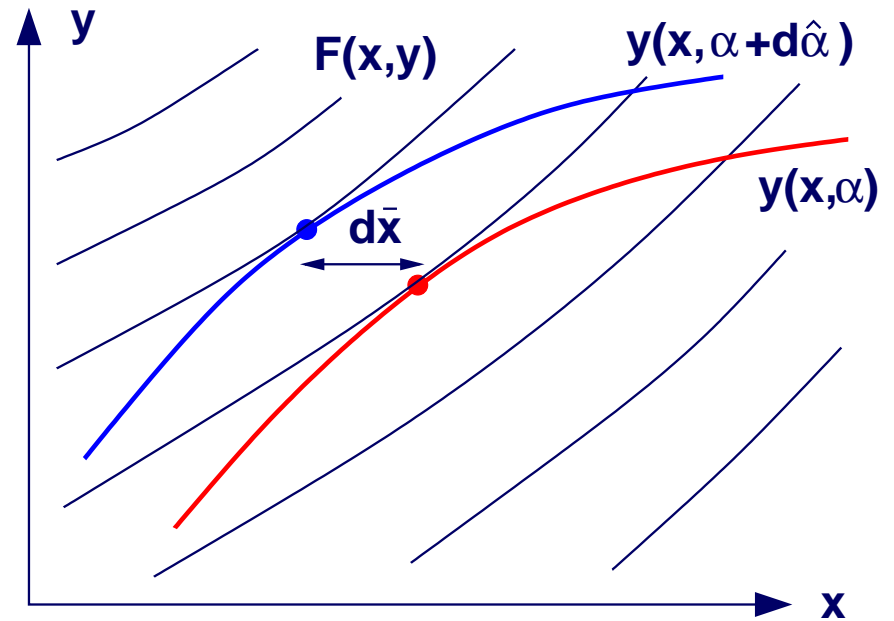
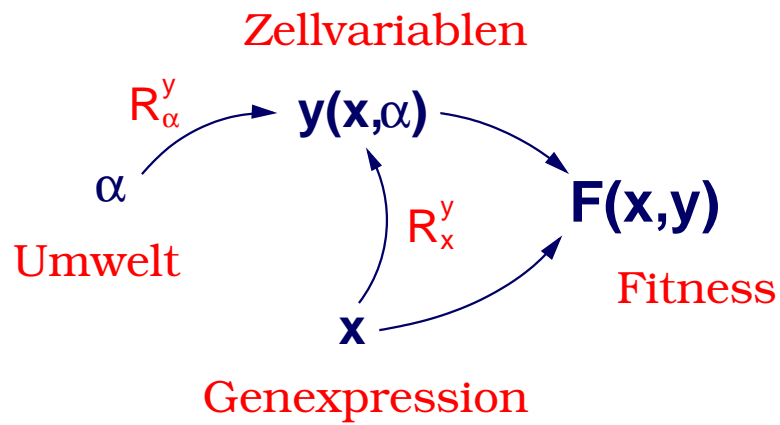
Effektive Fitness

$$G(x, \alpha) = F(x, y(x, \alpha))$$

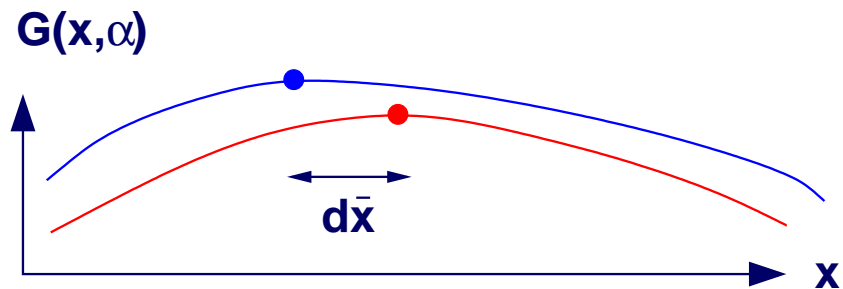
$G(x, \alpha)$



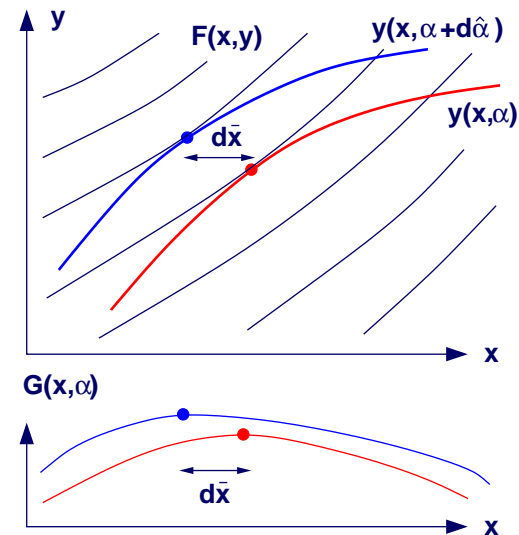
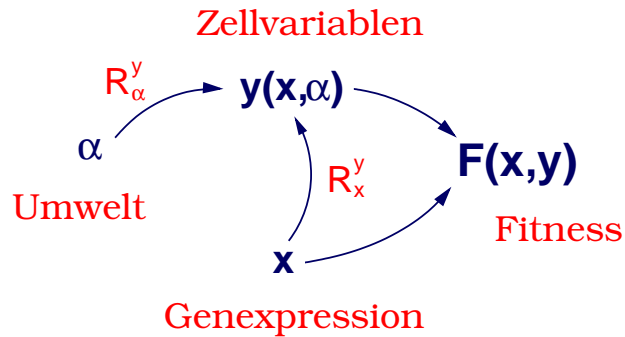
Das Modell



Differentielle Störung $d\hat{\alpha} \rightarrow$
Differentielle Expression $d\bar{x}$



Optimale differentielle Expression



Differentielle Störung $d\hat{\alpha} \rightarrow d\hat{y}, d\hat{R}_x^y$

Responsekoeffizienten

$$R_x^y = \frac{\partial y}{\partial x}, \quad R_\alpha^y = \frac{\partial y}{\partial \alpha}$$

Ableitungen der Fitness

$$F_x, F_{xx}, F_{yy}, G_x, G_{xx}, G_{x\alpha}$$

Optimalitätsbedingung

$$d\bar{G}_x = 0$$

Differentielles Expressionsprofil

$$d\bar{x} = -G_{xx}^{-1} G_{x\alpha} d\hat{\alpha}$$

Vereinfachende Annahme

$$F(x, y) = V(y) + U_1(x_1) + U_2(x_2) + \dots$$

Expression als Reaktion auf $d\hat{y}$ und $d\hat{R}_x^y$

$$d\bar{x} = -(G_{xx}^{-1} R_x^{yT} F_{yy}) d\hat{y} - (G_{xx}^{-1} F_y^T) d\hat{R}_x^y$$

Drei Anwendungen

1. **Optimale Feedbacksteuerung**

Wie müssen Gene auf Zellvariablen reagieren, damit optimale Expressionsmuster entstehen?

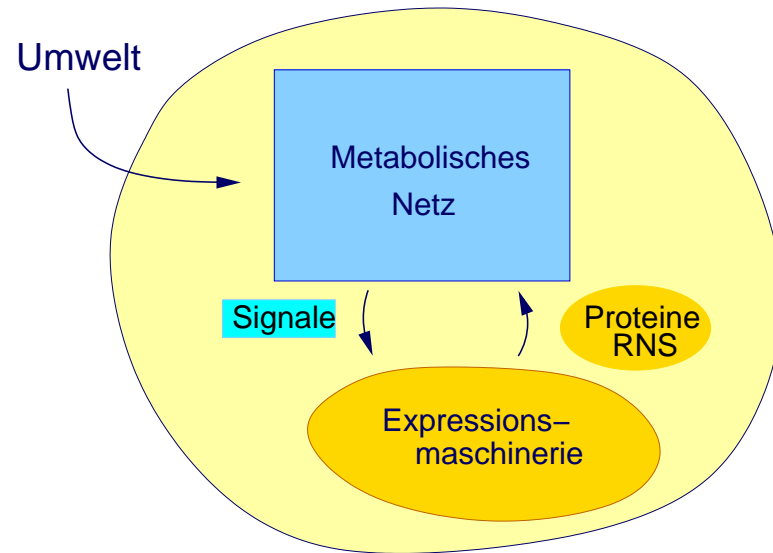
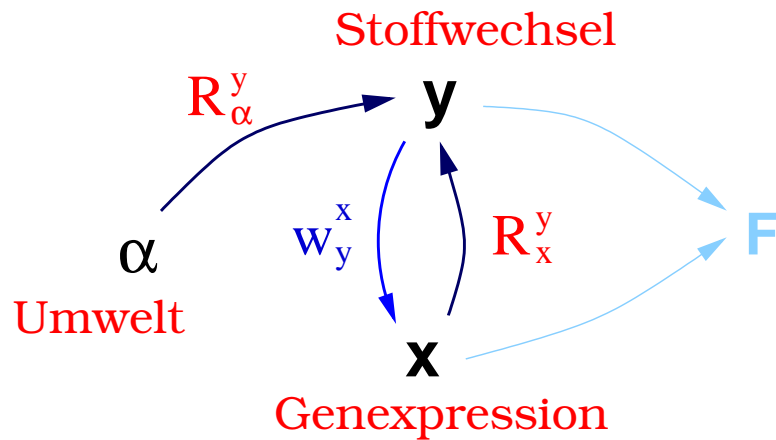
2. **Expressionsmuster und Responsekoeffizienten**

Wie läßt sich die korrelierte Expression im Stoffwechsel erklären?

3. **Kompensation von Gendeletionen**

Wie sollen die Expressionsmuster nach der Deletion einzelner Gene aussehen?

(1) Optimale Feedbacksteuerung



Lineare Beziehung

$$dx = w_y^x dy$$

Optimalitätsbedingung

$$d\bar{G}_x = 0$$

Annahme

$$d\hat{R}_x^y = 0$$

Optimale Einflüsse

$$w_y^x = -F_{xx}^{-1} R_x^{yT} F_{yy}$$

(2) Expression im Stoffwechsel

- **Stoffwechsel: Topologie vs. Funktion**

Topologische Beschreibung durch Stoffwechselwege?

Responsekoeffizienten \Rightarrow globale Wirkung eines Enzyms

Responsekoeffizienten spiegeln die Netzstruktur wider

- **Optimalität führt zu Koregulation**

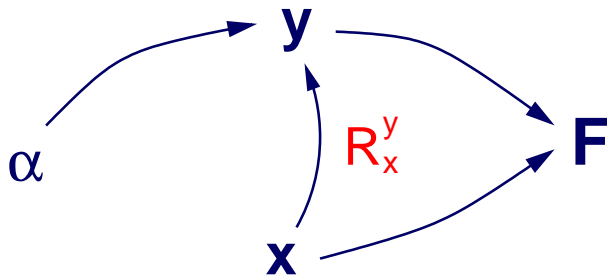
Teilprobleme:

1. Genexpression \Leftarrow Responsekoeffizienten

2. Responsekoeffizienten \Leftrightarrow Stoffwechselnetz

\Rightarrow **Summenregeln** für korrelierte Expression von benachbarten Reaktionen und auf stationären Flussmoden

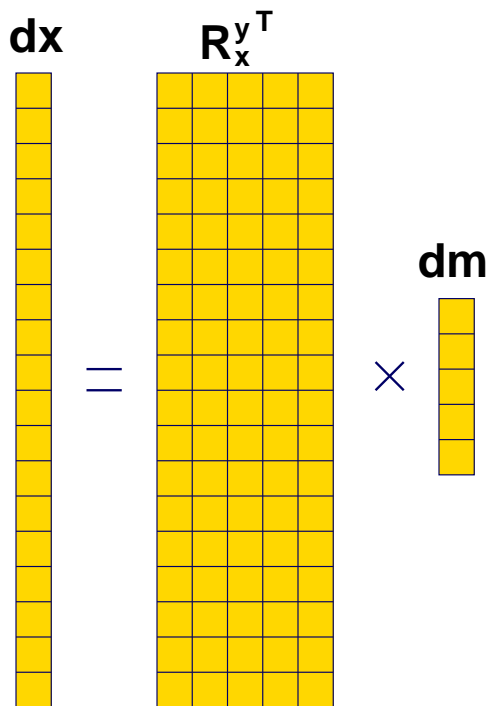
(2) Lineare Überlagerung von Responsekoeffizienten



Optimales Expressionsprofil

für gegebene Änderung dy :

$$d\bar{x} = F_{xx}^{-1} R_x^{yT} (R_x^y F_{xx}^{-1} R_x^{yT})^{-1} dy$$



Annahme

Fitnesskrümmungen aller Gene gleich

Optimales Expressionsprofil

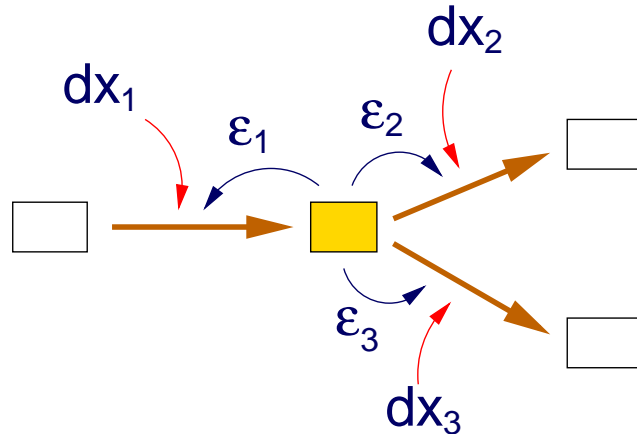
$$\begin{aligned} d\bar{x} &= R_x^{yT} dM \\ &= R_x^{y_1T} dm_1 + R_x^{y_2T} dm_2 + \dots \end{aligned}$$

⇒ Strukturen in Responsekoeffizienten können sich auf Expressionsprofile übertragen

(2) Summenregel 1: Benachbarte Reaktionen

Annahmen:

- Nur Reaktionsgeschwindigkeiten werden reguliert
- Keine Erhaltungsbeziehung zwischen Metaboliten



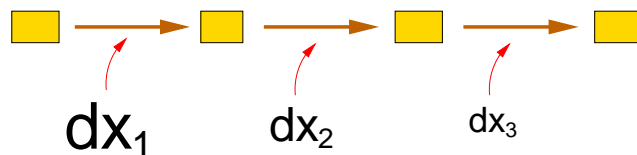
Summenregel

$$d\bar{x}^T \epsilon L = 0$$

aus dem Konnektivitätstheorem

Rings um einen Metaboliten gilt

$$\epsilon_1 d\bar{x}_1 + \epsilon_2 d\bar{x}_2 + \epsilon_3 d\bar{x}_3 + \dots = 0$$



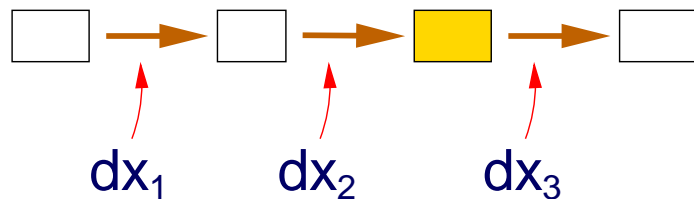
Lineare Reaktionskette:

Differentielle Expressionswerte sind proportional

(2) Summenregel 2: Reaktionen entlang Flußmoden

Annahmen:

- Nur Stoffkonzentrationen werden reguliert



Summenregel

$$d\bar{x}^T K = 0$$

aus dem Summationstheorem

Auf stationärem Flussmodus gilt

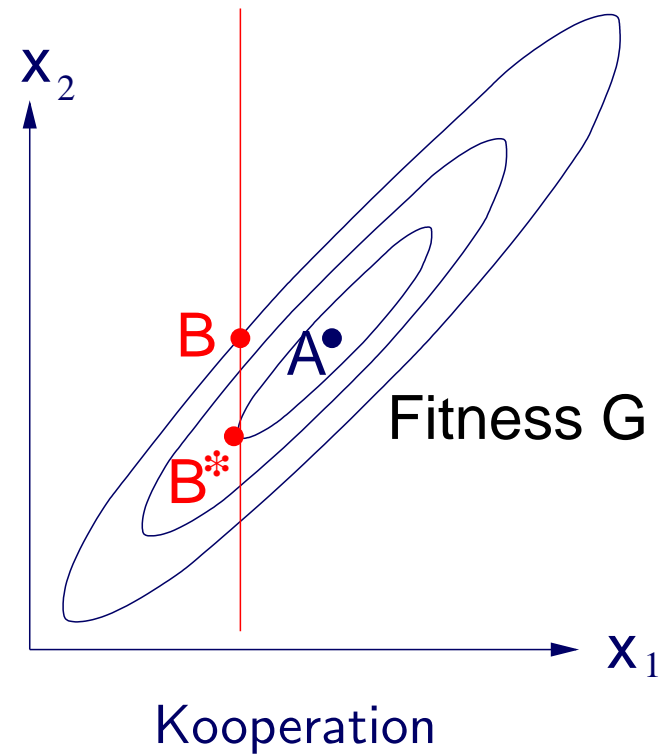
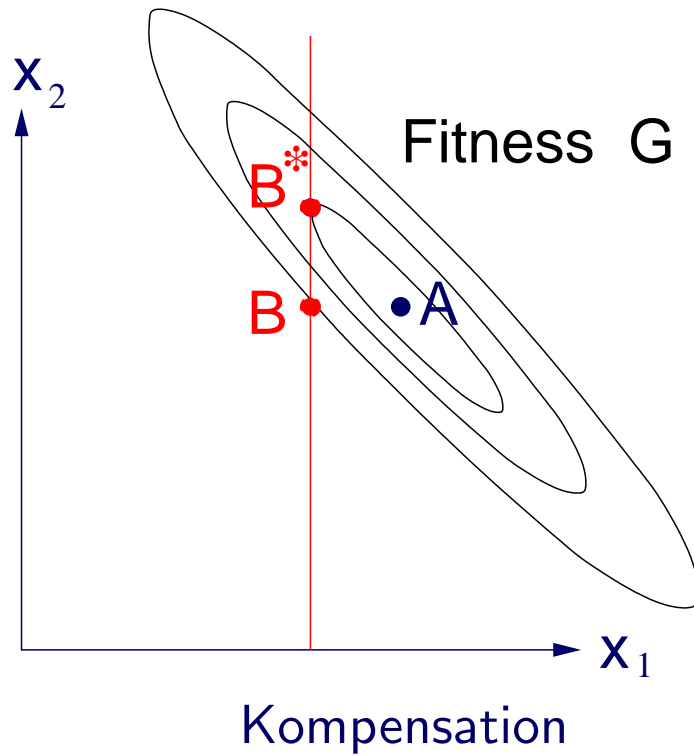
$$d\bar{x}_1 + d\bar{x}_2 + d\bar{x}_3 + \dots = 0$$

Lineare Reaktionskette:

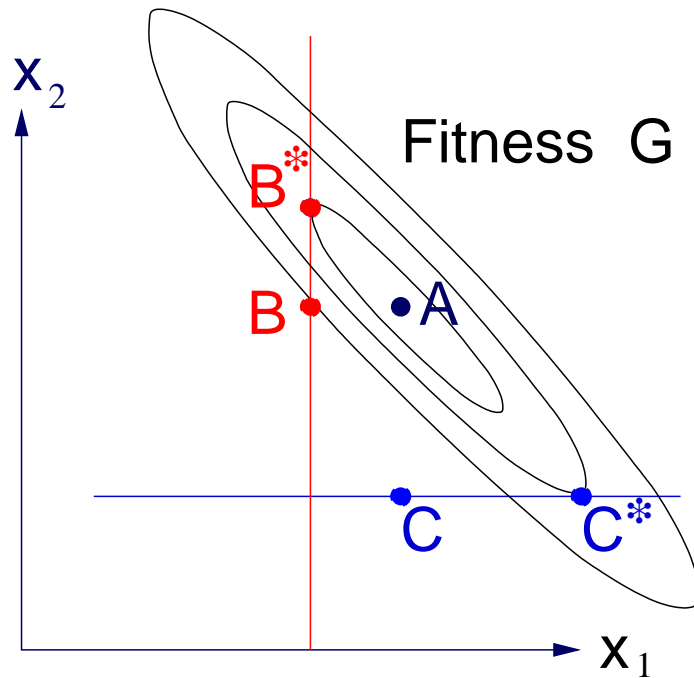
Gegenläufige Steuerung

Zur Erhöhung einer Konzentration wird die Enzymaktivität oberhalb des Metaboliten erhöht, unterhalb gehemmt.

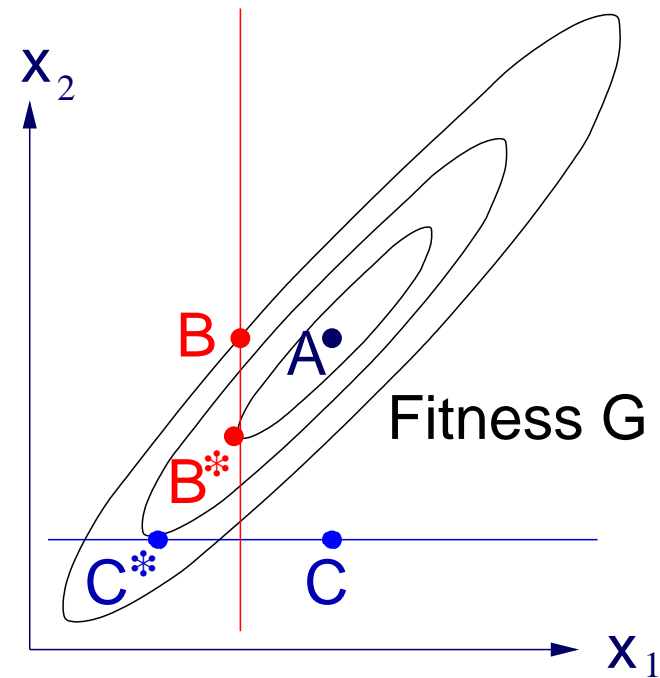
(3) Kompensation von Gendelektionen



(3) Kompensation von Gendeletionen



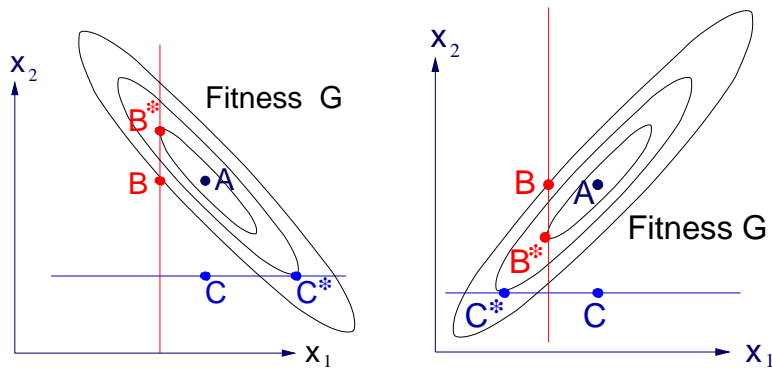
Kompensation



Kooperation

In beiden Fällen: Symmetrisches Verhalten der Gene

(3) Kompensation von Gendeletionen

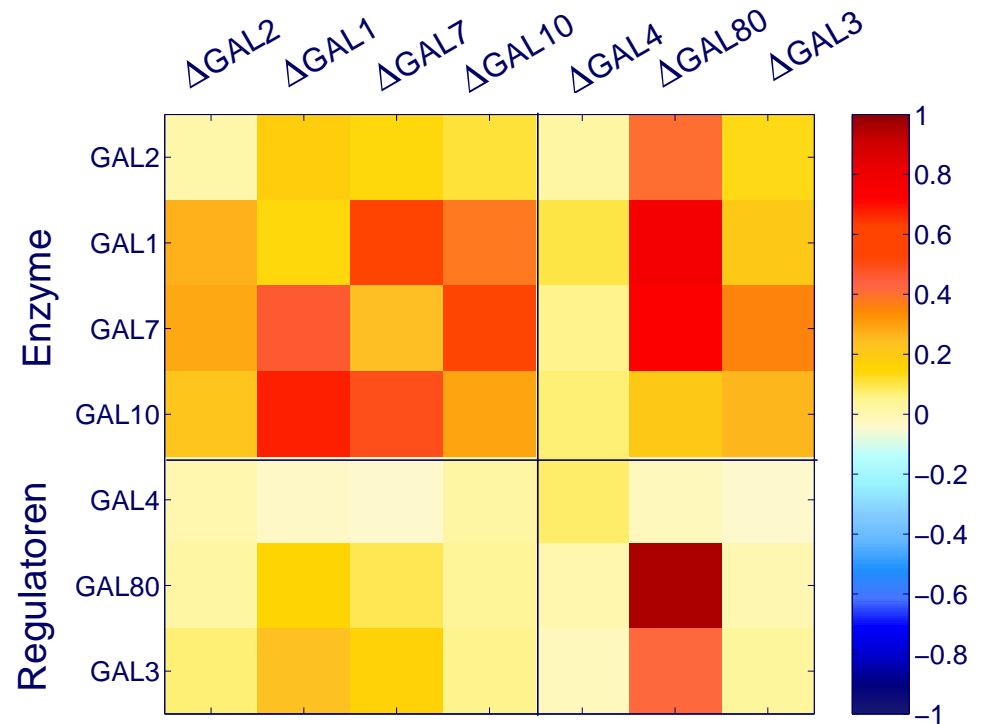


Modellvorhersage

für Expressionsmatrix X nach
Deletionen

$$X = G_{xx}^{-1} D$$

G_{xx}^{-1} ist symmetrisch, D ist diagonal.



Daten: + Galaktose/ - Galaktose

T. Ideker et al. (1997), *Integrated Genomic and proteomic analyses of a systematically perturbed metabolic network*, Science 292

Zusammenfassung

Ansatz

- Optimalitätsannahme \Rightarrow differentielle Expressionsprofile
- Koordination vieler Gene mit komplexen Wirkungen
- lokale Näherung \Rightarrow allgemeine Vorhersagen

Ergebnisse

- Optimale differentielle Expressionsmuster
- Regulation von Genen repräsentiert u.A. deren Funktion
- Korrelierte Expression von Stoffwechsellenzymen
- Symmetrische Antwort auf Deletionen



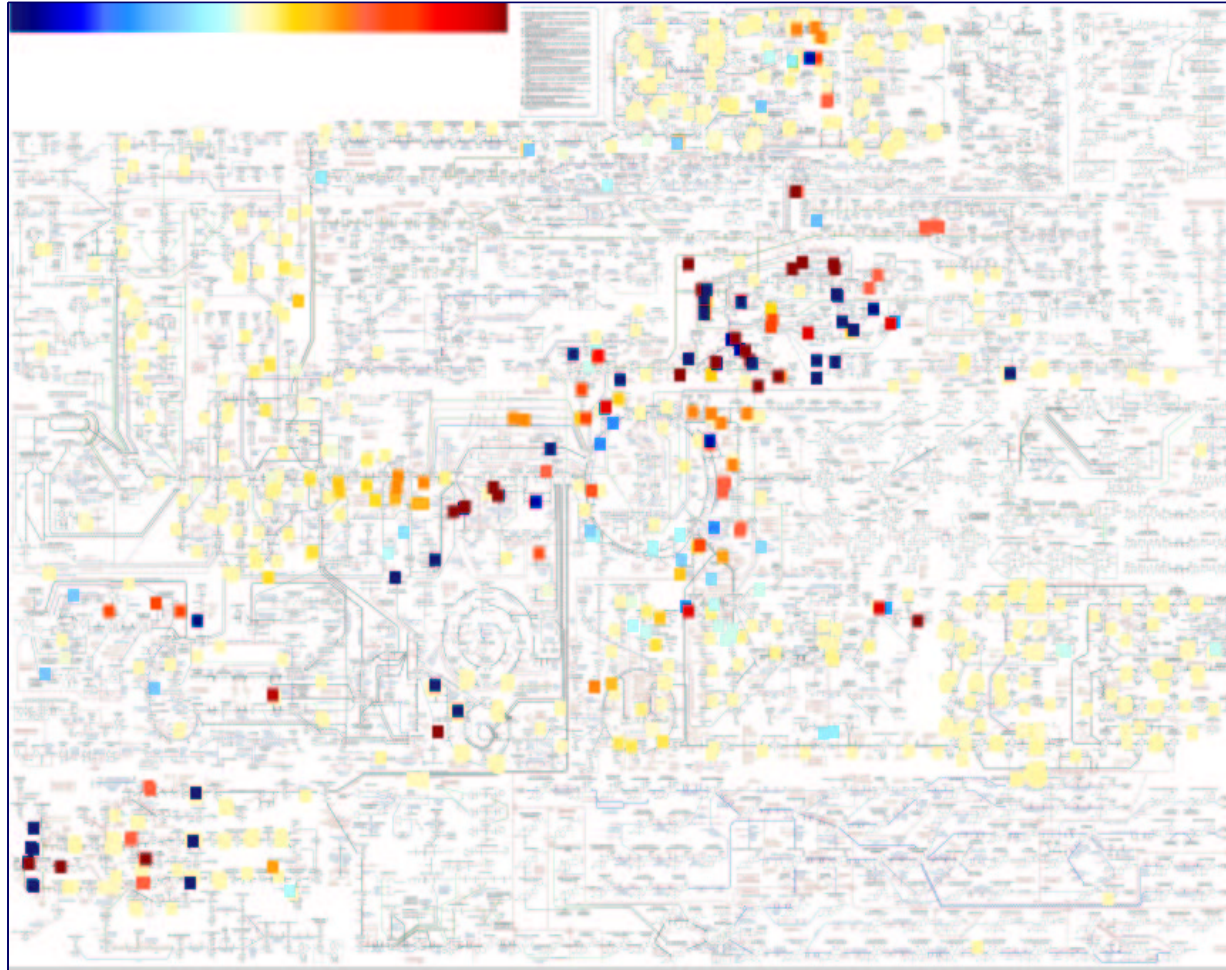
Analyse der optimalen differentiellen Genexpression

Wolfram Liebermeister

13. Januar 2004

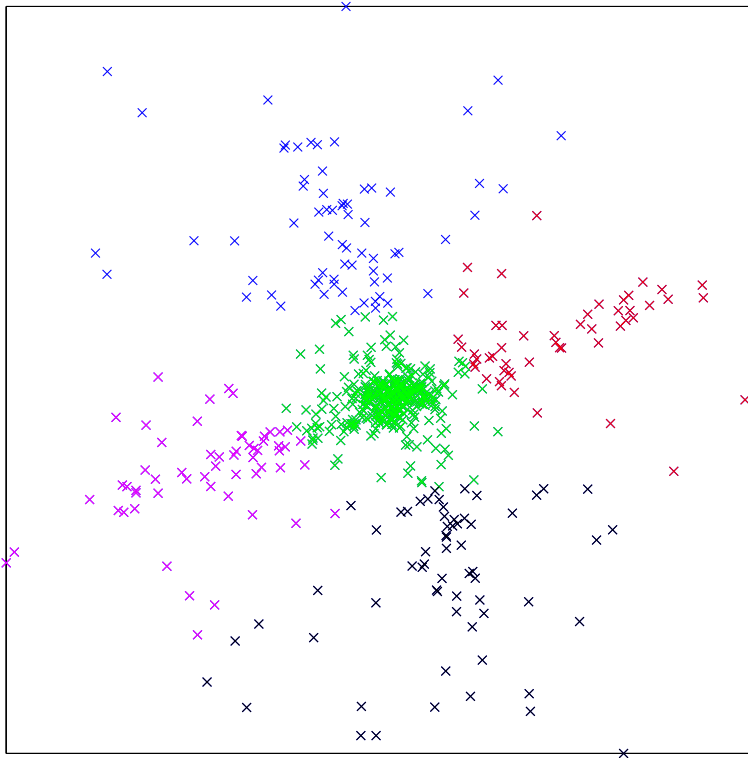
Humboldt-Universität zu Berlin
Theoretische Biophysik

Metabolische Kontrollkoeffizienten

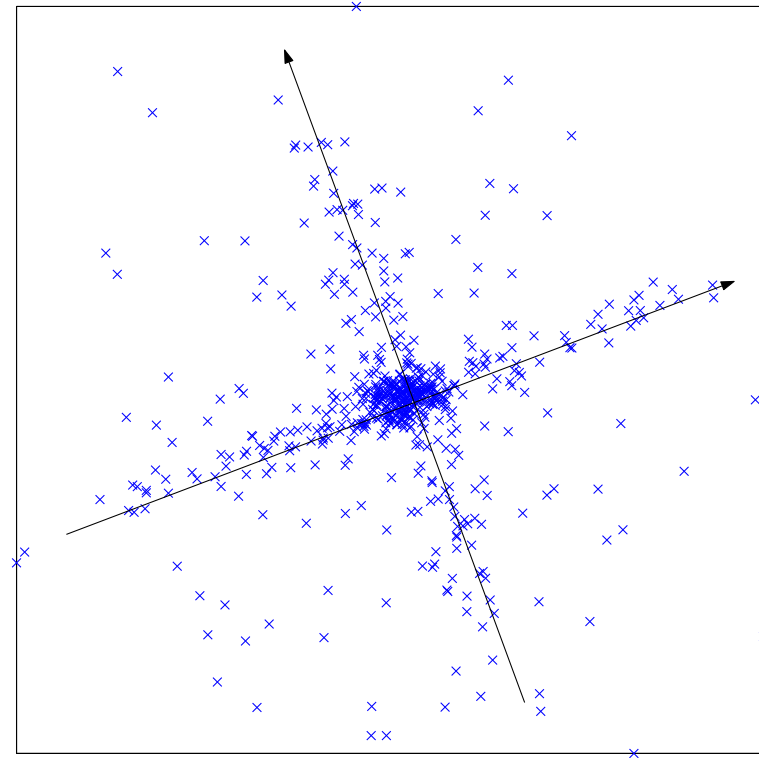


$R_i^{Ethanol}$ für Ethanol, basierend auf Hefe-Netz aus KEGG

Analyse von multivariaten Expressionsdaten

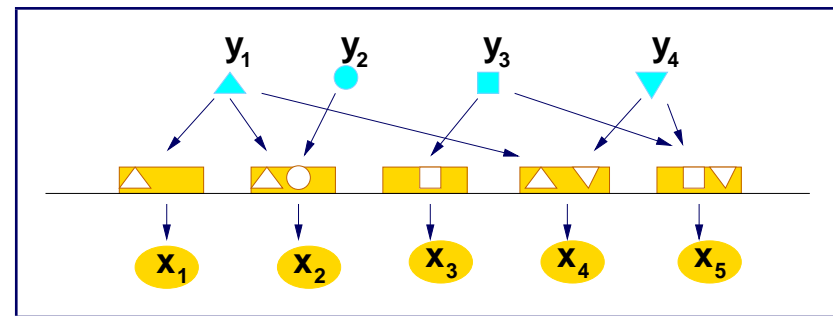
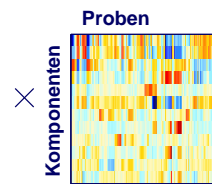
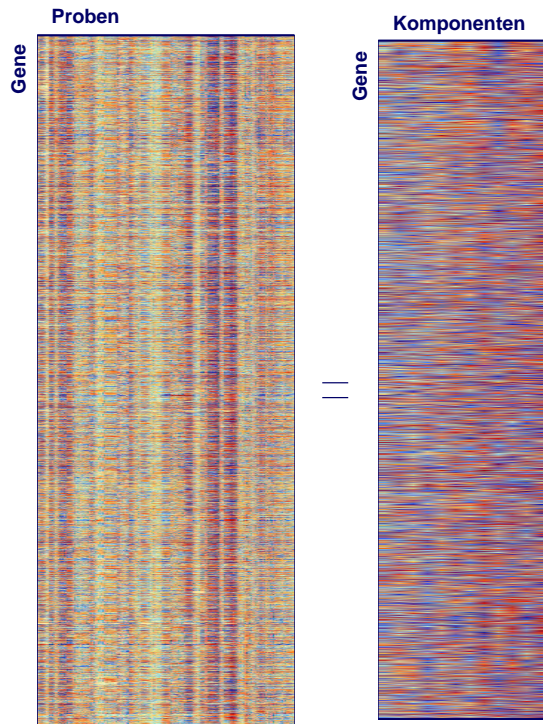


Clustering



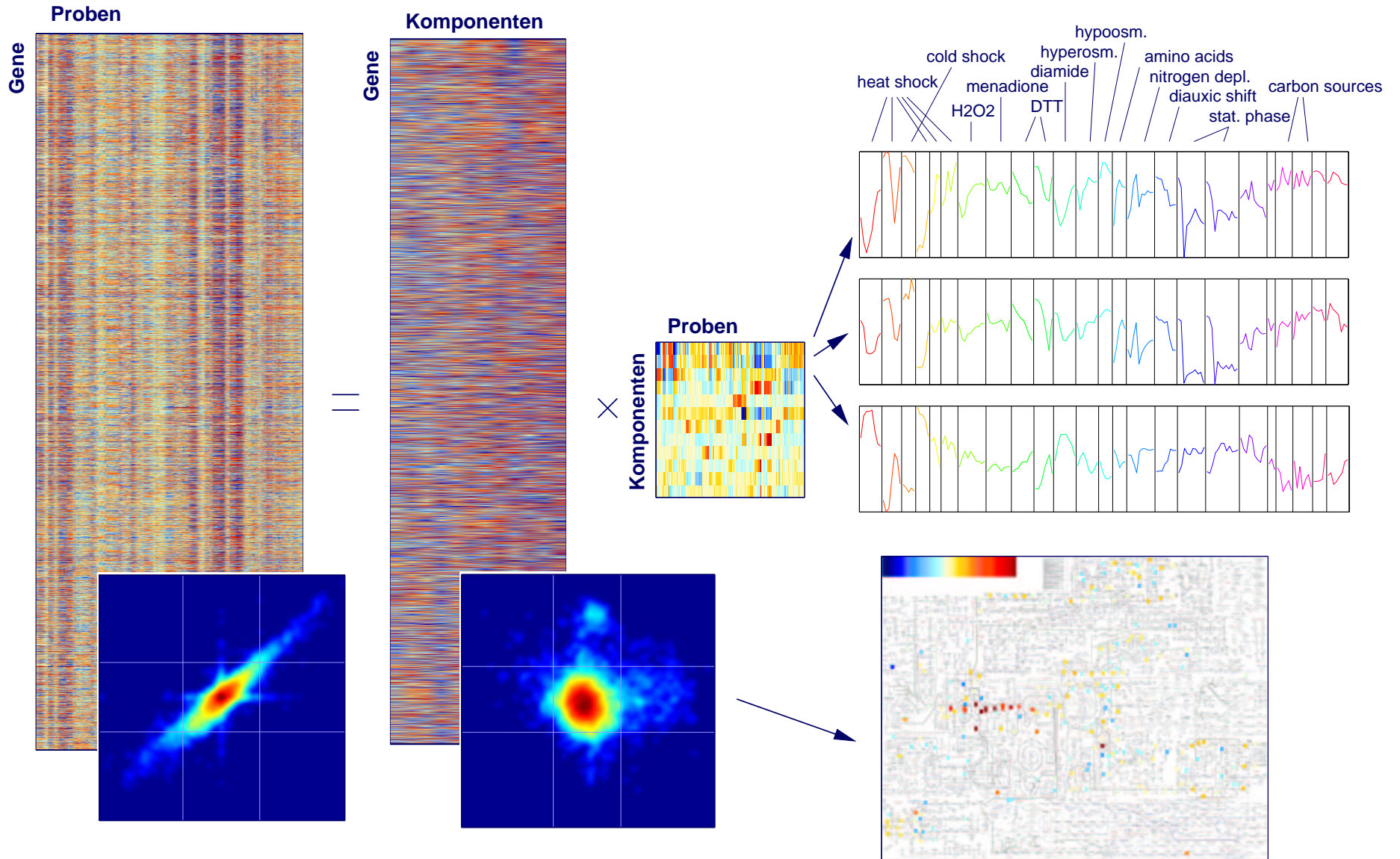
Lineares Modell (ICA)

Lineare Modelle

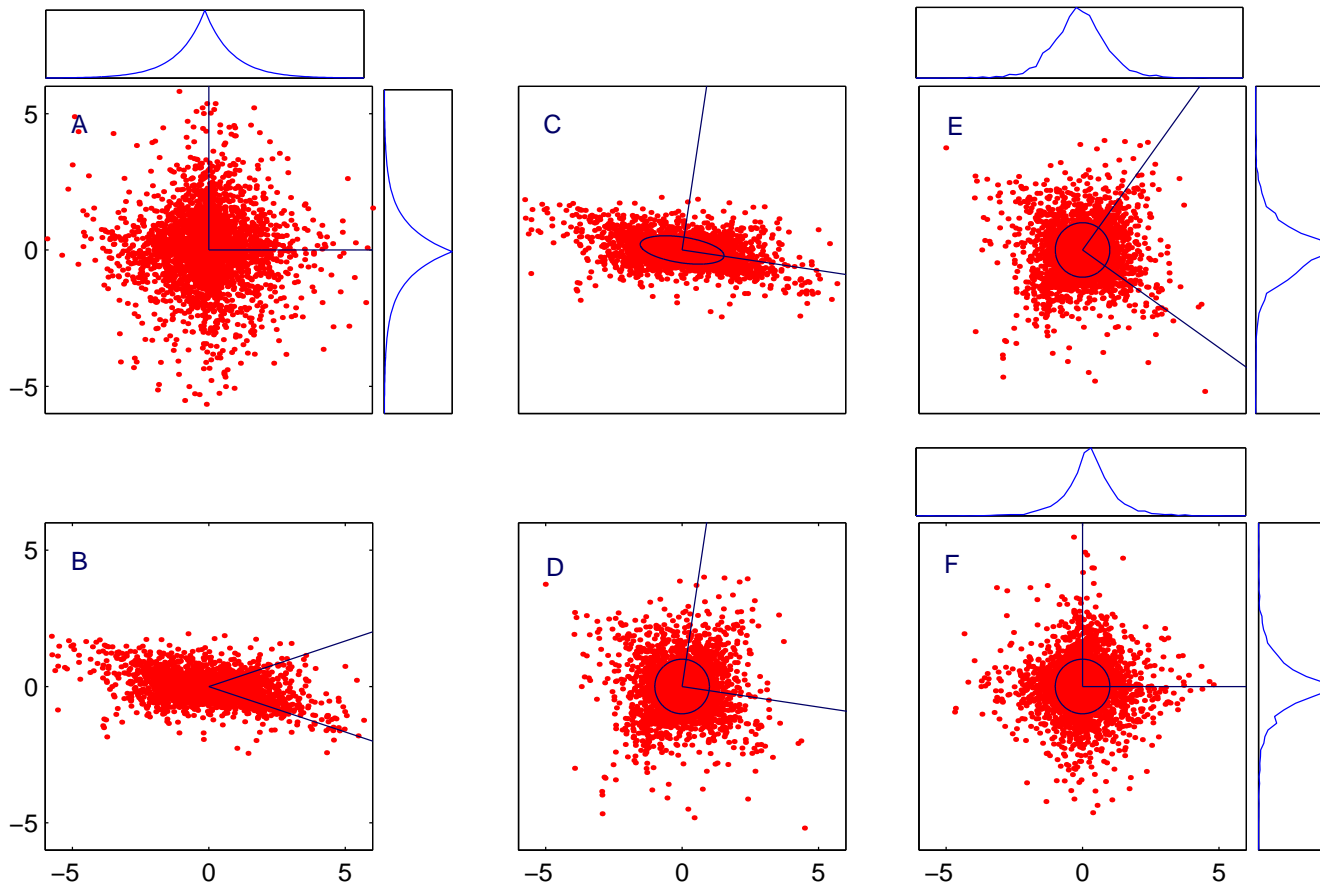


$$X = S \times Y + \eta$$

Independent component analysis (ICA)

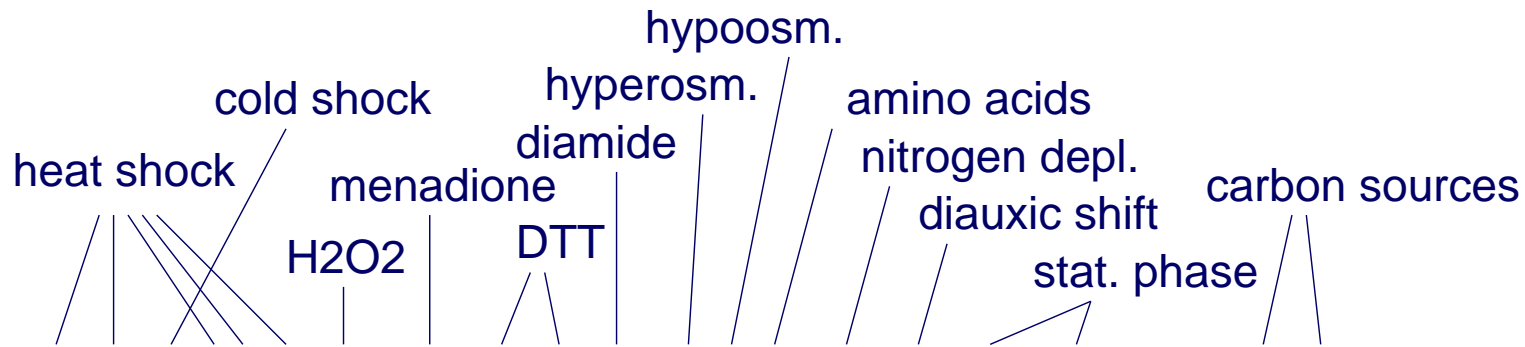


ICA

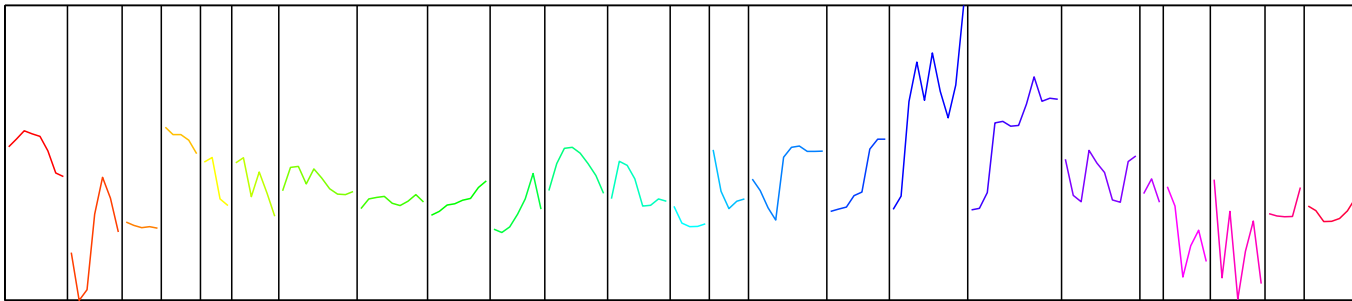


A. Wahre Komponenten C. Kovarianz E. Auffällige Randverteilungen
B. Künstliche Daten D. Entkorrelieren F. Rekonstruierte Komponenten

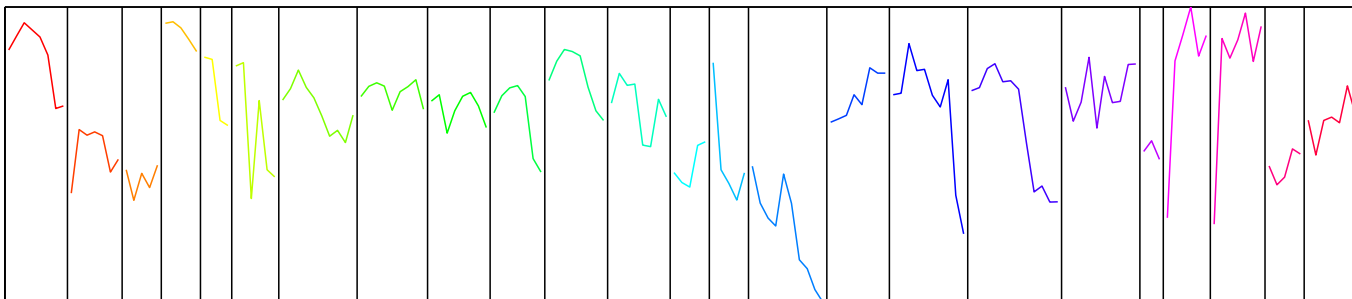
Nichtlineare ICA (Zellstress-Daten)



Mode 1

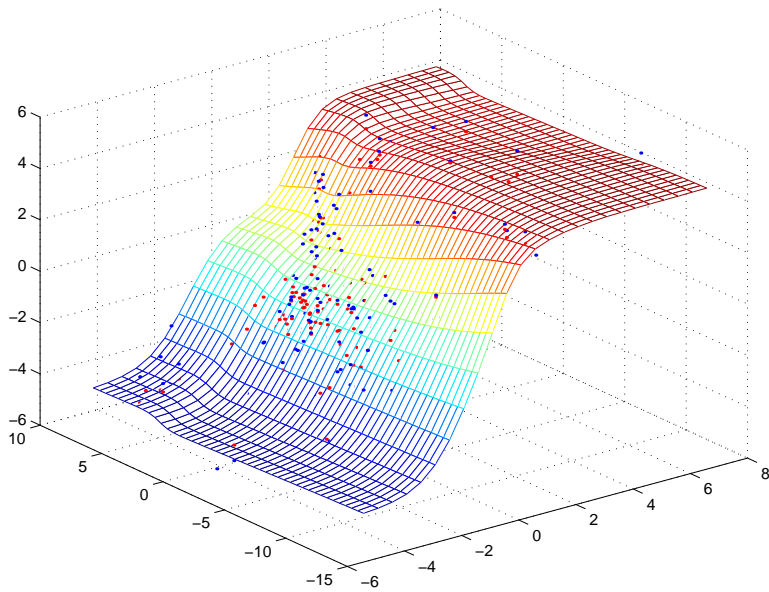


Mode 2

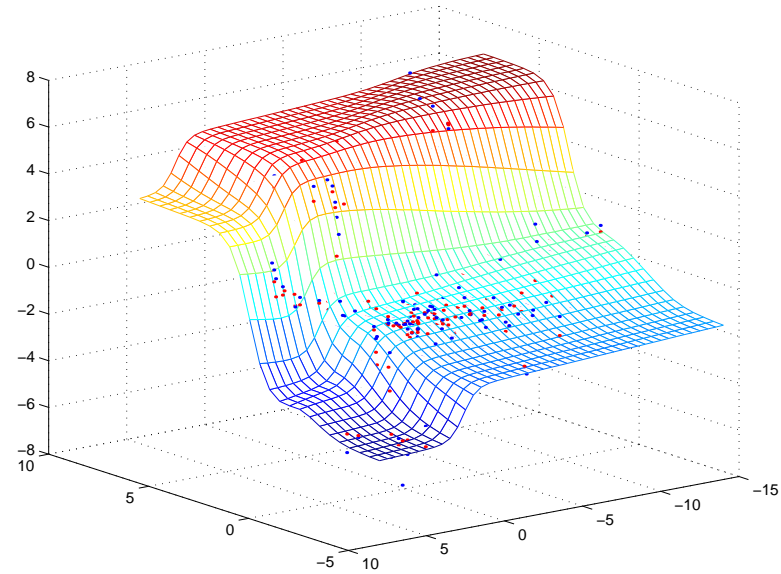


Nichtlineare Komponenten in Zellstress-Daten Gasch et al.

Nichtlineare ICA (Zellstress-Daten)



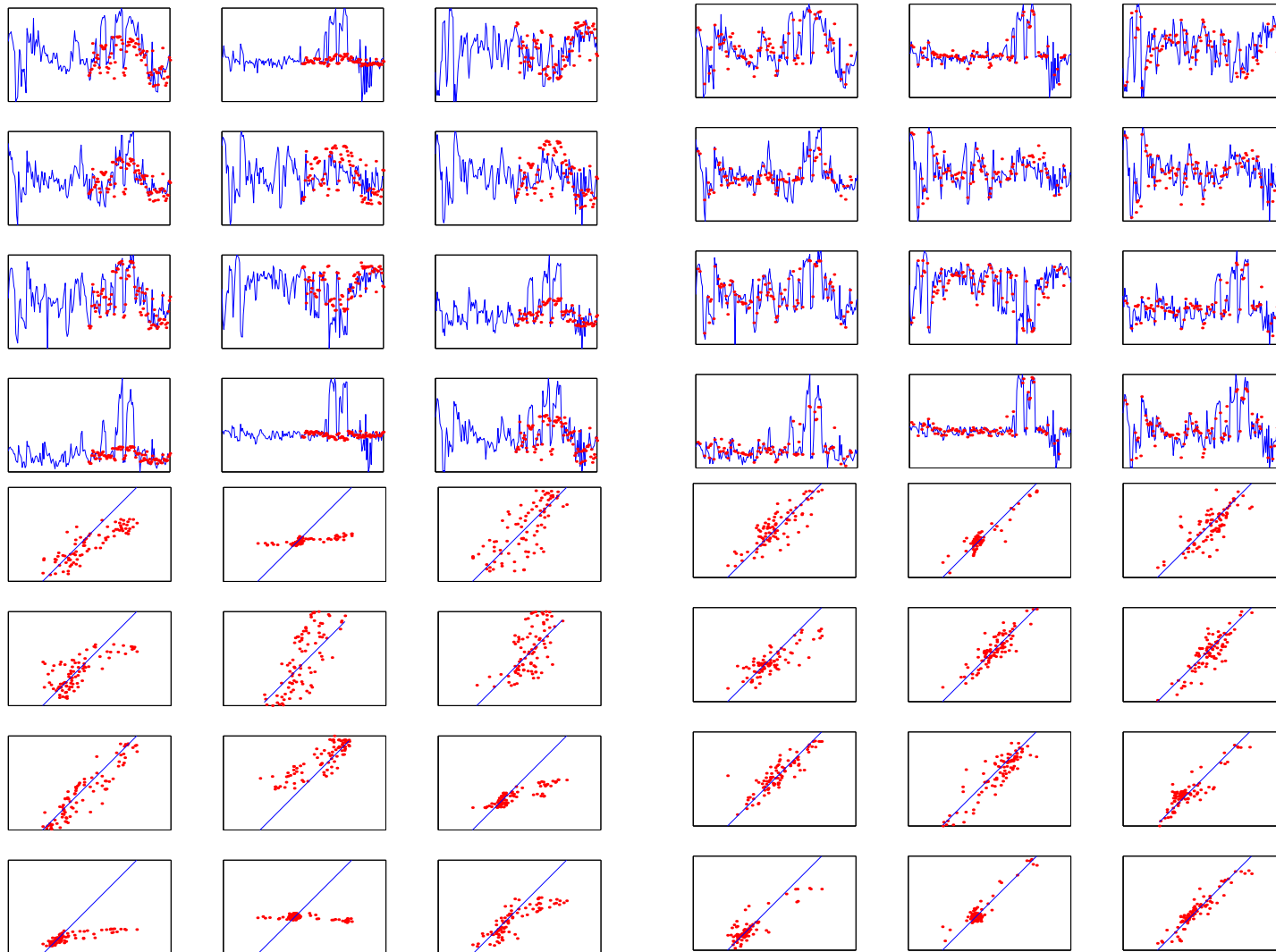
HSP26



JEN

Zellstress-Daten Gasch et al.

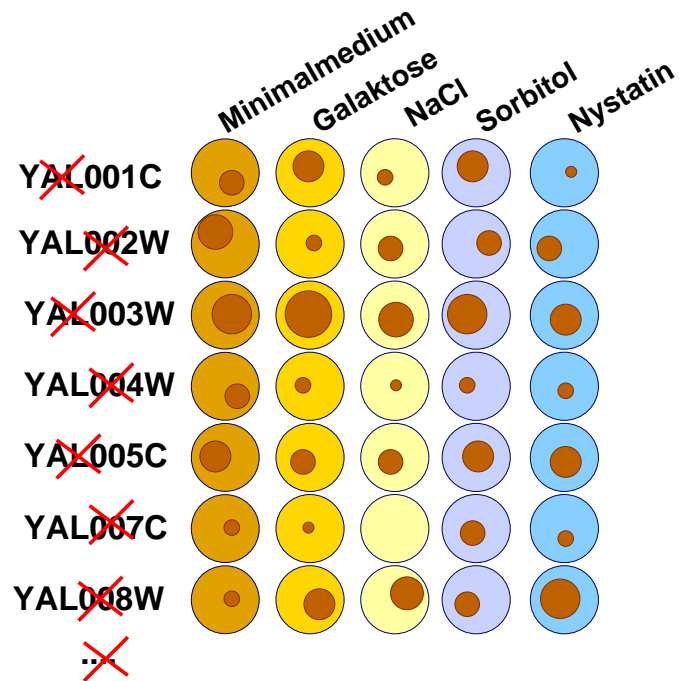
Nichtlineare ICA (Zellstress-Daten)



Blau: Genprofile

Rot: Vorhersage mit nichtlin. ICA

Expression und Fitnessverlust nach Deletionen



Interaktion Deletion/Medium

$$dx^T G_{x\alpha} d\alpha = \frac{d\hat{x}_i^T}{(G_{xx}^{-1})_{ii}} G_{xx}^{-1} G_{x\alpha} d\alpha$$

Diff. Expression ohne Deletion

$$d\bar{x} = -G_{xx}^{-1} G_{x\alpha} d\alpha$$

G. Giaever et al. (2002),

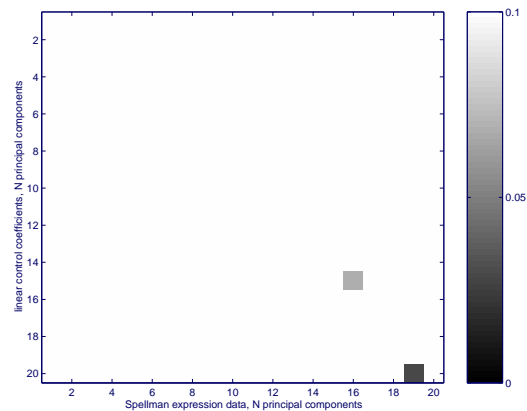
Functional profiling of the Saccharomyces cerevisiae genome, Nature 418

“The fact that such a small percentage of the genes that exhibit a significant increase in expression also exhibit a significant fitness defect was unexpected and warrants closer inspection.”

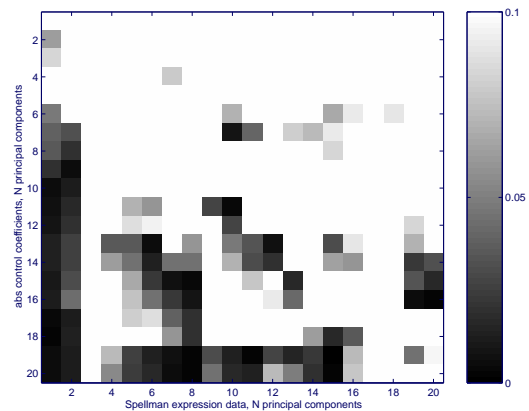
Expression und Kontrollkoeffizienten

- Vergleich von Expressionsdaten und Kontrollkoeffizienten
- Maß für Ähnlichkeit:
Winkel zwischen den Räumen der ersten Hauptkomponenten
- p-Werte aus Permutationstest

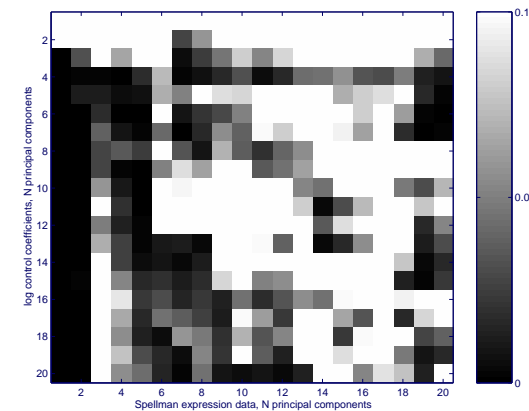
Kontrollkoeffizienten C^J



$|C^J|$



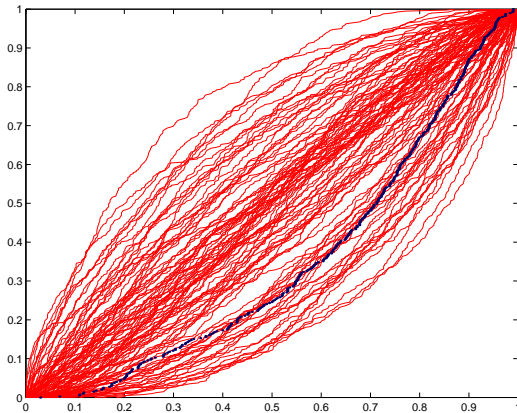
$\log|C^J|$



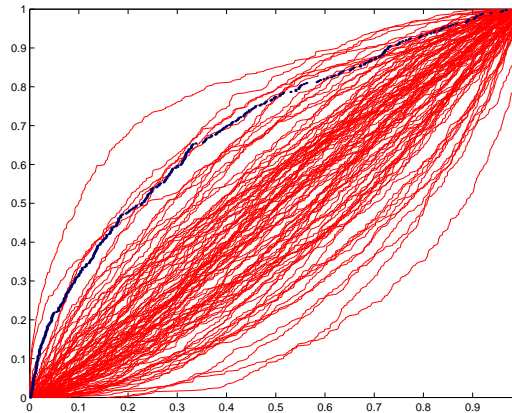
Expression und Kontrollkoeffizienten

Verteilung der p-Werte

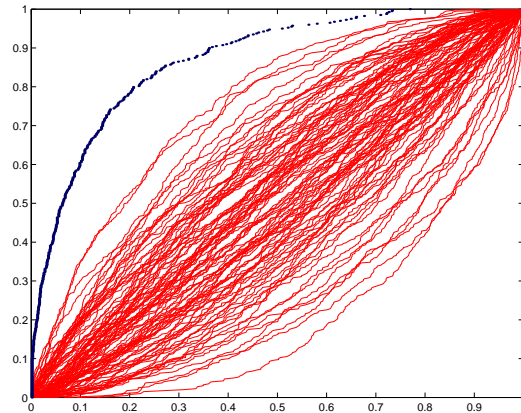
Kontrollkoeffizienten C^J



$|C^J|$



$\log|C^J|$

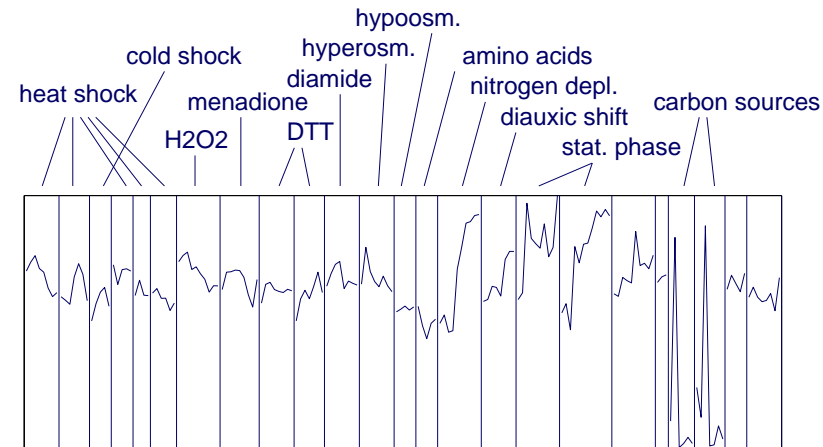
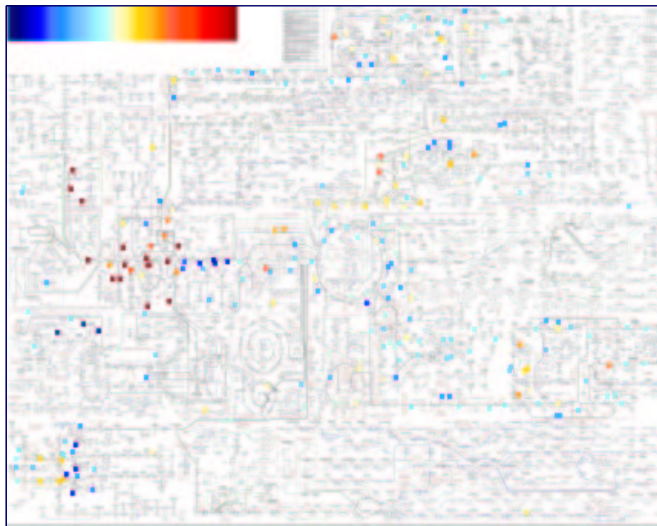


Schwarz: Kumulative Verteilungen der p-Werte

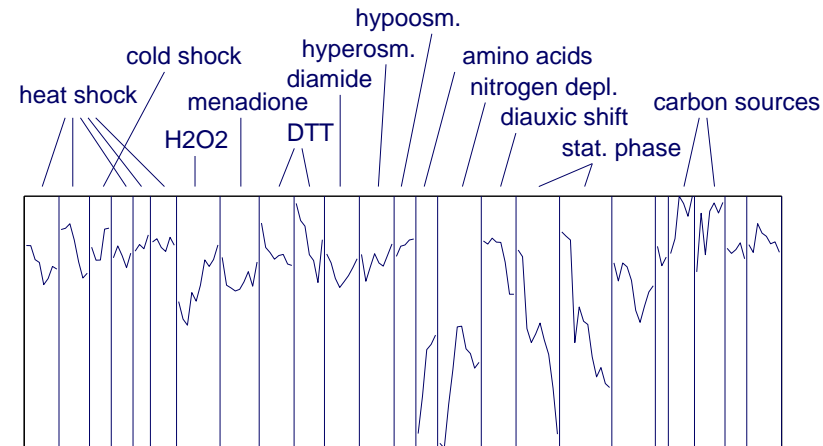
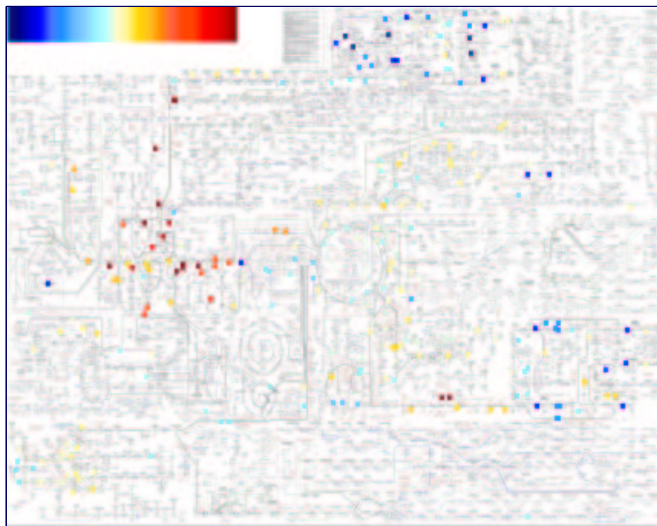
Rot: Ergebnis unter Nullhypothese (Permutationstest)

Daten: Hefe-Zellzyklus, Spellman et al. 1998

Expression und Kontrollkoeffizienten: ICA

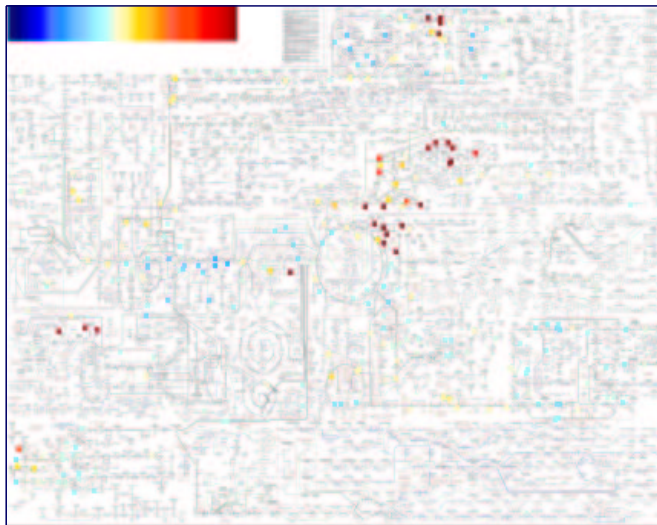


Galaktoseabbau und obere Glycolyse (Wachstum auf Galaktose)

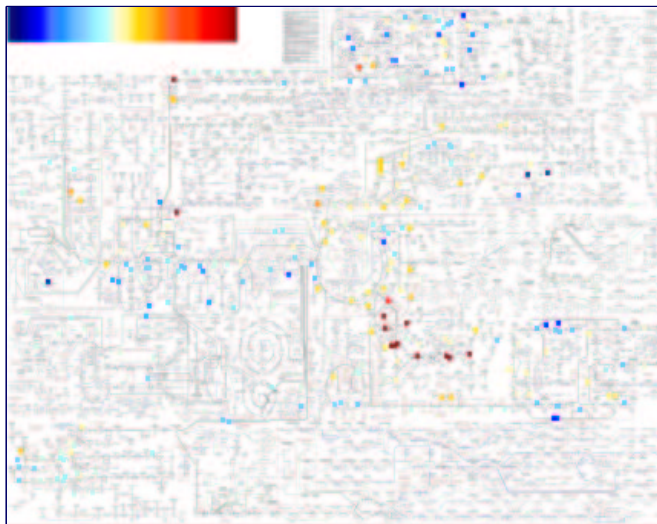
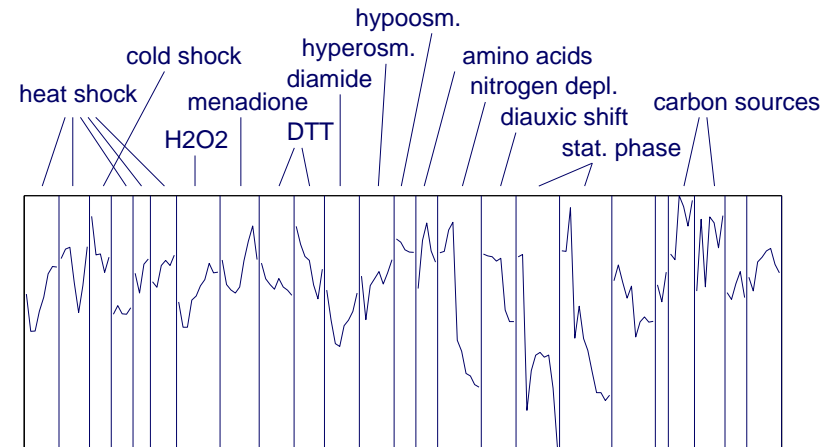


+ Glykolyse, - Nukleotidsynthese

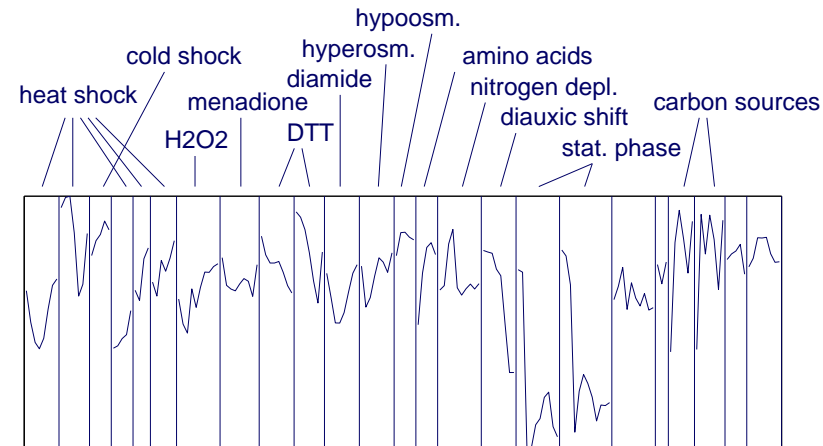
Expression und Kontrollkoeffizienten: ICA



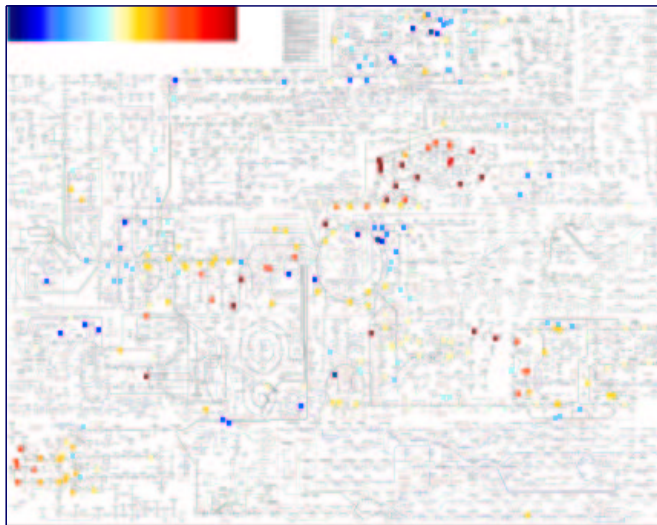
Aminosäuresynthese



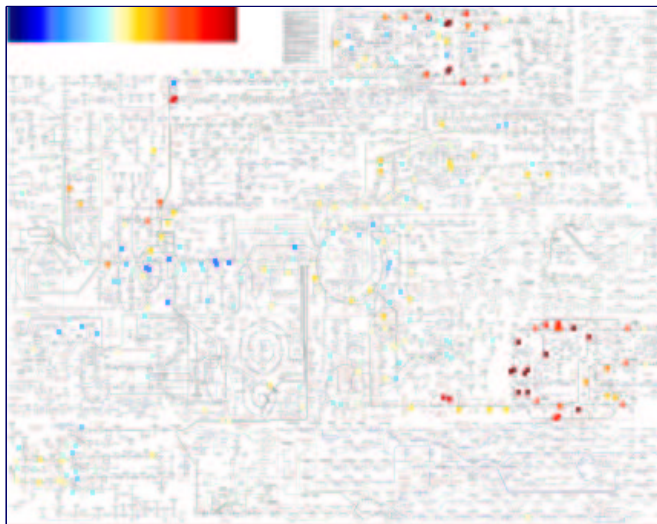
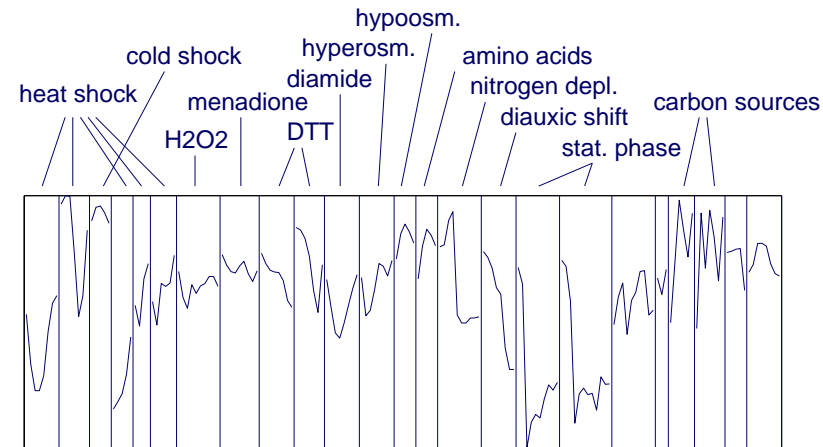
Prolinweg



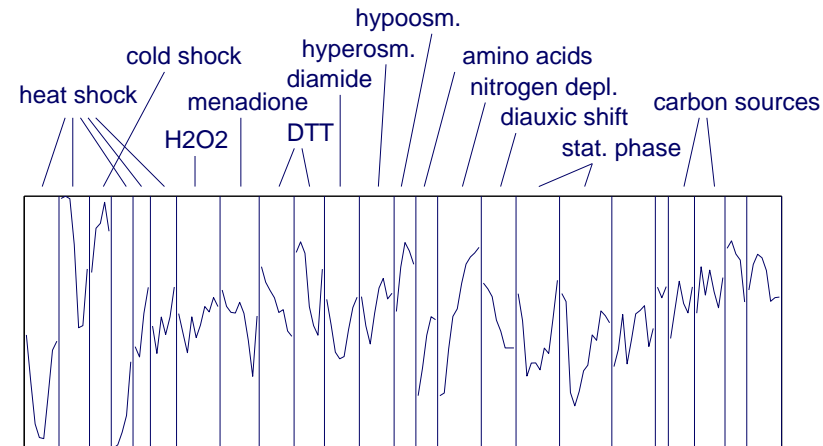
Expression und Kontrollkoeffizienten: ICA



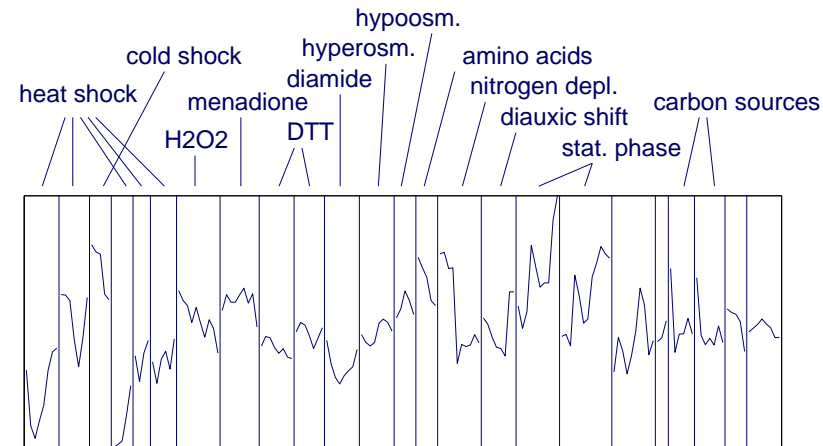
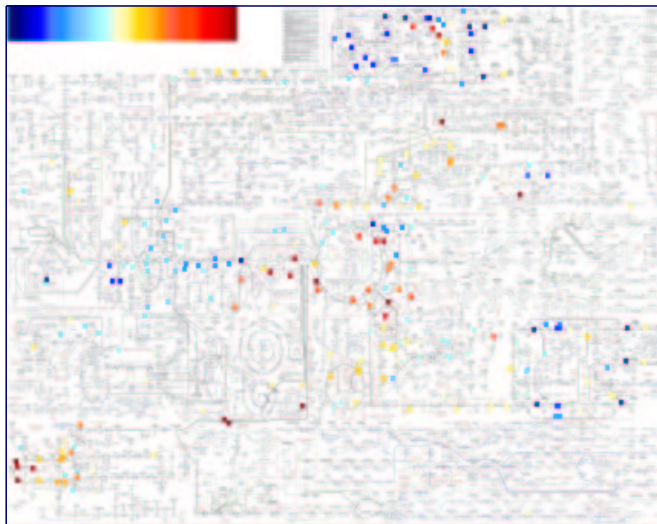
Aminosäuresynthese



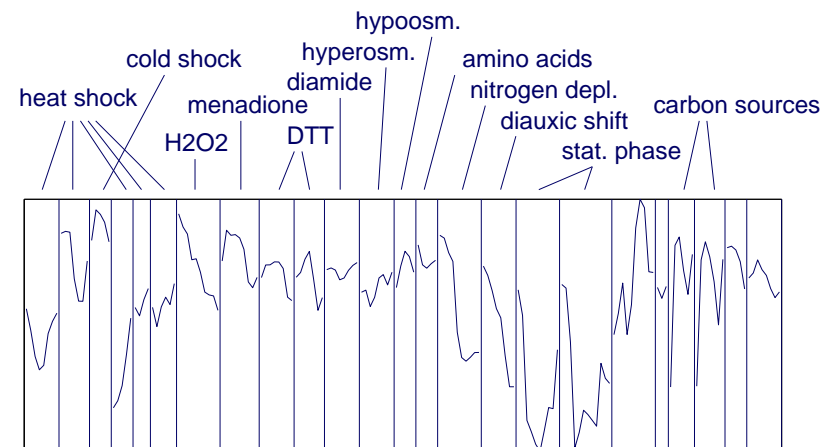
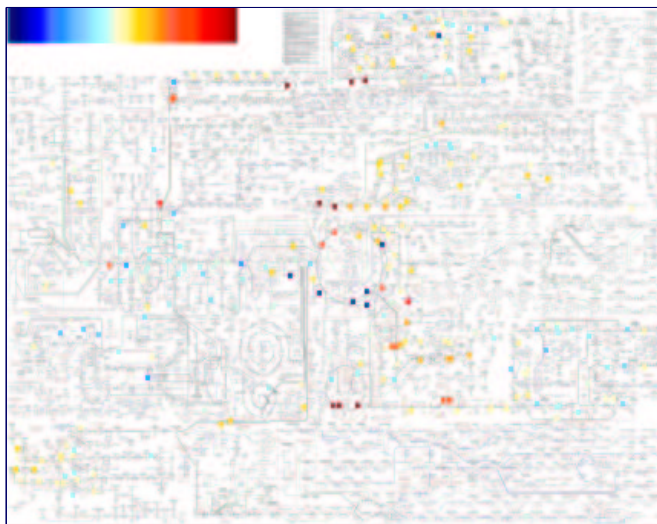
Nukleotidsynthese



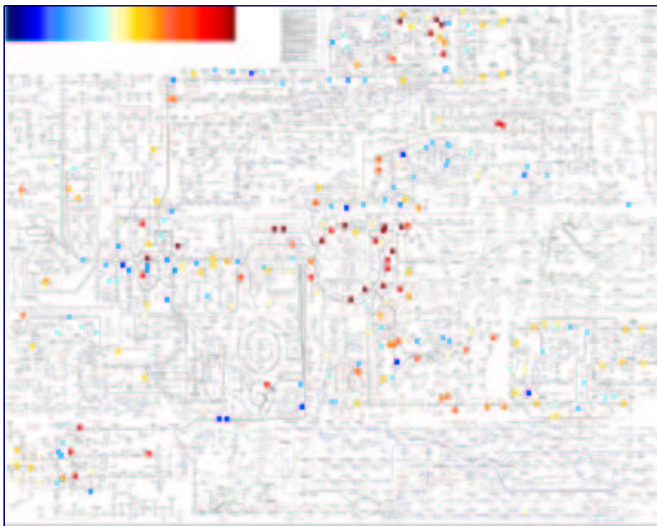
Expression und Kontrollkoeffizienten: ICA



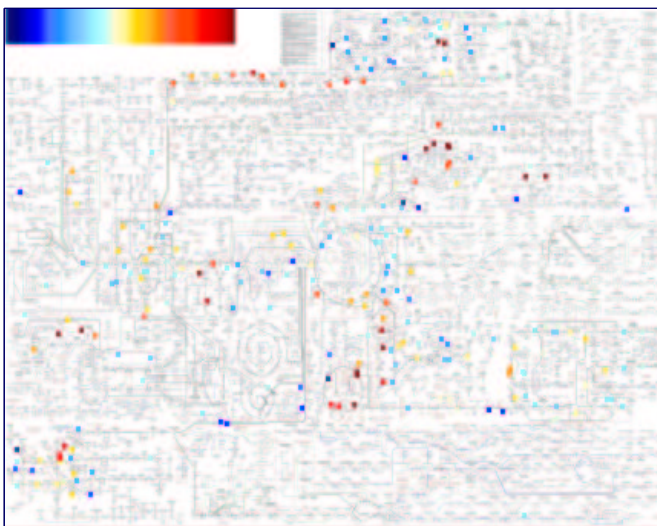
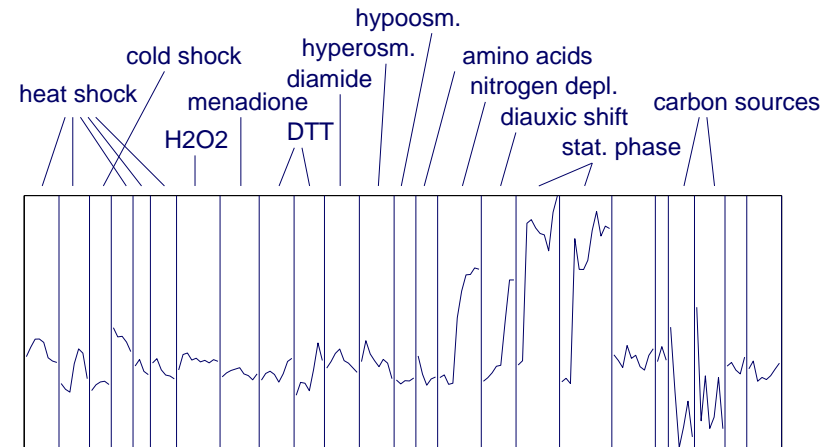
+ Zitratzyklus, - Glykolyse -Nukleotidsynthese



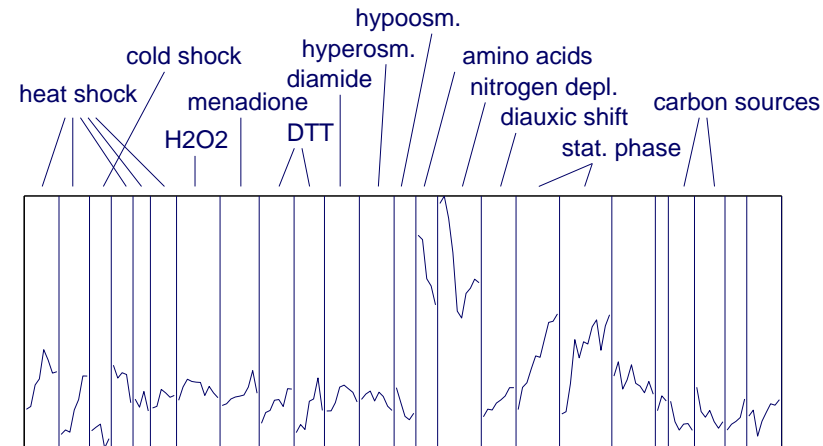
Expression und Kontrollkoeffizienten: ICA



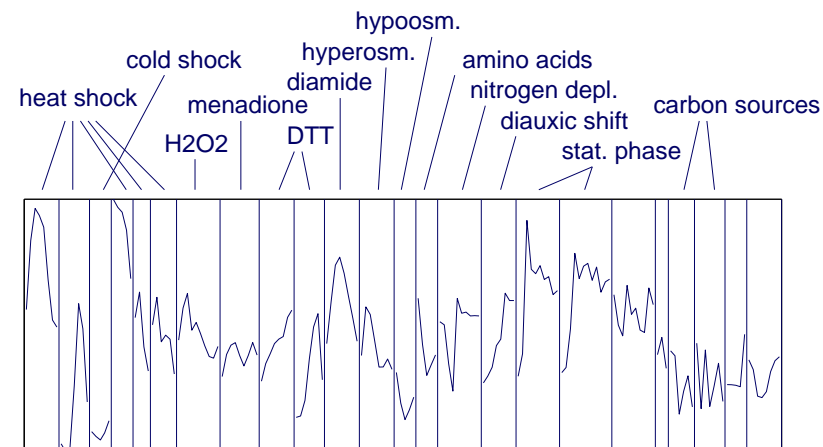
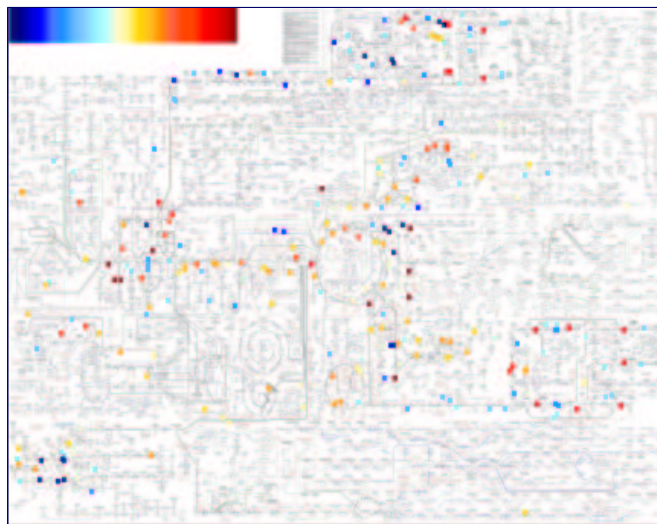
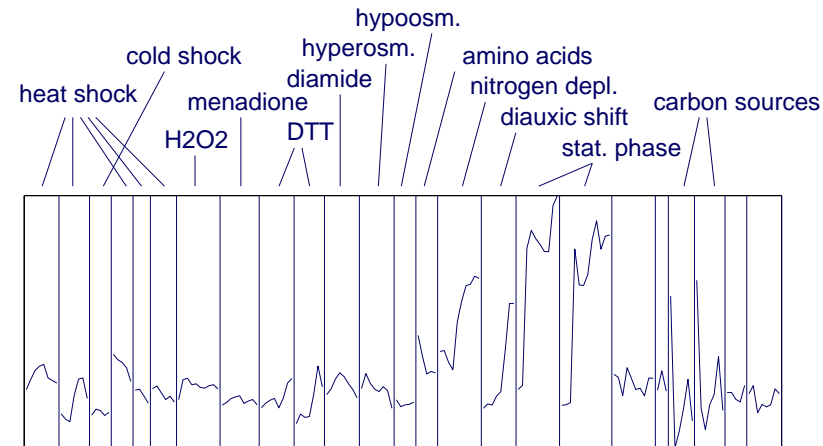
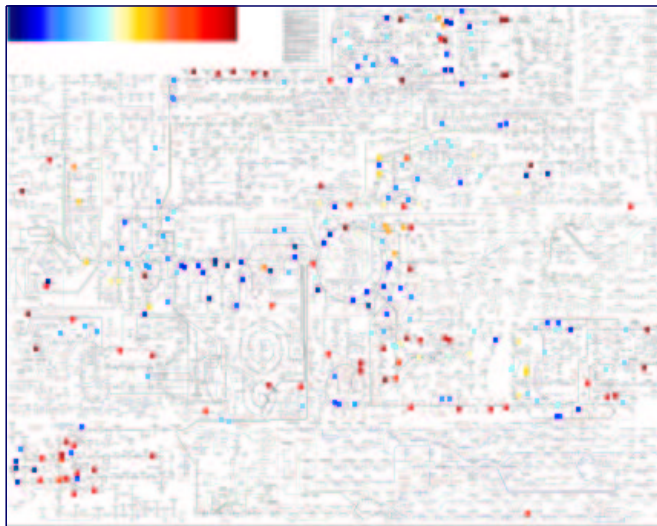
Zitratzyklus



Reaktion auf Minimalmedium (- Aminosäuren,- Stickstoff)



Expression und Kontrollkoeffizienten: ICA



Ähnlich dem environmental stress response cluster (Gasch et al., 2000)