

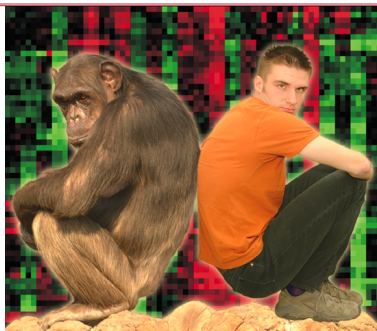
Auf molekularer Ebene, was unterscheidet ...



- Hautkrebs und gesunde Haut?
- Behandelbare und unheilbare Tumore?
- Menschen mit und ohne Krebsrisiko?
- Patienten mit und ohne Cystischer Fibrose?

Und was unterscheidet eigentlich ...

...Mensch



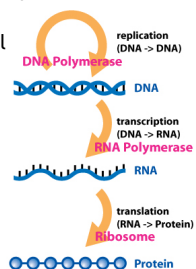
...und Schimpanse?

Von DNA zur Expression



- Vergleich nur auf genomischer Ebene nicht ausreichend
- Vergleich darüber hinaus auf der Expressionsebene
- Die **Expression** der genetischen Information in einem DNA-Molekül findet in zwei Stufen statt:

- (i) der **Transkription**:
 $\text{DNA} \Rightarrow \text{mRNA}$
- (ii) der **Translation**:
 $\text{mRNA} \Rightarrow \text{Protein}$



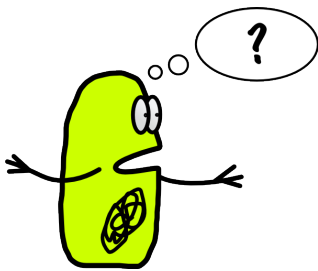
Transkriptom



Genomweite Genexpression:

wann und in *welcher Menge* wird jedes Gen einer Zelle exprimiert

Microarrays?



Was ist ein Microarray?



Ein **Microarray** ist ein Miniatur-Labor auf einem Chip, das erlaubt, sehr große Mengen biologischen Materials in einem sogenannten **Hochdurchsatzverfahren** zu untersuchen.

Es gibt sehr viele Varianten:

- DNA-Microarrays
- Protein-Microarrays
- Gewebe-Microarrays
- Antikörper-Microarrays

DNA-Microarrays

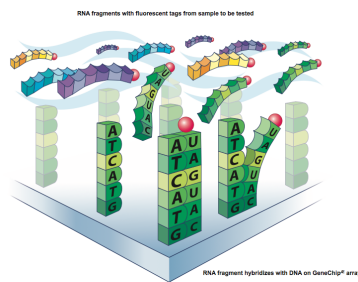


- Ein DNA-Microarray besteht aus einer kleinen Membran oder Glasplatte auf der DNA-Sonden in geordneter Weise aufgebracht sind.
- Markierte cDNA-Sequenzen aus einer biologischen Probe **hybridisieren** mit ihrem komplementären Gegenpart auf dem Array
- Microarrays erlauben u.a. die parallele Analyse der **Expression** mehrerer tausend Gene bzw. des gesamten **Transkriptoms** einer Zelle
- Synonyme: Chip, Biochip, DNA-Array, Gene-Array, GeneChip® (geschützter Name der Firma Affymetrix)

Definition Microarray:



Prinzip von DNA-Microarrays:
komplementäre **Hybridisierungseigenschaften**
von Nukleinsäuren

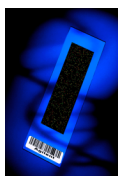


www.affymetrix.com

DNA-Microarrays



Affymetrix GeneChip



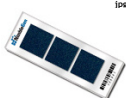
Agilent
http://www.agilent.com/about/newsroom/fscs/images/library/images/cag_43_microarray_dna.jpg



ABI



Illumina BeadChip



Nimblegen
<http://www.nimblegen.com/>

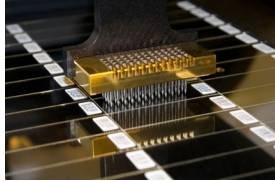


Febit
http://www.xconomy.com/wordpress/wp-content/images/2008/03/dna_cartridge_earth_icon.jpg

Zwei Typen von Microarrays



- **Spotted Arrays:** Sonden (cDNA mit 500-5,000 Basen oder 25-60bp Oligos) werden mit Hilfe von Robotern auf Array aufgebracht und immobilisiert.



http://www.digitallapoptosis.com/archives/science/Microarray_Printer.jpg

Zwei Typen von Microarrays



- **in situ Arrays:** Proben = Oligos werden *in situ* via **Photolithographie** auf dem Array direkt synthetisiert.

Beispiel: **GeneChip®** von Affymetrix

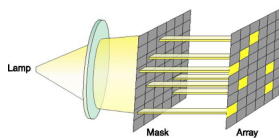


Image courtesy of Affymetrix

Einsatz



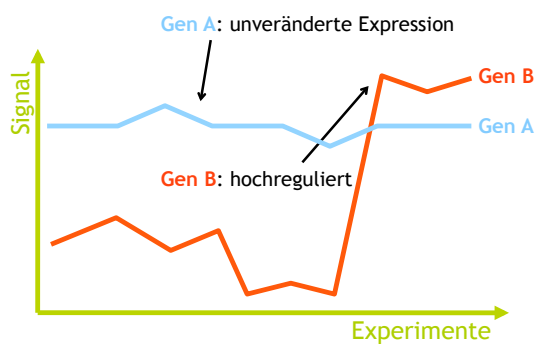
- Ein häufiger Gebrauch von DNA-Arrays ist die Expression zwei Zellpopulationen zu vergleichen:
 - welche Gene sind aktiviert (hochreguliert)
 - welche Gene sind reprimiert (herunterreguliert)

Beispiele



- Vergleich der Expressionstätigkeit von Genen in gesunden Zellen und in Tumorzellen
- Vergleich der Expressionstätigkeit von Genen in Hefe, wenn diese unter anaeroben (Bier) bzw. aeroben (Brot) Bedingungen wächst.

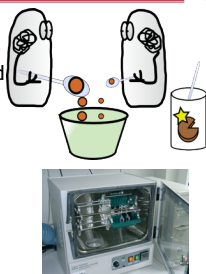
Expressionsprofile



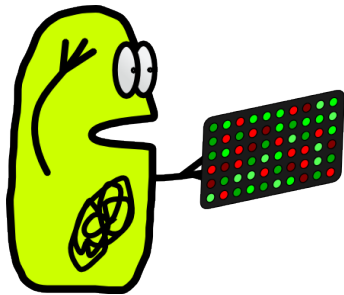
Versuchsablauf



- Wir isolieren mRNA aus den jeweiligen Proben
- Diese wird in cDNA umgewandelt und mit Farbstoffen markiert
- Dann wird jede markierte cDNA in einem speziellen Ofen mit dem Microarray für ca. 16 Stunden ...
- ... „gebacken“ (hybridisiert).
- Mit Laserlicht wird das Ergebnis beleuchtet: an jeder Stelle, an der cDNA auf dem Microarray gebunden wurde, leuchtet es. Je heller die Intensität, desto mehr cDNA hat gebunden



Datenanalyse



Datenanalyse



Microarrays erzeugen **große** Datenmengen

- Beispiel:
10 Experimente: 2x5 Replikate,
200 000 oligos (RNA/DNA-Segmente)
auf dem Microarray
=> $200\,000 \times 10 = 2\text{Mio}$ Datenpunkte

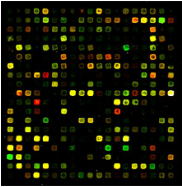
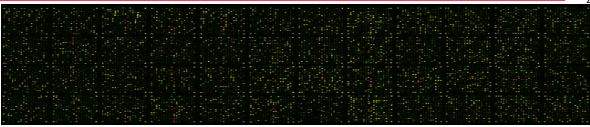
Microarrays und Bioinformatik



Bioinformatik auf verschiedensten Ebenen notwendig:

- Bildanalyse
- Visualisierung
- Clustering
- Statistik
- Datenbanken
- ...

Bildverarbeitung: Dual Channel Arrays



Schwarz:
Gen nicht
exprimiert

Rot:
Gen nur in A
exprimiert

Grün:
Gen nur in B
exprimiert

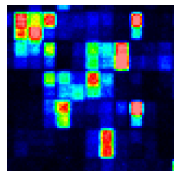
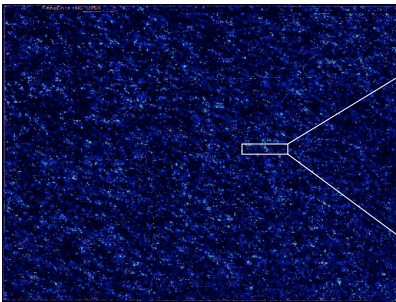
Gelb:
Gen in A und B
exprimiert

- Zwei Samples, zwei Farbmarkierungen (mit spez. Anregungs-Wellenlänge $\neq$ Emissions-Wellenlänge)
- Pro Farbkanal ein Scan
- Integration beider „Kanäle“ -> ein Bild mit „Pseudofarben“.
- Typischerweise werden die Farben rot, gelb, grün und schwarz benutzt, da diese die eingesetzten Farbstoffe Cy3 (grün) und Cy5 (rot) symbolisieren.

Bildverarbeitung: Single-Channel Array



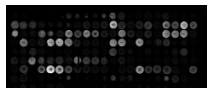
Hier: GeneChip von Affymetrix



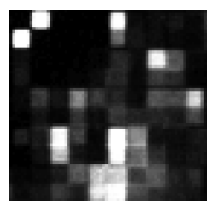
Bildverarbeitung



- Schwierig für spotted arrays



- Leichter für *in situ* arrays

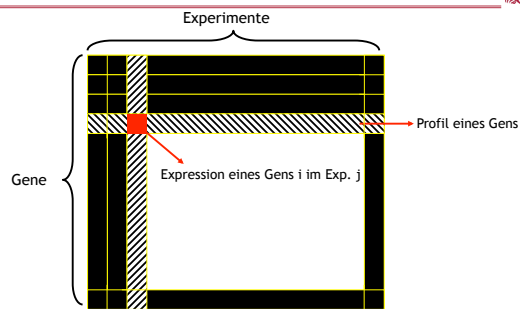


Normalisierung



- Microarrayexperiment: gesucht sind Gene, die z.B. zwischen zwei Bedingungen **unterschiedliche Expression** zeigen
- Jedes Microarray-Experiment birgt technisch bedingte Fehlerquellen, die zur Variation des gemessenen Signals beitragen
 - Falsche Kalibrierung der Messgeräte
 - Scanning-Parameter sind unterschiedlich
 - Farbstoffeigenschaften
- Computergestützte Datentransformation um technisch bedingte Fehler zu korrigieren

Expressionsmatrix

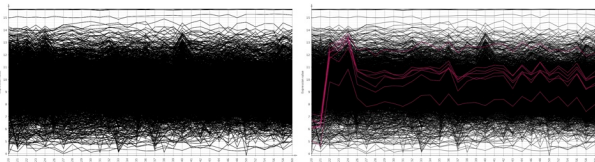


Wertebereich: $1 \cdot 2^{16}$ bzw. nach $\log_2$ -Transformation zwischen 0-16

Expressionsprofile: Profilplot



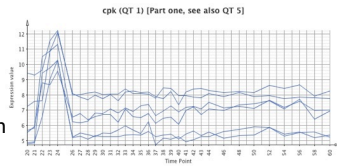
- Problem: zu viele Profile in einem Plot => Profile mit gemeinsamem Muster werden verdeckt



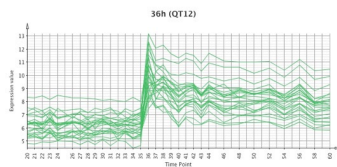
Expressionsprofile: Profilplot



- Profilplots von Gruppen von Genen mit ähnlichem Profil
- -> Co-Expression
- -> evtl. Co-Regulation



- Anwendung von Clusteringmethoden auf Expressionsdaten



Clustering: Definition



- Gegeben eine Menge von n Objekten, die jeweils p Werte aufweisen
- Ein Clustering ist eine Klassifikation der Objekte in **Klassen = Cluster** derart, dass
 - Objekte innerhalb eines Clusters eine kleine Distanz haben
 - Objekte in verschiedenen Clustern eine große Distanz haben

Clustering



Zwei Zutaten:

- **Distanzmaß**
z.B. Euklidische Distanz
- **Cluster-Algorithmus**
z.B. k-means

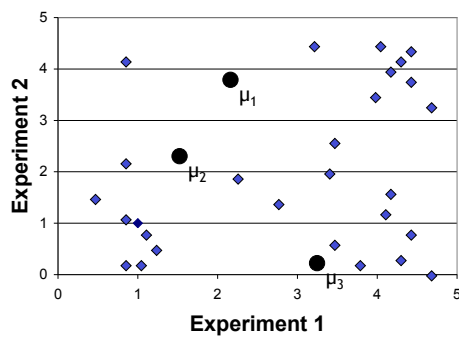
k-means (Lloyd's) Algorithmus



- 1) Wähle zufällig k Datenpunkte als Clusterzentren $\mu_1, \dots, \mu_k$ aus
- 2) Für jedes Gen x berechne seinen nächstgelegenen Clusterzentroiden:

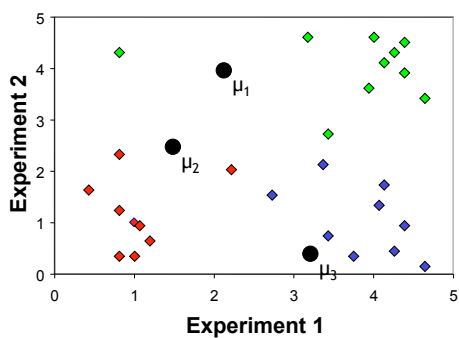
$$C(i) = \operatorname{argmin}_{1 \leq l \leq k} d(x_i, \mu_l)^2$$
- 3) Für jeden Cluster berechne neuen Clusterzentroiden
- 4) Wiederhole Schritte 2-3 bis Algorithmus konvergiert

k-means Algorithmus - Demo

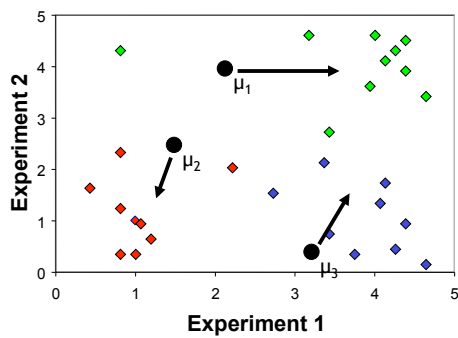


Setze k , hier $k=3$, Initialisierung: hier zufällige Punkte als Zentroide

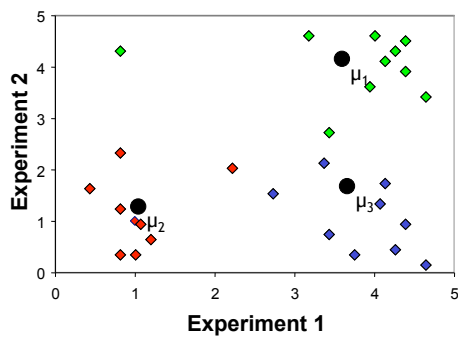
k-means Demo: Schritt (2)



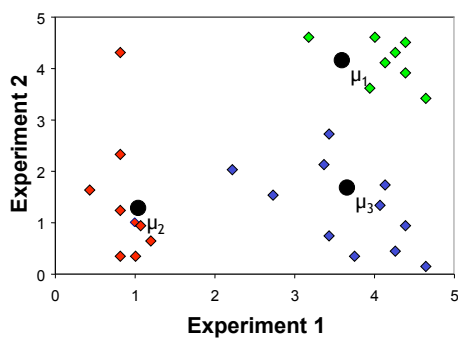
k-means Demo: Schritt (3)



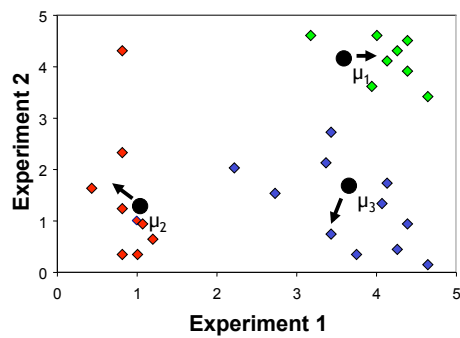
k-means Demo: Erg. Schritt (3)



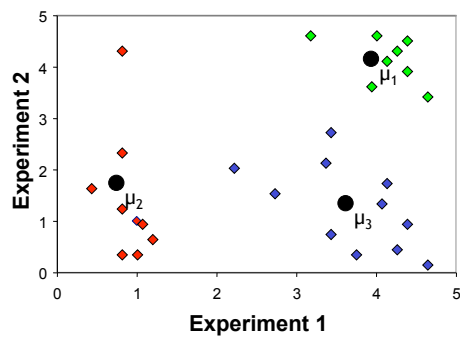
k-means Demo: Schritt (2)



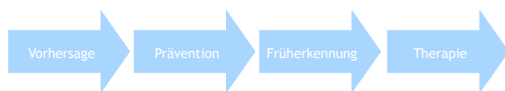
k-means Demo: Schritt (3)



k-means Demo: Erg. Schritt (3)



Anwendung: Personalisierte Medizin



Genetische Merkmale berücksichtigen

- Risiken vorhersagen
- Risiken minimieren
- Krankheiten früh erkennen
- Bessere Therapie

Jeder Mensch ist einzigartig



Unverändert SNP
 Gen: AAG-CGA-ATT-AGG AAG-GGA-ATT-AGG
 Protein: Lys - Gln -Ile -Arg Lys - Gly -Ile -Arg

0,1% von 3.000.000.000
 ergeben 3.000.000 Unterschiede

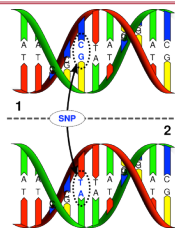
Polymorphismen einzelner Basenpaare (SNPs)

- Häufigster Typ von genomischer Variation bei Menschen
- Erbkrankheiten (6000 bekannt)

Microarray-Variante: SNP-Arrays



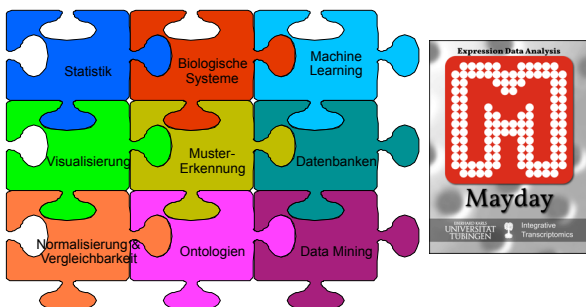
Sonden enthalten
 bekannte (bis zu ca.
 500.000) SNPs



Beispiel BRCA1 und BRCA2

- „breast cancer susceptibility gene“ 1 und 2.
- Hohes Risiko für Brustkrebs bei bestimmten Mutationen
- Vorsorge möglich

Viele Aufgaben - eine Plattform



<http://www-ps.informatik.uni-tuebingen.de/mayday/>
