



Vorlesung: Einführung in die Bioinformatik (Sommersemester 2014)

Übungsblatt Nr. 5: It is hip to chip

Elektronische Abgabe an: Joachim.Fritscher, per Email an joachim.fritscher@student.uni-tuebingen.de

1. k-Means Clustering

Führen Sie ein k -means Clustering “per Hand” (wie in der Vorlesung) auf folgenden fiktiven Expressionsdaten, welche die Expressionsstärke von 7 Genen in 2 Bedingungen (einmal mit Sauerstoff, einmal ohne Sauerstoff) zeigt, durch:

Gen	mit O_2	ohne O_2
Gen 1	1.1	60
Gen 2	8.2	20
Gen 3	4.2	35
Gen 4	1.5	21
Gen 5	7.6	15
Gen 6	2.0	55
Gen 7	3.9	39

Visualisieren Sie zunächst die Gene in der $(+O_2, -O_2)$ -Ebene. Führen Sie für $k = 2, 3, 4$ jeweils ein k -means Clustering mit zufälligen Startzuweisungen durch. Wie sehen die Cluster aus (sprich welche Gene sind in einem Cluster für ein gewähltes k)?

2. Co-Expression und Profilplot

Zeichnen Sie einen Profilplot der folgenden Expressionsdaten. Bestimmen Sie anhand der Profile, welche Gene co-exprimiert sind.

Gen	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Gen 1	3.2	3.1	10.5	3.1
Gen 2	8.2	8.2	4.0	4.2
Gen 3	4.5	11.6	5.6	10.9
Gen 4	3.7	10.8	4.0	11.2
Gen 5	2.8	2.9	11.0	3.1
Gen 6	5.1	5.3	5.2	5.5
Gen 7	9.0	9.3	4.5	4.6

3. Expressionsanalysen von Leukämie

Lesen Sie zur Bearbeitung dieser Aufgabe das Paper von Golub *et al.*¹ und informieren Sie sich über die beiden Krebstypen *akute lymphoblastische Leukämie* (ALL) und *akute myeloische Leukämie* (AML).

Im Folgenden sollen einige der Gene, die von den Autoren als Markergene für entweder Typ ALL oder Type AML identifiziert wurden, näher betrachtet werden. Dabei gelten Gene als Marker für Typ AML, wenn sie in diesem und nicht in AML aktiv, also hoch exprimiert sind und umgekehrt.

- (a) Ergänzen Sie dazu zuerst mit Hilfe des Papers (und der dort vorhandenen Abb. 3) die nachfolgende Tabelle.

Index	Gen-Name	Gen aktiv in Krebstyp
A1	Zyxin	
A2	Cyclin D3	
A3	Myosin light chain	
A4	HOX A-9	
A5	SNF 2 A	
B1	Adenosine deaminase	
B2	Leptin receptor	
B3	OP 18	
B4	CD 18	nicht aktiv
B5	SRP9	
C1	Actin	AML + ALL
C2	IL7 receptor-33	
C3	CD-33	
C4	MCM 3	
C5	LYN	
D1	Myc 3	nicht aktiv
D2	ATPase	
D3	Adipsin	
D4	CD 19	nicht aktiv
D5	Catalase	
E1	IL8	
E2	Lysozyme	
E3	Topoisomerase II	
E4	Acyl-CoA dehydrogenase	
E5	Glucose-6-phosphat	AML + ALL

- (b) Versuchen Sie dann eine Diagnose (also ob der Patient an AML oder an ALL leidet) für die drei angegebenen Patienten zu erstellen. Dabei bedeutet ein *x*, dass das entsprechende Gen in dem Patienten “aktiv”, sprich exprimiert ist.

Patient 1:

	1	2	3	4	5
A		x	x		
B	x		x		
C	x			x	
D					
E			x		x

Patient 2:

	1	2	3	4	5
A				x	
B		x			
C	x				
D					
E					x

Patient 3:

	1	2	3	4	5
A	x			x	
B		x			
C	x		x		
D		x	x		
E					x

¹Golub *et al.* 1999, <http://www.sciencemag.org/content/286/5439/531.full>, pdf aus dem Uni-Netz zugänglich