



# Einführung in die Bioinformatik

Prof. Dr. Enno Ohlebusch, Prof. Dr. Hans Kestler, Dr. Karlheinz Holzmann,  
Tobias Badura

WS 16/17

## Übungsblatt 1

Abgabe bis 30.11.2016, 12:00. Lösungen bitte elektronisch an [tobias.badura@uni-ulm.de](mailto:tobias.badura@uni-ulm.de).

Die Übungsaufgaben sollen mit der Skript-Sprache R (<http://www.r-project.org>) durchgeführt werden.

### 1. Aufgabe (3): Matrix-Indizierung

(a) Erstellen Sie eine  $9 \times 9$  Matrix  $X$  mit

$$x_{ij} = \begin{cases} 1, & \text{wenn } j \in \{i, 10 - i\} \\ 0, & \text{sonst} \end{cases}$$

(b) Erstellen Sie eine  $9 \times 9$  Matrix  $X$  mit  $x_{ij} = d(i, j)$ , wobei

$$d(i, j) = \min(|i - 5|, |j - 5|)$$

(c) Erstellen Sie eine  $9 \times 9$  Matrix  $X$  mit

$$x_{ij} = \begin{cases} 1, & \text{wenn } i \text{ und } j \text{ gerade oder } i \text{ und } j \text{ ungerade} \\ 0, & \text{sonst} \end{cases}$$

### 2. Aufgabe (3): Imperative vs. funktionale Programmierung

Schreiben Sie eine Funktion „binarize“, die eine Abbildung eines Vektors  $x \in \mathbb{R}^d$  auf einen Vektor  $y \in \mathbb{B}^d$  nach folgender Vorschrift berechnet:

$$y_i = \begin{cases} 1, & \text{wenn } x_i > \frac{1}{d} \sum_{i=1}^d x_i \\ 0, & \text{sonst} \end{cases}$$

(a) Implementieren Sie zwei Möglichkeiten, diese Funktion zeilenweise auf eine Datenmatrix anzuwenden:

1. Verwenden Sie eine for-Schleife.
2. Verwenden Sie „apply“.

(b) Laden Sie in R den „Iris“-Datensatz mit dem Kommando „data(iris)“ und wandeln Sie ihn in eine Datenmatrix um („data.matrix“). Entfernen Sie die letzte Spalte. Wenden Sie „binarize“ auf den reduzierten Datensatz an. Beschreiben Sie ihr Ergebnis.

### 3. Aufgabe (7): Dotplot

Schreiben Sie eine Funktion die ein Dotplot generiert (in R). Eingabe sind zunächst zwei Nukleotid-Sequenzen. Erweitern Sie Ihre Funktion um die Filterung nach der k-Wortmethode. Testen Sie Ihre Funktion mit den Sequenzen  $S_1$  und  $S_2$ . Erzeugen Sie zusätzlich Dotplots für  $k = 2$  und  $k = 3$ .

- > Sequenz S\_1  
AACTTTGGAGAAT
- > Sequenz S\_2  
AACTTGCGAGAAC

---

Suchmaschinen bieten direkten Zugriff auf Informationen zu (fast) allen Themen. Einige bekannte Ausgangspunkte einer (Ihrer) Recherche sind:

- yahoo, google, msn, wikipedia, ...
- Entrez, PIR, PubMed, scholar ...

#### **4. Aufgabe (4): Wnt signalling**

Was bedeutet *Wnt signalling*? Verwenden sie Suchmaschinen, Datenbanken, (online) Literatur um sich zu informieren. Beantworten sie die folgenden Fragen:

1. Was bezeichnet Wnt signaling? Ein Protein? Organismus, Zelle, Netzwerk, ...?
2. Wo (und wann) kommt es vor?
3. Was ist die Aufgabe / Funktion des Wnt signaling?
4. Welcher Zusammenhang besteht zwischen *Wnt signaling* und Stammzellen?

Geben sie jeweils an wie, und wo Sie Antworten gefunden haben.

#### **5. Aufgabe (3): Literatursuche**

Finden sie Literatur, die ihr Wissen unterstützt. Literaturhinweise finden Sie typischerweise in den Datenbankeinträgen. Hilfreich ist auch das „Durchstöbern“ bekannter (einschlägiger) Journale.

1. Finden Sie die Web-Adresse der Zeitschriften Nucleic Acid Research (NAR) und Bioinformatics.
2. Finden Sie einen Artikel zu Wnt signalling. Können Sie auf die vollständige Arbeit zugreifen? Welcher Teil ist direkt in PubMed? Führen Sie ein korrektes Zitat an.
3. Die Zeitschrift NAR veröffentlicht jährlich Berichte über biologische Datenbanken. Finden Sie die aktuelle (neueste) Liste der Datenbanken. In welchen Datenbanken kann man weitere Informationen zu Wnt Signalling finden (vermuten)?

#### **6. Aufgabe (3): Obesity protein**

In den Neunzigern des letzten Jahrhunderts wurde von der Arbeitsgruppe um Jeff Friedman an der Rockefeller University ein Hormon identifiziert, das übergewichtige Mäuse abnehmen lässt.

1. Suchen Sie in PubMed nach allen Publikationen zum „obesity gene“ im Zeitraum 1990–1999. Geben Sie die gestellte Suchanfrage und die Anzahl der Publikationen an.
2. Welches weitere Gen wird hierbei häufig genannt und welches Hormon kodiert es?
3. Zitieren Sie die Originalpublikation, in der die Sequenz erstmals genannt wird.

#### **7. Aufgabe (3): Taxonomien**

In der „NCBI taxonomy database“ kann man die Taxonomien bestimmter Organismen durchsuchen.

1. Suchen Sie nach Einträgen zum Neandertaler. Wie kann es eine Proteinsequenz zu dieser längst ausgestorbenen Spezies geben?
2. Geben Sie ein solches Protein ohne das Stichwort „mitochondrion“ an. Was kann man darüber im Vergleich zum modernen Menschen sagen?
3. Finden Sie ein weiteres Beispiele für eine Sequenz eines ausgestorbenen Lebewesens.