

Übung

Name des Studierenden:

Datum:

Einführung

Bitte bereiten Sie folgende Probleme vor der Übung am Mittwoch. Die Klausur wird ähnliche Probleme haben (s. auch Probleme in den ersten zwei Übungen). Während matlab kein Prüfungsgegenstand ist, sind die in dem matlab-Tutorial besprochenen Themen prüfungsrelevant.

1 Polygene Krankheiten

Erkrankte Geschwisterpaare In dieser Übung wollen wir die erste publizierte affected-sib-pair-Analyse durcharbeiten. Für ein zufällig gewähltes Locus ist eine zufällige Segregation zu erwarten:.

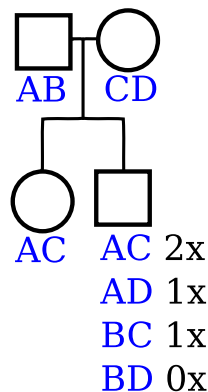


Abbildung 1: Zufällige Segregation.

- kein elterliches Allel gemeinsam: $1/4$ (AC – BD)
- ein elterliches Allel gemeinsam: $1/2$ (AC – AD und AC – BC)
- zwei elterliche Allele gemeinsam $1/4$ (AC – AC)
- Im Allgemeinen (ungeachtet des genetischen Modells) ist eine Abweichung von den Häufigkeiten $1/4 - 1/2 - 1/4$ dafür, dass ein Geschwisterpaar 0 – 1 – 2 elterliche Allele gemeinsam haben

Die ASP-Analyse prüft auf eine Abweichung von diesem Muster.

Beim Diabetes Typ 1 kommt es aufgrund einer Autoimmunreaktion zur Zerstörung der Beta-Zellen der Bauchspeicheldrüse, welche normalerweise das Insulin produzieren. Eine Reihe von Studien zeigten eine Assoziation von bestimmten HLA-Typen¹ mit dem Diabetes mellitus Typ 1. Das HLA-System verfügt über sehr viele Allele, so dass die Identifikation von Identität durch Abstammung (wie in der Abbildung) häufig möglich ist. Daher untersuchten Cudworth und Woodrow² HL-A-Haplotypen bei einer Reihe von Familien mit zwei oder mehr vom Diabetes mellitus Typ 1 betroffenen Geschwistern.

¹Das humane Leukozytenantigen-System (HLA-System, HL-Antigene, engl. Human Leukocyte Antigene) ist der Name des humanen Haupthistokompatibilitätskomplex (MHC). Unterschiedliche HLA-Allele sind mit verschiedenen Autoimmunerkrankungen assoziiert.

²Cudworth AG, Woodrow JC (1975) Evidence for HL-A-linked genes in "juvenile" diabetes mellitus. British Medical Journal 3:133–135.

Aufgabe: Bestimme die Anzahl Geschwisterpaare mit jeweils 0,1 oder 2 IBD-Allele. Berechne die erwarteten Anteile unter der Voraussetzung, dass die HL-A Allele keine Verbindung mit dem Diabetes haben. Verwende den χ^2 -Test (s. unten), um die statistische Signifikanz der Ergebnisse zu prüfen.

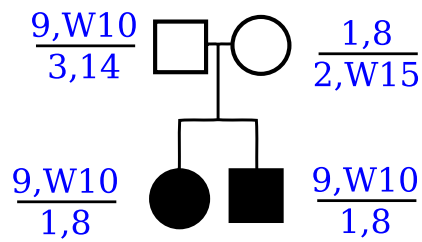


Abbildung 2: Familie mit zwei diabetischen Kindern mit Angabe des HLA-Haplotyps (die zwei Loci sind durch Komma getrennt). Es liegt eine Identität durch Abstammung (identity by descent, IBD) für zwei Allele vor.

<i>Familie</i>	<i>Kind #1</i>	<i>Kind #2</i>	Anzahl gemeinsamer Haplotypen (0–1–2)
1	2,12 / 2,12	2,12 / 2,12	
2	W28,W18 / -,W10	W28,W18 / -,W10	
3	2,12 / 9,8	2,12 / 9,8	
4	2,7 / 9,14	2,7 / 9,14	
5	11,W22 / 1,8	11,W22 / 2,8	
6	-,W10 / 1,8	-,W10 / 1,8	
7	3,7 / 3,W10	1,17 / 3,W10	
8	9,5 / 2,W15	9,5 / 2,W15	
9	2,W10 / 10,8	2,W10 / 9,W5	
10	2,W15 / 11,W18	2,W15 / 11,W18	
11	W28,W22 / 9,W18	W28,W22 / 9,W18	
12	2,W15/2,W10	2,W15/2,W10	
13	2,8/11,7	1,8/9,14	
14	-,12/-,W18	-,12/-,W18	
15*	1,8/2,27	1,8/2,27	
16	2,W15/1,8	2,W15/1,8	
16	W29,12/1,8	2,W15/1,8	
16	W29,12/1,8	2,W15/1,8	
17	1,8/W29,8	1,8/W29,8	
17	1,8/W29,8	2,12/W29,8	
17	1,8/W29,8	2,12/W29,8	
17	2,12/W29,8	2,12/W29,8	
17	2,12/W29,8	1,8/W29,8	
17	1,8/W29,8	2,12/W29,8	

Tabelle 1: Daten von Tabelle 11, Cudworth und Woodrow (1975). *) In dieser Familie war der Elternteil, welcher den Haplotyp 1,8 verebte hatte, verstorben, so dass IBD und IBS nicht zu unterscheiden waren. Daher ist hier 1 gemeinsames Allel zu zählen. Familie 16 hatte drei betroffene, Familie 17 vier betroffene Kinder.

Insgesamt sind 24 Geschwisterpaare in der Tabelle eingetragen. Was ist die erwartete Verteilung der Paare mit 0, 1 und 2 gemeinsamen IBD-Haplotypen unter der Nullhypothese?

Erwartete Zahlen von Geschwisterpaaren			
IBD-Allele	0	1	2
Beobachtete Zahlen von Geschwisterpaaren			
IBD-Allele	0	1	2

Der χ^2 -Wert lässt sich nun berechnen mit der Formel (O_i beobachtet, E_i erwartet):

$$\chi^2 = \sum_{i=0}^2 \frac{(O_i - E_i)^2}{E_i}$$

p -Werte für eine χ^2 -Verteilung mit zwei Freiheitsgraden finden sich in der folgenden Tabelle:

	p -Wert			
	.05	.025	.01	.001
χ^2	5.99	7.38	9.21	13.82

Fragen

Liegt eine signifikante Kopplung an den HL-A-Lokus nach der ASP-Methode vor?

Was könnten die Gründe sein, dass der HL-A-Haplotyp nicht bei jedem erkrankten Geschwisterpaar übereinstimmt?

Warum setzt die ASP-Methode eine Kenntnis des genetischen Modells bzw. des Vererbungsmodus nicht voraus? Vergleichen Sie sie mit der lod-Wert-Analyse.

Lösung

- Geschwisterpaare mit 2 IBD-Allelen: 13
- Geschwisterpaare mit 1 IBD-Allel : 10
- Geschwisterpaare mit 0 IBD-Allelen: 1

Insgesamt 24 Geschwisterpaare, daher die Erwartung 6–12–6 für 0–1–2 geteilte IBD-Allelen. Der χ^2 -Werte kann berechnet werden als

$$\begin{aligned}
 \chi^2 &= \sum_{i=0}^2 \frac{(O_i - E_i)^2}{E_i} \\
 &= \frac{(1-6)^2}{6} + \frac{(10-12)^2}{12} + \frac{(13-6)^2}{6} \\
 &= \frac{25}{6} + \frac{4}{12} + \frac{49}{6} \\
 &= 12.67
 \end{aligned}$$

dies entspricht einem p -Wert < 0.01 . Daher liegt eine signifikante Kopplung vor.

Doppelverdau-Problem

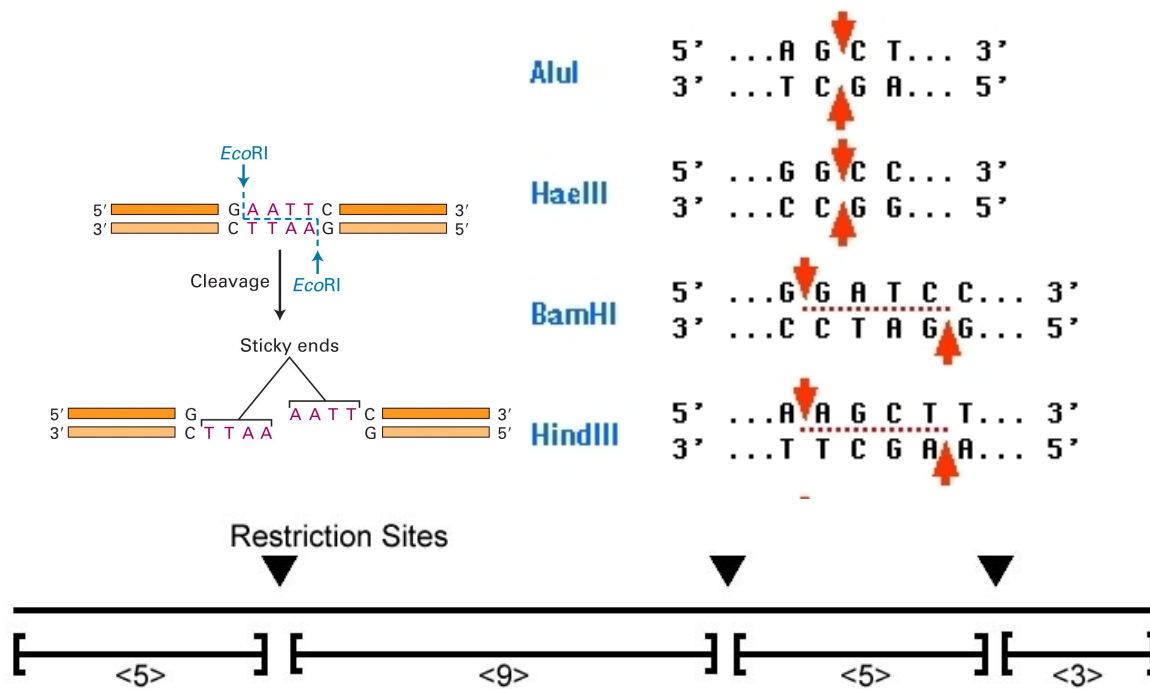


Abbildung 3: Restriktionsverdau. Links EcoRI. Rechts verschiedene Enzyme. Unten: Nach Verdau durch ein Restriktionsenzyme können je nach DNA-Sequenz mehrere Fragmente entstehen.

Das Doppelverdau-Problem:

- Dasselbe DNA-Segment mit zwei unterschiedlichen Restriktionsenzym verdauen (allein und in Kombination), d.h. man erhält drei Gruppen von Verdaufragmenten
- Gesucht wird eine Liste von Restriktionsstellen (Positionen, wo geschnitten wird) für beide Enzyme³

Aufgabe: Lösen Sie das Doppelverdauprobem für folgende Fragmentlängen:

- Restriktion durch Enzym A: 3, 10
- Restriktion durch Enzym B: 2, 5, 6
- Restriktion durch beide Enzyme: 1,2,4,6

Bemerkung: Die Länge der ursprünglichen DNA: 13

Das folgende Schnittmuster (ggf. mit der umgekehrten Reihenfolge) ist eine gültige Lösung:

```

2  1    4    6
--B-A----B-----

```

Das heißt, A schneidet nach Nukleotid 3 und B schneidet nach Nukleotid 2 und 7.

³Wir nehmen der Einfachheit halber an, dass es jeweils nur ein Fragment mit derselben Länge gibt.

2 Genomaufbau

Beschreiben Sie die Rolle der reversen Transkriptase und der Integrase im Lebenszyklus der LINEs Die LINE-mRNA wird mittels Reverser Transkriptase in DNA umgeschrieben und mittels Integrase in das Genom der Wirtszelle eingefügt.

Wie könnten Alu-Sequenzen die Entstehung einer Translokation zwischen zwei Chromosomen begünstigen? Nicht allelische homologie Rekombination

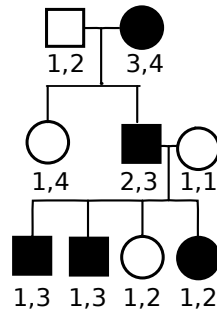
Definieren Sie ein Pseudogen und beschreiben Sie den Unterschied zwischen prozessierten und nichtprozessierten Pseudogenen Pseudogene sind schadhafte Kopien von Genen (verkürzte oder mutierte Fragmente). **Nichtprozessierte Pseudogene:** Duplikation eines Gens mit Exons, Introns, Promoter. **Prozessierte Pseudogene:** Schadhafte Kopien von Genen, die die Exonsequenzen eines funktionellen Gens enthalten sowie am Ende eine Oligo(dA)-Sequenz. Solche prozessierte Pseudogene sind auf der Ebene der cDNA durch Retrotransposition kopiert worden (wahrscheinlich unter Beteiligung des LINE-1-Transpositionssystems).

Definieren Sie die Begriffe ortholog, homolog und paralog **Homologie** : Beschreibt eine Beziehung zwischen zwei Sequenzen, die einen gemeinsamen evolutionären Ursprung haben. Homologie umfasst Orthologie und Paralogie. **Ortholog** sind Gene, die durch Speziation aus einem einzigen Gen vom gemeinsamen Vorfahren entstanden sind. **Paraloge** sind Gene einer bestimmten Spezies, die durch Duplikation entstanden sind.

3 lod-Score

Definieren Sie den Begriff Haplotyp Ein Block von physikalisch gekoppelten Genloci, die zusammenvererbt werden, z.B. mehrfache Genloci auf in einem Bereich auf einem Chromosom.

Warum übersteigt die Rekombinationshäufigkeit nie 50%? Die Rekombinationshäufigkeit zwischen zwei Genloci ist der Anteil an Nachkommen, bei denen eine Rekombination zwischen den beiden Loci während der Meiose stattgefunden hat. Die Rekombinationshäufigkeit kann 50% nicht übersteigen, weil die zufällige Verteilung der Gene eine Rekombinationshäufigkeit von 50% erzeugt. Zum Beispiel, hat ein Elternteil selber AB und ab geerbt und sind diese Loci nicht gekoppelt, dann treten die Kombinationen AB, ab, Ab und aB jeweils mit gleicher Wahrscheinlichkeit in den Gameten auf. Die Hälfte ist rekombinant. Wären die Loci miteinander gekoppelt, dann wäre die Rekombinationshäufigkeit noch geringer.



Bestimmen Sie die Phase, zählen Sie die Rekombinationen und berechnen Sie den lod-Score für folgende Familie für eine Rekombinationshäufigkeit von $\theta = 0.2$ Phase: Der Phänotyp wird von der ersten zur zweiten Generation zusammen mit dem Allel 3 vererbt. Wir können vier Meiosen von der zweiten zur dritten Generation beobachten. In 2 von 3 Fällen wird das Allel 3 mit dem Phänotyp vererbt und in 1 von 1 Fall wird das andere Allel (2) an ein gesundes Kind vererbt. Daher liegen 3 nicht rekombinante und eine rekombinante Meiose vor. Die lod-Formel ist daher:

$$\begin{aligned} lod &= \log_{10} \frac{(1-\theta)^{4-1} \theta^1}{(1/2)^4} \\ &= \log_{10} \frac{0.8^3 \times 0.2}{(1/2)^4} \end{aligned}$$

Es war nicht notwendig, die genaue numerische Antwort zu berechnen.

4 RNA

Beschreiben Sie die Unterschiede zwischen DNA und RNA. Nennen Sie und beschreiben Sie die Funktionen von mindestens 4 Arten von RNA

RNA-Sekundärstruktur

- G–C: drei Wasserstoffbrücken
- A–U: zwei Wasserstoffbrücken
- G–U: Eine Wasserstoffbrücke ("wobble pair" → Wackelpaar)

Bestimmen Sie, unter Beachtung folgender Bedingungen:

- Bilden Nukleotide i und j eine Wasserstoffbrücke, dann $|i - j| > 1$
- Sind (i, j) und (p, q) zwei Wasserstoffbrücken (Paare von Nukleotiden), wobei $i < p < j$, dann gilt $q < j$

die gültigen Strukturen für die RNA-Sequenz AACUGAUAGC. Welche Struktur ist die energetisch günstigste?

S. bitte die Vorlesungsfolien für einige Beispiele. Eine richtige Antwort wäre 2 oder mehr gültige Strukturen sowie die Berechnung der Anzahl der Wasserstoffbrücken. Die Struktur mit den meisten Wasserstoffbrücken wäre dann die energetisch günstigste.

Beschreiben Sie die Funktionen von Drosha, Dicer und den Argonaut-Proteinen pri-miRNAs werden durch die Endonuclease Drosha verarbeitet. Das Ergebnis sind ~ 70 nt Haarnadeln namens prä-miRNAs, prä-miRNAs werden durch Exportin-5 ins Zytoplasma transportiert. Nach der Verabreichung durch Dicer werden die miRNAs in einen Ribonukleoproteinkomplex namens miRNA-induzierter Silencing-Komplex (miRISCs) eingetaucht. Die wichtigsten Proteinbestandteile des miRISC sind Proteine der Argonaut-Familie.

5 Mutationen

Beschreiben Sie die Folge der Mutation c.723_724insT Insertion eines T zwischen den Nukleotiden 723 und 724.

Beschreiben Sie die Folge der Mutation c.921delA Deletion des A an der Position 921 nach dem ersten Nukleotid des Startkodons.

Welche der folgenden ist eine Nonsense-Mutation? ein Missense-Mutation? p.G107C, pG107X p.G107C: Missense pG107X: Nonsense

Beschreiben Sie den Unterschied zwischen stillen Mutationen, konservativen und nichtkonservativen Aminosäuresubstitutionen **Stille Mutation:** Veränderung der Nukleotidsequenz ohne Veränderung der Aminosäuresequenz. **konservative Aminosäuresubstitution:** Mutante Aminosäure ist nach physikochemischen Gesichtspunkten ähnlich wie die Wildtypamino­säure, z.B. Leuzin und Isoleuzin. **Nichtkonservative Aminosäuresubstitution:** Mutante Aminosäure ist nicht ähnlich.