

Übung "Statistische Aspekte der Analyse
molekularbiologischer und genetischer Daten"

Inhaltsverzeichnis

1	Übung 1: Biologische Grundlagen – Teil 1	1
1.1	Aufgabe 1	1
1.2	Aufgabe 2	1
1.3	Aufgabe 3	1
1.4	Aufgabe 4	1
1.5	Aufgabe 5	2
1.6	Aufgabe 6	3
2	Übung 2	4
2.1	Aufgabe 1	4
2.2	Aufgabe 2	4
2.3	Aufgabe 3	5
2.4	Aufgabe 4	5
3	Übung 3	7
3.1	Aufgabe 1	7
3.2	Aufgabe 2	8
3.3	Aufgabe 3	8

1 Übung 1: Biologische Grundlagen – Teil 1

1.1 Aufgabe 1

- zu a: siehe Codonsonne¹
AUG (ATG) als Startcodon, UGA (TGA) als Stopcodon
5' - ATG GTT AAA CAC GTG CAC GAG TGA - 3'
3' - TAC CAA TTT GTG CAC GTG CTC ACT - 5'
- zu b:
5' - AUG GUU AAA CAC GUG CAC GAG UGA - 3'
- zu c: tRNA für Valin, Lysin, Histidin, Valin, Glutamin, Glutaminsäure (das komplementäre der RNA)
- zu d: unpolar/neutral, positiv/basisch, positiv/basisch, unpolar/neutral, polar/neutral, negativ/sauer

1.2 Aufgabe 2

1.3 Aufgabe 3

- E. coli: $4,6 \cdot 10^6$ Basen, 4500 Gene
- Bäckerhefe: $2 \cdot 10^7$ Basen, 6000 Gene
- Ackerschmalwand: 10^8 Basen, 25500 Gene
- Fruchtfliege (Drosophila Melanogaster): $2 \cdot 10^8$ Basen, 13500 Gene
- Menschen: $3,27 \cdot 10^9$ Basen, 23000 Gene

1.4 Aufgabe 4

- SNP²:
 - Single Nucleotide Polymorphism - Einzelnukleotid-Polymorphismus
 - Variation eines einzelnen Basenpaares in einem DNA-Strang
 - SNPs sind geerbte und vererbte genetische Varianten. Begrifflich davon abzugrenzen ist der Begriff der Mutation, der in der Regel eine neu aufgetretene Veränderung bezeichnet
 - Laktosetoleranz: durch einen SNP im Intron des Gens mcm6 entwickelt, welches 5' von LCT(Lactase) liegt
- CNV³:
 - Copy number variation - Kopienzahlvariation
 - struktureller Variation des Erbguts, die Abweichungen der Anzahl der Kopien eines bestimmten DNA-Abschnittes innerhalb eines Genoms erzeugt
- Chromosomen-Mutationen⁴:
 - strukturelle Veränderung eines Chromosoms, 5 Arten

¹<https://de.wikipedia.org/wiki/Code-Sonne>

²<https://de.wikipedia.org/wiki/Einzelnukleotid-Polymorphismus>

³https://de.wikipedia.org/wiki/Gene_copy_number_variants

⁴<https://de.wikipedia.org/wiki/Chromosomenmutation>

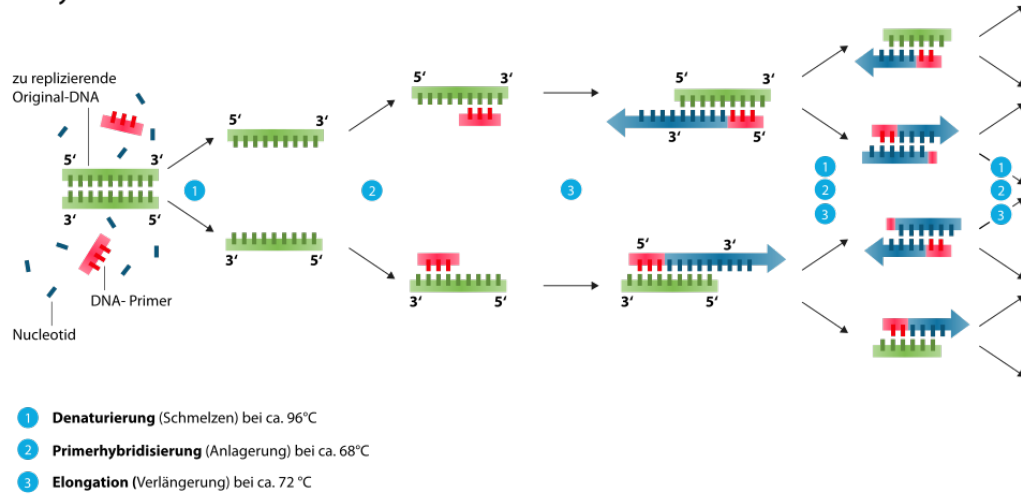
- Deletion: Ein Teilstück des Chromosoms (Endstück oder mittlerer Abschnitt) geht verloren
- Translokation: Chromosomen können auseinanderbrechen und dabei Teilstücke verlieren, welche in die Chromatide eines anderen Chromosoms angeheftet werden
- Duplikation: Ein Abschnitt des Chromosoms ist doppelt vorhanden, da ein auseinandergebrochenes Teilstück in die Schwesterchromatide eingegliedert wurde
- Inversion: Innerhalb eines Chromosoms kann sich nach einem doppelten Bruch ein Stück wieder umgekehrt einfügen
- Insertion (auch: Addition): Hier besitzt ein Chromosom ein zusätzliches Teilstück

1.5 Aufgabe 5

- PCR⁵: Polymerase-Kettenreaktion (polymerase chain reaction)
- Prozess besteht aus etwa 20–50 Zyklen, jeder Zyklus besteht aus drei Schritten
 1. Denaturierung (Melting, Schmelzen): Zunächst wird die doppelsträngige DNA auf 94–96 °C erhitzt, um die Stränge zu trennen. Die Wasserstoffbrückenbindungen, die die beiden DNA-Stränge zusammenhalten, werden aufgebrochen. Im ersten Zyklus wird die DNA oft für längere Zeit erhitzt (Initialisierung), um sicherzustellen, dass sich sowohl die Ausgangs-DNA als auch die Primer vollständig voneinander getrennt haben und nur noch Einzelstränge vorliegen. Manche (sogenannte Hot-Start-) Polymerasen müssen durch eine noch längere anfängliche Erhitzungsphase (bis zu 15 Minuten) aktiviert werden. Danach wird schnell auf 65 °C abgekühlt, um die Rückbildung der Doppelhelix zu verhindern.
 2. Primerhybridisierung (primer annealing): Die Temperatur wird ca. 30 Sekunden lang auf einem Wert gehalten, der eine spezifische Anlagerung der Primer an die DNA erlaubt. Die genaue Temperatur wird hierbei durch die Länge und die Sequenz der Primer bestimmt (bzw. der passenden Nukleotide im Primer, wenn durch diesen Mutationen eingeführt werden sollen = site-directed mutagenesis). Wird die Temperatur zu niedrig gewählt, können sich die Primer unter Umständen auch an nicht hundertprozentig komplementären Sequenzen anlagern und so zu unspezifischen Produkten („Geisterbanden“) führen. Wird die Temperatur zu hoch gewählt, ist die thermische Bewegung der Primer u. U. so groß, dass sie sich nicht richtig anheften können, so dass es zu gar keiner oder nur ineffizienter Produktbildung kommt. Die Temperatur, welche die beiden oben genannten Effekte weitgehend ausschließt, liegt normalerweise 5–10 °C unter dem Schmelzpunkt der Primersequenzen; dies entspricht meist einer Temperatur von 55 bis 65 °C.
 3. Elongation (Extending, Polymerisation, Verlängerung, Amplifikation): Schließlich füllt die DNA-Polymerase die fehlenden Stränge mit freien Nukleotiden auf. Sie beginnt am 3'-Ende des angelagerten Primers und folgt dann dem DNA-Strang. Der Primer wird nicht wieder abgelöst, er bildet den Anfang des neuen Einzelstrangs. Die Temperatur hängt vom Arbeitsoptimum der verwendeten DNA-Polymerase ab (68–72 °C). Dieser Schritt dauert etwa 30 Sekunden je 500 Basenpaare, variiert aber in Abhängigkeit von der verwendeten DNA-Polymerase. Übliche Thermocycler kühlen die Reaktionsansätze nach Vollendung aller Zyklen auf 4–8 °C, so dass eine PCR am Abend angesetzt werden kann und die Proben am Morgen darauf weiterverarbeitet werden können.

⁵<https://de.wikipedia.org/wiki/Polymerase-Kettenreaktion>

Polymerasekettenreaktion - PCR



zu amplifizierende Sequenz:

5'ACCGCGGCTT AGGAAAXXXX XXXXXCCCG GGGCGTATGC TGACGG3'
 3'-CGAA TCCTTT-5' 3'-GGGC CCCGCA-5'

1.6 Aufgabe 6

Didesoxymethode nach Sanger⁶:

- Didesoxynukleotide weil: wird als Stopp-Nukleotiden benutzt, an Ribose (Zucker) an Position 2' und 3' desoxidiert ist. Dadurch fehlt am 3'-Kohlenstoff-Atom die Hydroxygruppe, an der bei der Polymerisation das nächste Nukleotid angehängt wird.
- auch Desoxynukleotide weil: sonst funktioniert die Verlängerung nicht
- Ergebnis nur Didesoxynukleotide: es gibt keine Verlängerung

nur Didesoxynukleotide

⁶https://de.wikipedia.org/wiki/DNA-Sequenzierung#Didesoxymethode_nach_Sanger

2 Übung 2

2.1 Aufgabe 1

a.)

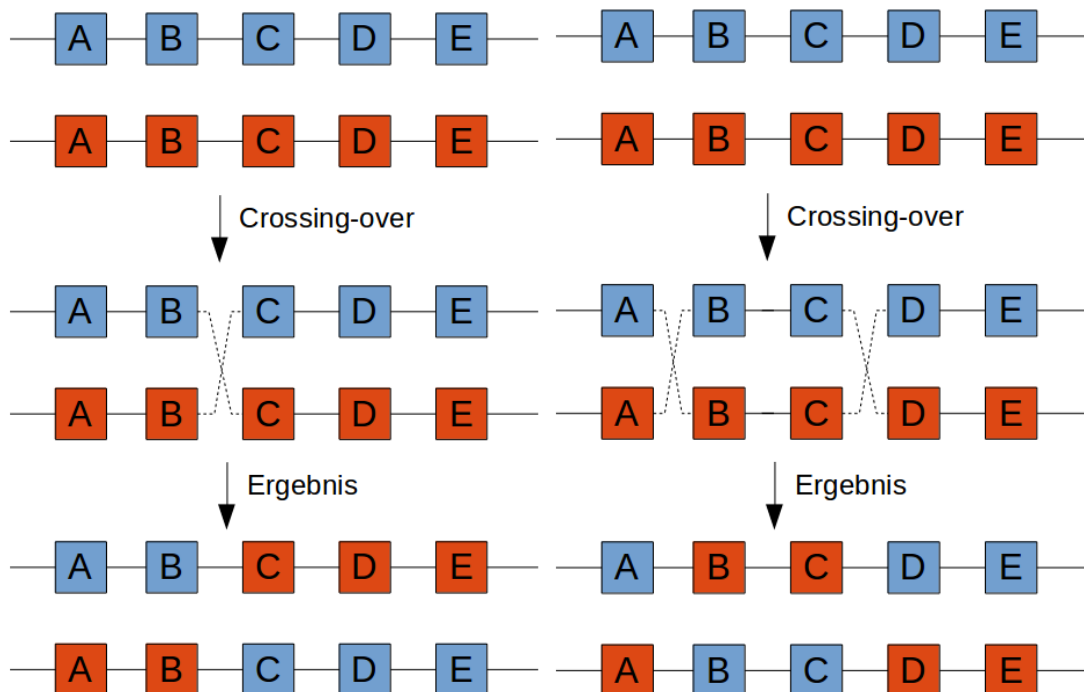
Als Crossing-over⁷ wird in der Genetik eine kreuzweise Überlagerung zweier Chromatiden mit nachfolgendem, gegenseitigem Austausch von Abschnitten bezeichnet, wie er zwischen väterlichen und mütterlichen homologen Chromosomen bei einer Meiose auftreten kann.

b.) A und B sind rekombiniert zu C,D,E

c.) A, D,E sind rekombiniert mit B,C

zu b.)

zu c.)



2.2 Aufgabe 2

Gen: ABO⁸ rs8176719⁹:

- (-;-): likely to be of blood type O
- (-;G): most likely to be of blood type A or B
- (G;G): most likely to be of blood type A, B or AB

rs8176747¹⁰:

- G führt zu Blutgruppe A, C zu Blutgruppe B

rs8176750¹¹: definiert Untergruppe von A

⁷<https://de.wikipedia.org/wiki/Crossing-over>

⁸<http://www.snpedia.com/index.php/ABO>

⁹<http://www.snpedia.com/index.php/rs8176747>

¹⁰<http://www.snpedia.com/index.php/rs8176747>

¹¹<http://www.snpedia.com/index.php/rs8176750>

- (-;C): A1
- (-;-): A2

Kombinationsmöglichkeiten:

- praktisch durch Allele vorgegeben: $3 \cdot 2 \cdot 2 = 12^{12}$
- theoretisch: $5^3 = 125$
- Musterlösung: 3 SNPs auf einem Allel \rightarrow 8 Kombinationen; 2 Allele: 36 Möglichkeiten

A und B kodominant, Faktor 0 rezessiv

2.3 Aufgabe 3

- a.)
- b.)
- c.)

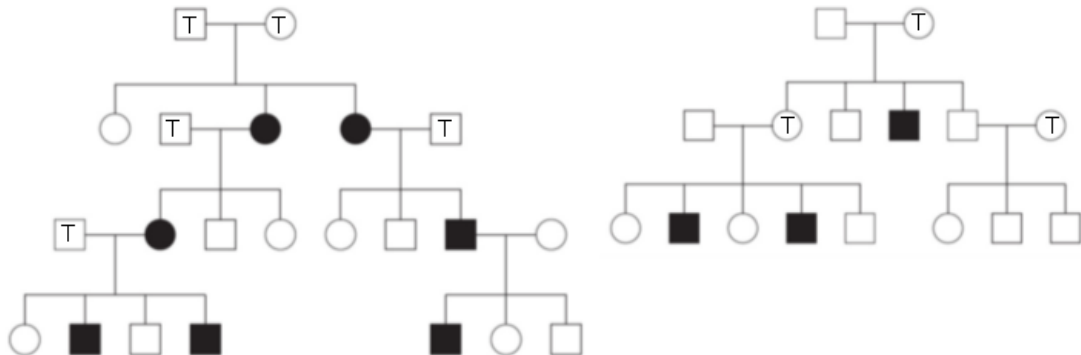
2.4 Aufgabe 4

- a.)

rezessiv:¹³ bedeutet in der Genetik „zurücktretend“ oder auch „nicht in Erscheinung tretend“
dominant:¹⁴ ein dominantes Allel setzt sich in der Merkmalsausprägung gegenüber einem rezessiven Allel durch

Penetranz:¹⁵ prozentuale Wahrscheinlichkeit, mit der ein bestimmter Genotyp zur Ausbildung des zugehörigen Phänotyps führt

- b.)



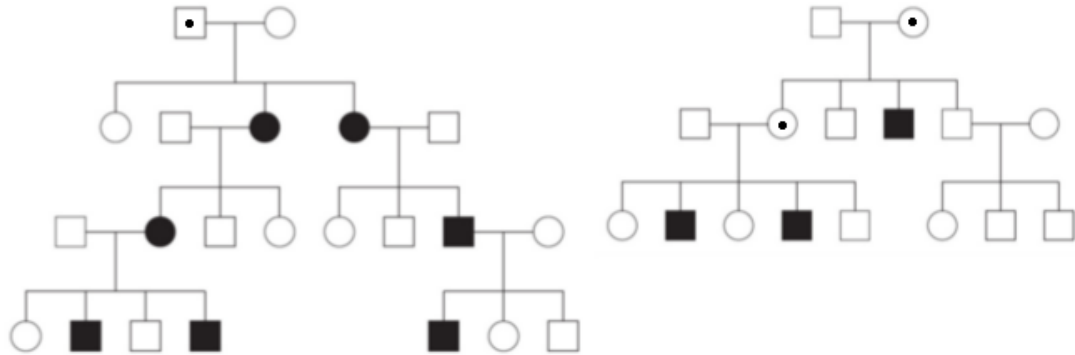
¹²<https://sites.google.com/site/abobloodgroup/14.aboalleles%28oalleles%29>

¹³<https://de.wikipedia.org/wiki/Rezessiv>

¹⁴[https://de.wikipedia.org/wiki/Dominanz_\(Genetik\)](https://de.wikipedia.org/wiki/Dominanz_(Genetik))

¹⁵[https://de.wikipedia.org/wiki/Penetranz_\(Genetik\)](https://de.wikipedia.org/wiki/Penetranz_(Genetik))

aus Musterlösung:



c.)

links: autosomal rezessiv, aus Musterlösung: autosomal dominant mit reduzierter Penetranz, weil:

- beide Geschlechter betroffen
- in jeder Generation
- etwa die Hälfte der Kinder betroffen

rechts: genosomal rezessiv, auf einem X-Chromosom der Mutter

3 Übung 3

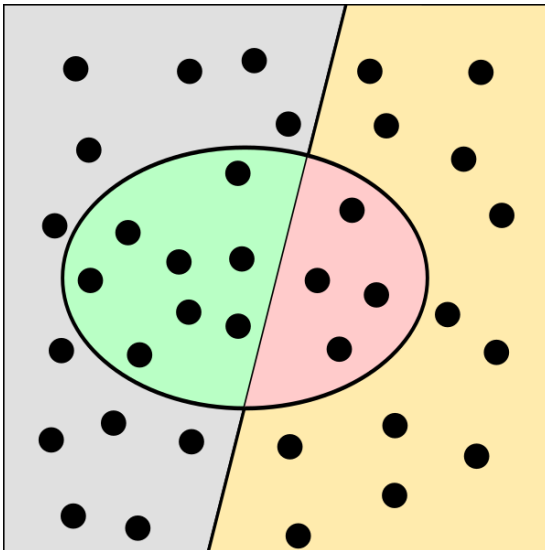
3.1 Aufgabe 1

a.)

- Sensitivität: gibt den Anteil der korrekt als positiv klassifizierten Objekte an der Gesamtheit der tatsächlich positiven Objekte an
- Spezifität: gibt den Anteil der korrekt als negativ klassifizierten Objekte an der Gesamtheit der in Wirklichkeit negativen Objekte an
- Prävalenz: welcher Anteil der Menschen einer bestimmten Gruppe (Population) definierter Größe zu einem bestimmten Zeitpunkt an einer bestimmten Krankheit erkrankt ist
 $\text{Prävalenz} = \text{Anzahl der zum Untersuchungszeitpunkt Kranken} / \text{Anzahl der in die Untersuchung einbezogenen Individuen}$

Vierfeldertafel

	Person ist krank ($r_p + f_p$)	Person ist gesund ($f_p + r_n$)
Test positiv ($r_p + f_p$)	richtig positiv (r_p)	falsch positiv (f_p)
Test negativ ($f_n + r_n$)	falsch negativ (f_n)	richtig negativ (r_n)



b.)

gegeben:

- $K = \{\text{Patient ist krank}\}$
- $P = \{\text{Test ist positiv}\}$
- Sensitivität: $\mathbb{P}(P|K) = 0,95$
- Spezifität: $\mathbb{P}(\bar{P}|\bar{K}) = 0,90$
- Prävalenz: $\mathbb{P}(K) = 0,1$

gesucht:

- positiv prädiktiver Wert (PPW):

$$\mathbb{P}(K|P) = \frac{\mathbb{P}(P|K) \cdot \mathbb{P}(K)}{\underbrace{\mathbb{P}(P)}_{\text{Satz von Bayes}}} = \frac{\mathbb{P}(P|K) \cdot \mathbb{P}(K)}{\underbrace{\mathbb{P}(P|\bar{K}) \cdot \mathbb{P}(\bar{K}) + \mathbb{P}(P|K) \cdot \mathbb{P}(K)}_{=1 - \mathbb{P}(\bar{P}|\bar{K})}}$$

$$\mathbb{P}(K|P) = \frac{0,95 \cdot 0,1}{0,1 \cdot 0,9 + 0,95 \cdot 0,1} = \underline{\underline{0,513513514}}$$

- negativ prädiktiver Wert (NPW):

$$\mathbb{P}(\bar{K}|\bar{P}) = \frac{\mathbb{P}(\bar{P}|\bar{K}) \cdot \mathbb{P}(\bar{K})}{\underbrace{\mathbb{P}(\bar{P})}_{1 - \mathbb{P}(P)}} = \frac{\mathbb{P}(\bar{P}|\bar{K}) \cdot \mathbb{P}(\bar{K})}{1 - (\mathbb{P}(P|\bar{K}) \cdot \mathbb{P}(\bar{K}) + \mathbb{P}(P|K) \cdot \mathbb{P}(K))}$$

$$\mathbb{P}(\bar{K}|\bar{P}) = \frac{0,9 \cdot 0,9}{1 - (0,1 \cdot 0,9 + 0,95 \cdot 0,1)} = \underline{\underline{0,993865031}}$$

c.)

gegeben:

- Sensitivität: $\mathbb{P}(P|K) = 0,95$
- Spezifität: $\mathbb{P}(\bar{P}|\bar{K}) = 0,90$
- Prävalenz: $\mathbb{P}(K) = 0,05$

gesucht:

- positiv prädiktiver Wert (PPW) = 0,33
- negativ prädiktiver Wert (NPW) = 0,997084548104956

d.) siehe R-Script

3.2 Aufgabe 2

3.3 Aufgabe 3

siehe R-Script