

Funktionelle Genomanalysen 2023 (09-CRT-A006)

Übung 1: Grundlagen der genetischen Statistik

Dr. Janne Pott

09.-11. Juni 2023

Allgemeine Hinweise:

- Die Aufgaben werden in der Übung gemeinsam bearbeitet.
- Zur Lösung von manchen Aufgaben wird ein Taschenrechner o.ä. benötigt.
- Am Ende des Moduls wird eine Musterlösung bereitgestellt.

Aufgabe 1: Crossing-over & Linkage-Disequilibrium

- Definieren Sie anhand der Abbildung 1 den Begriff **Crossing-over**.
- Erläutern Sie den Zusammenhang zwischen der **Crossing-over** und **LD-Struktur** des Genoms.
- Betrachten Sie Tabelle 1. Bestimmen Sie die **Randverteilungen** und berechnen Sie das **LD-Maß** r^2 ! Formel:
$$r^2 = \frac{(p_{00}p_{11} - p_{01}p_{10})^2}{p_{0.}p_{.0}p_{1.}p_{.1}}$$
- Interpretieren Sie das Ergebnis! Was sind die **häufigen Haplotypen**? Was bedeutet dies für ein doppelt heterozygoten Individuum?
- Würden Sie zwischen SNP 1 und SNP 3 ein höheres oder niedrigeres r^2 erwarten? Begründen Sie Ihre Entscheidung!

Table 1: 4-Felder-Tafel der beiden biallelischen SNPs: SNP 1 (Allele A1/B1) und SNP 2 (Allele A2/B2) aus Daten von 500 gemessenen diploiden Individuen

	SNP 1 - Allel A1	SNP 1 - Allel B1
SNP 2 - Allel A2	570	15
SNP 2 - Allel B2	25	390

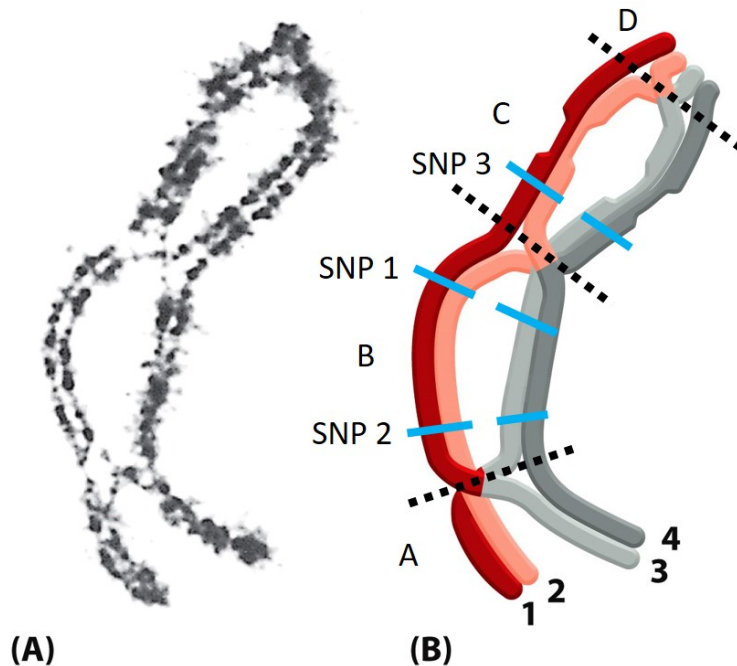


Figure 21-10 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Figure 1: Crossing-over eines Chromosoms. A) Elektronenmikroskopische Aufnahme. B) Schematische Darstellung. Die schwarz gestrichelten Linien kennzeichnen ein Crossing-over, die blauen Linien die grobe Position von SNPs. Insgesamt sind 4 Segmente (A-D) eingetragen. Die etwas dünneren Stellen im Segment C kennzeichnen das Zentromer. Modifiziert aus Alberts et al., Molecular Biology of the Cell, 2008

Aufgabe 2: Hardy-Weinberg-Gleichgewicht

Für den biallelischen SNP 1 mit Allelen A und B wird folgende Genotypverteilung beobachtet:

Genotyp	AA	AB	BB	Missing
Häufigkeit	824	1326	463	87

Table 2: Genotypverteilung eines gemessenen SNPs mit Allelen A/B in n=2700 diploiden Individuen. Missing bedeutet, dass kein Genotyp vom Algorithmus bestimmt werden konnte.

- Welche Modellannahmen werden Hardy-Weinberg-Gleichgewicht (HWE) getroffen (Stichwort **ideale Population**)?
- Betrachten Sie Tabelle 2. Bestimmen Sie auf drei Nachkommastellen genau die
 - die **Callrate** des SNPs,
 - die **Allelfrequenzen** für A und B, und
 - die **erwartete Genotypverteilung** im HWE!
- Zusatz: Testen Sie auf HWE mit 5% Irrtumswahrscheinlichkeit. Stellen Sie dazu die **Nullhypothese** auf. Berechnen Sie die **Teststatistik** für diese und interpretieren Sie das Ergebnis (s. Tabelle 3 für die Quantile). Formel:

$$\sum_i \frac{(O_i - E_i)^2}{E_i}, i \in AA, AB, BB$$

Table 3: Wichtige Quantile der χ^2 -Verteilung nach Freiheitsgraden df und Wahrscheinlichkeit α

	$\alpha=0.99$	$\alpha=0.975$	$\alpha=0.95$	$\alpha=0.05$	$\alpha=0.025$	$\alpha=0.01$
df=1	0.00016	0.00098	0.0039	3.841	5.024	6.635
df=2	0.020	0.051	0.103	5.991	7.378	9.210
df=3	0.115	0.216	0.352	7.815	9.348	11.340
df=4	0.297	0.484	0.711	9.488	11.140	13.280
df=5	0.554	0.831	1.150	11.070	12.830	15.090

Aufgabe 3: Genetische Modelle & Stammbäume

- a) Definieren Sie die Begriffe **dominant**, **rezessiv** und **Penetranz**.
- b) Betrachten Sie die drei Stammbäume in Abbildung 2 und geben Sie folgendes an:
- eine Legende,
 - die Träger/in (soweit möglich),
 - wahrscheinlichstes Segregationsmuster (mit Begründung)

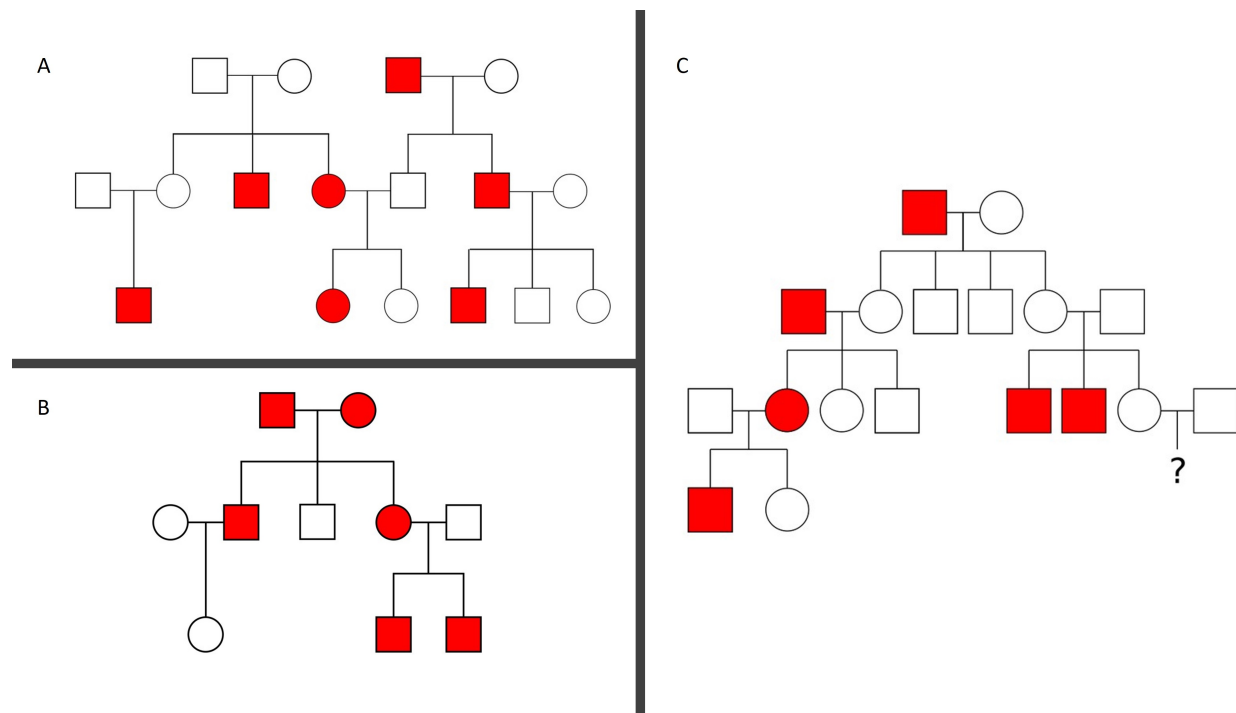


Figure 2: Drei Stammbäume. Frei aus dem Internet