



QTL-Mapping und Genomweite Assoziationsstudien

Birgit Gredler-Grandl

Prüfung 18.12.2015

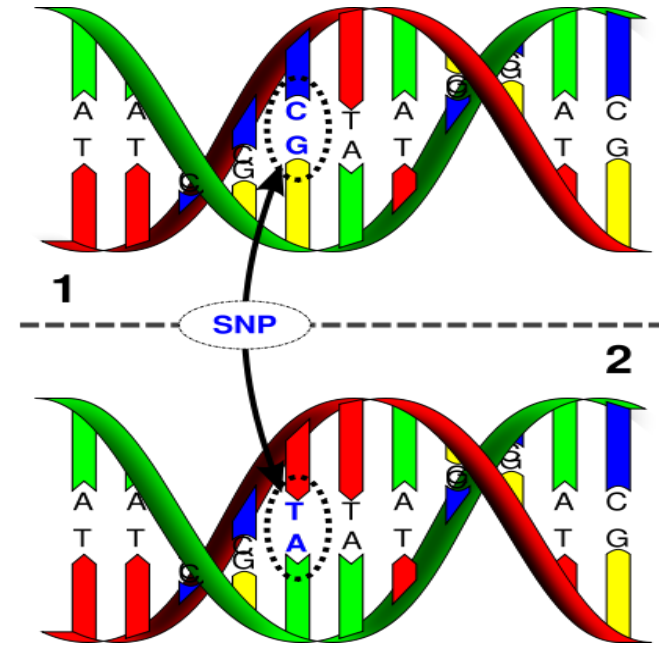
- 90 min, 09:15 – 10:45
- Keine Hilfsmittel
- Nicht programmierbarer Taschenrechner erlaubt
- Legitimationskarte mitbringen!
- Eine gemeinsame Prüfung für Züchtungslehre I und II

Heutige Vorlesung

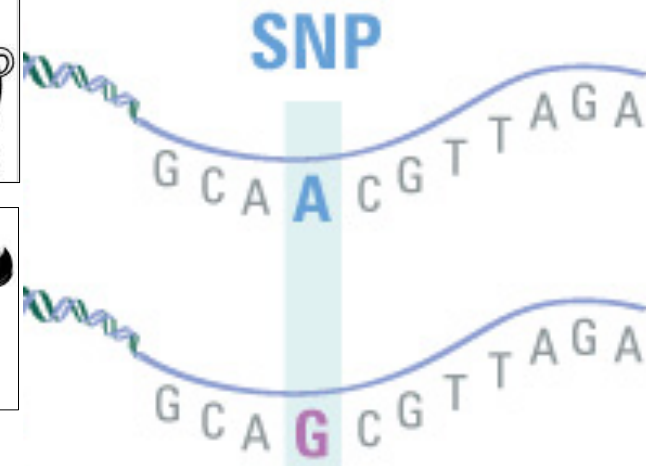
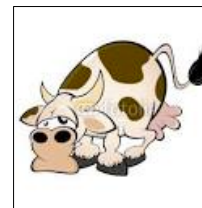
- Frage von der letzten Einheit:
 - Wie funktioniert die SNP-Genotypisierung?
- QTL-Mapping
- Genomweite Assoziationsstudien

SNP

- Punktuelle Veränderungen einer Base in der DNA-Sequenz (Punktmutation)
- Base **Cytosin** wurde durch Base **Thymin** ersetzt
- SNP sind diallel – treten nur in 2 Allel-Varianten auf
- 3 Genotypen: AA – AT – TT

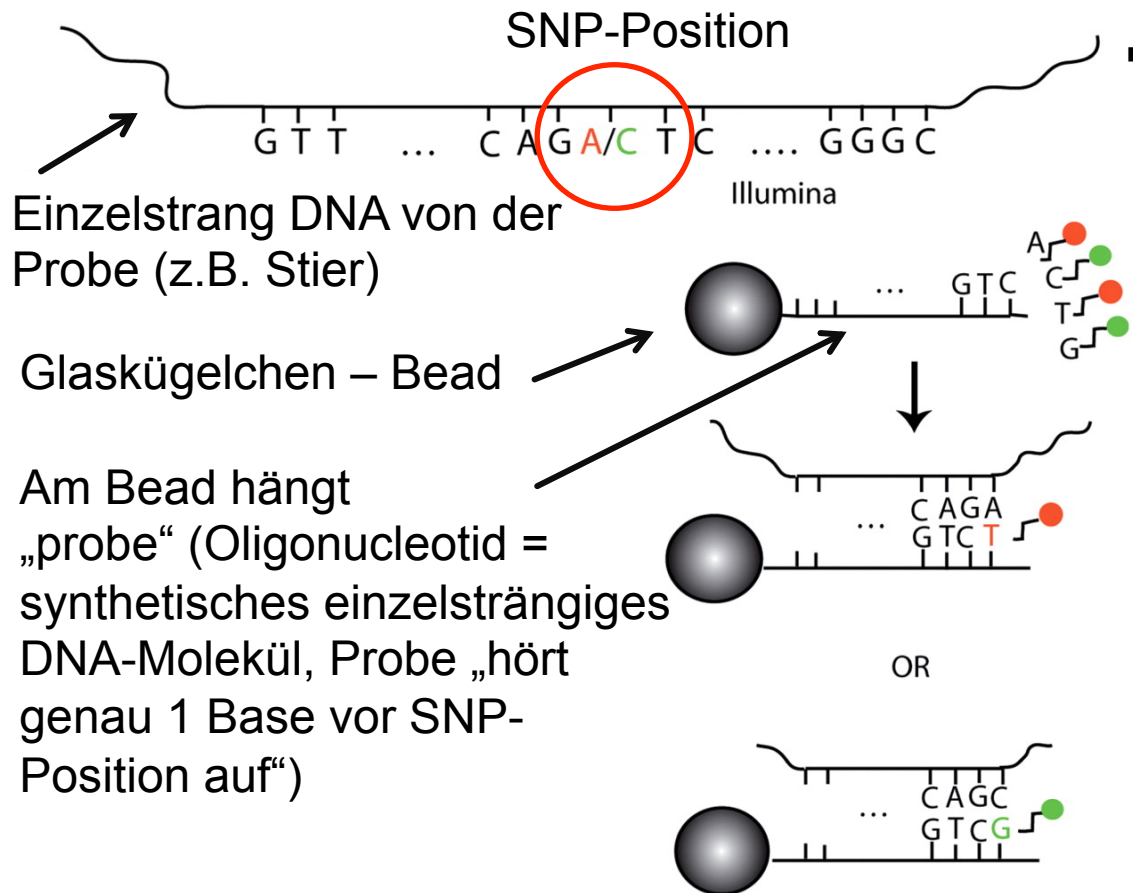


wikipedia.org



Seefried, 2015

SNP-Genotypisierung



- Methode stützt sich auf dem Prinzip, dass DNA-Basen komplementär sind (A-T, C-G)

SNP-Genotypisierung – Ausgangspunkt bildet biologisches Material

Haarprobe entnehmen



50-100 Haarwurzeln



Haare auf Haarkarte kleben



Tier-ID aufkleben



Versand der Probe zu Geneseek (USA) zur DNA-Extraktion und Genotypisierung

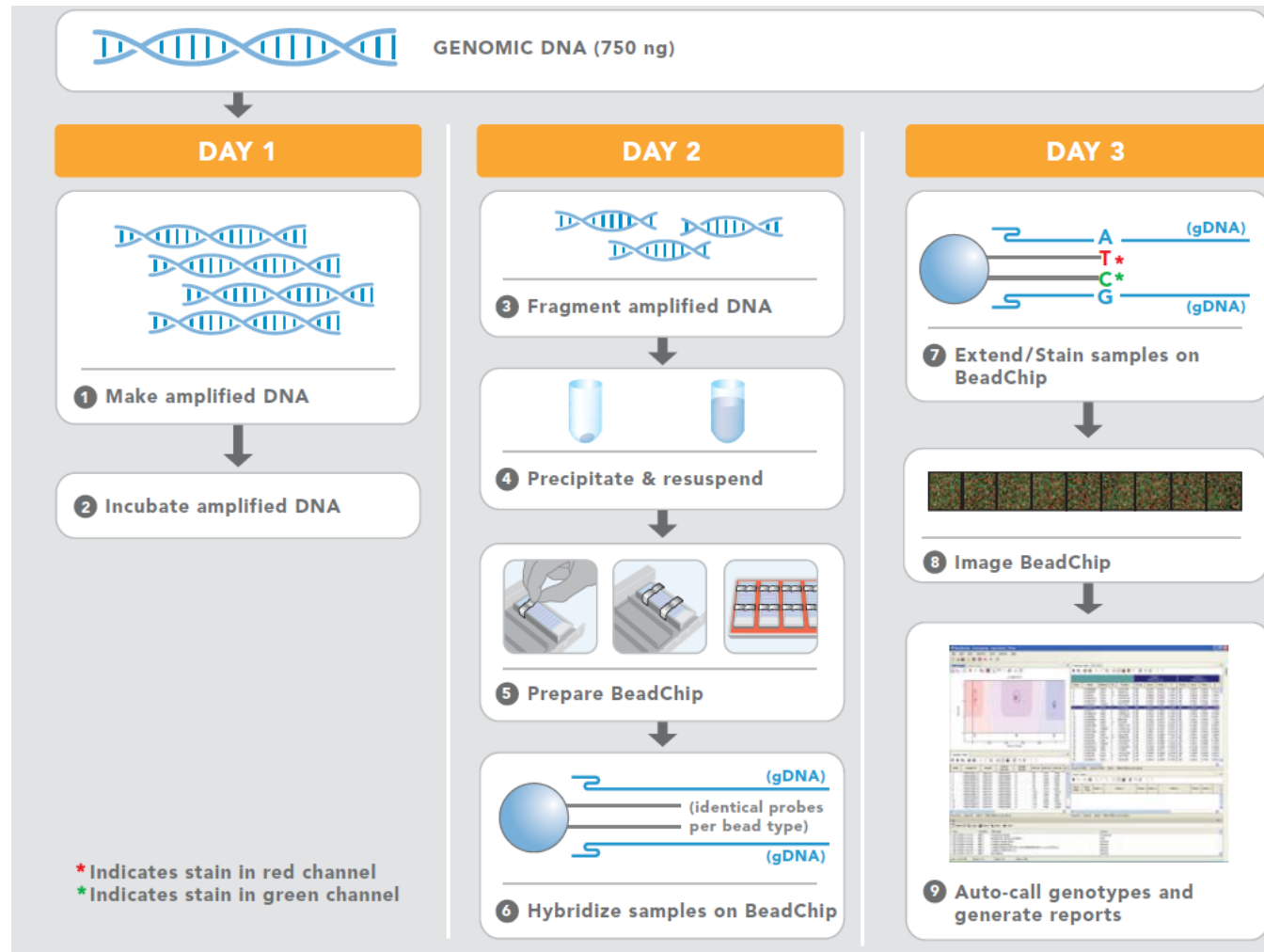


http://homepage.braunvieh.ch/documents/Probenahme_Haarprobe.pdf

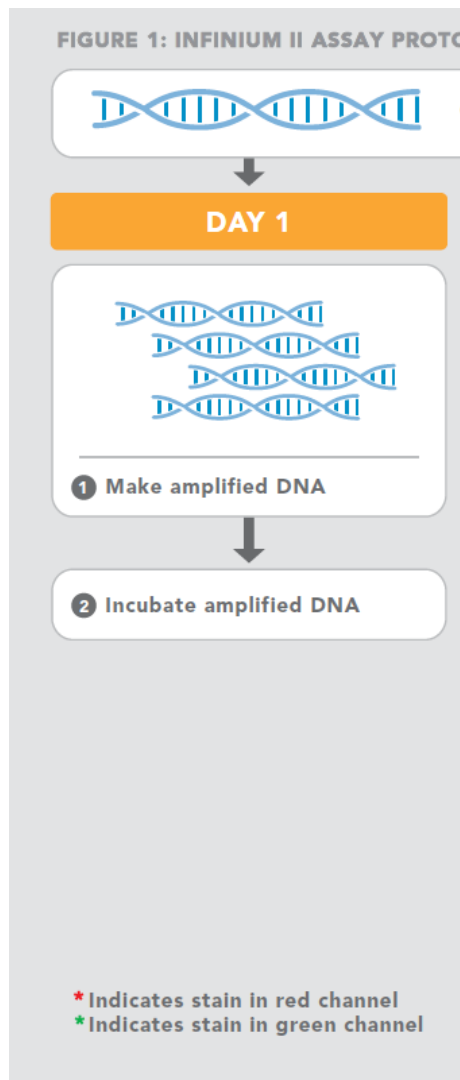
<http://www.neogen.com/Genomics/>

Illumina Typisierung Workflow

http://www.illumina.com/documents/products/workflows/workflow_infinium_ii.pdf

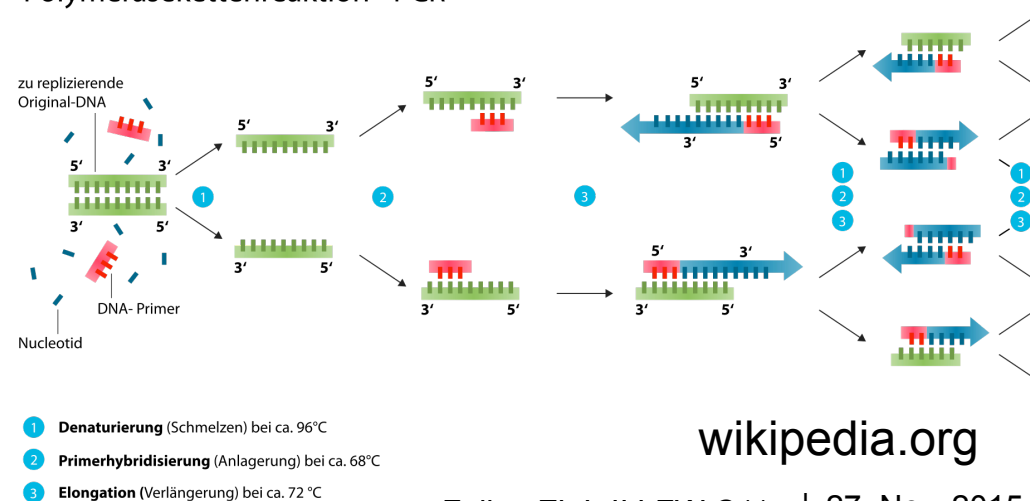


SNP-Genotypisierung

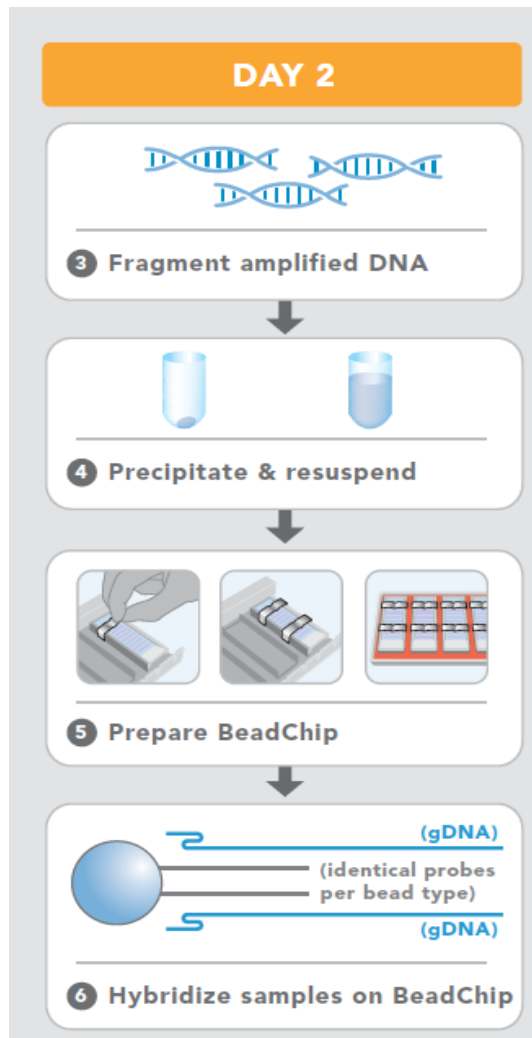


- 750 ng DNA werden benötigt
- 1 Nanogramm (ng) = 1 Milliardstel Gramm
- DNA wird amplifiziert
- DNA-Amplifikation: DNA wird vervielfältigt (kopiert)
- z.B. Polymerasekettenreaktion (PCR)

Polymerasekettenreaktion - PCR

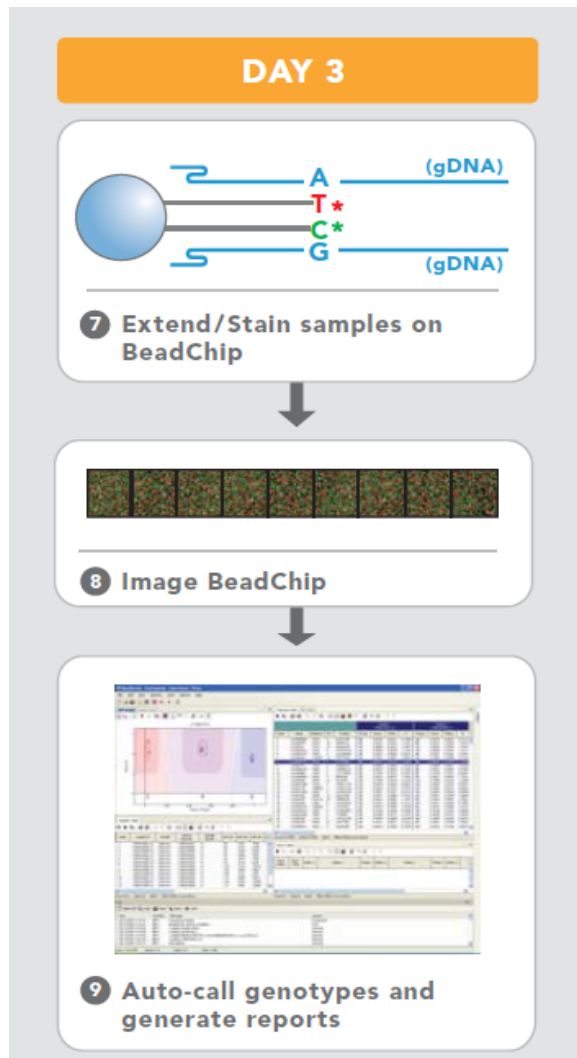


SNP-Genotypisierung



- DNA wird fragmentiert (3)
- Alkoholfällung und Resuspension (4)
- Vorbereitung Bead-Chip (5)
- (6) Beadchip wird mit Proben-DNA versetzt und es kommt zur Hybridisierung (komplementärer Einzelstrang lagert sich an); Es befinden sich sehr, sehr viele Oligonucleotide auf dem Bead
- Für jeden SNP je ein spezifischer Bead mit Oligonucleotide

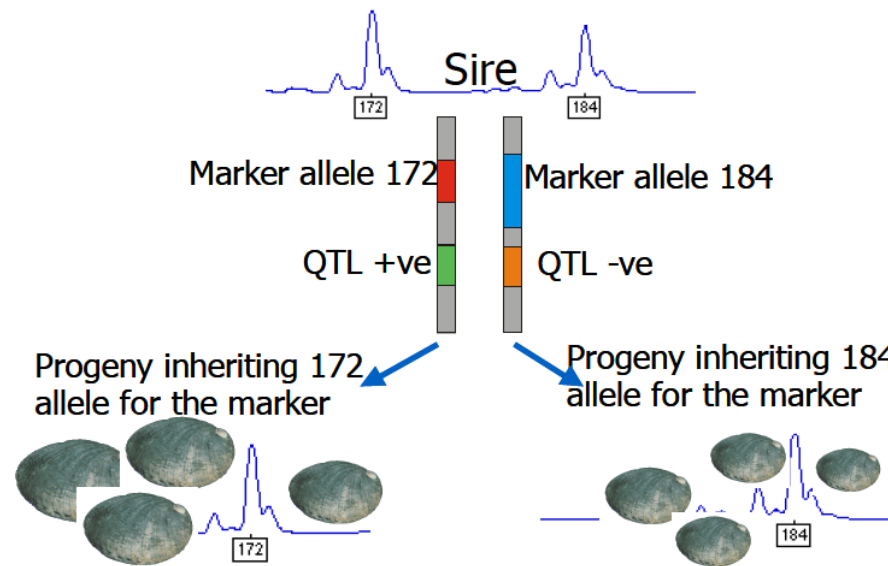
SNP-Genotypisierung



- Zugabe von DNA-Einzelbasen
- Einfärben mit Fluoreszenzmittel
- Unterschiedliche Farbe je nach Base
- Farbintensitäten werden ausgewertet

QTL-mapping

Grundprinzip: Vergleich von Mittelwerten



Versuchsdesigns QTL-mapping

- Mapping von QTLs welche **ZWISCHEN** Populationen segregieren
 - QTLs, welche Unterschiede zwischen Populationen erklären
 - Populationen mit extrem unterschiedlichen Phänotypen
 - z.B. Kreuzungen von Inzuchtlinien
- Mapping von QTLs welche **INNERHALB** von Populationen segregieren
 - QTLs, welche genetische Variation innerhalb Population erklären
 - Zuchtpopulationen (Rind, ...)
 - z.B. Töchterdesign (daughter design), Enkelinnendesign (granddaughter design)

Kreuzungen von Inzuchtlinien

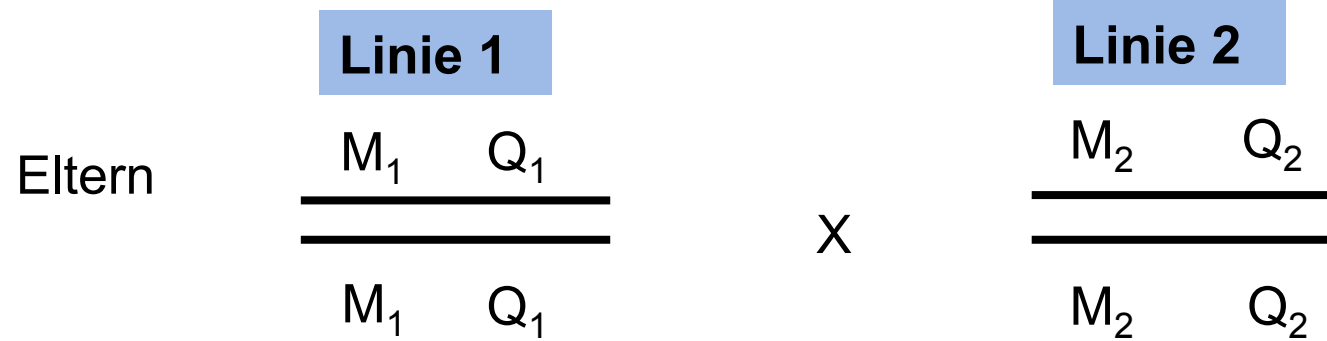
- Pflanzen
- Modellorganismen (z.B. Mäuse, Drosophila)
- Ziel: QTL finden, welche in den Linien unterschiedlich sind
- Sehr oft wird ein **F1 backcross Design** angewandt



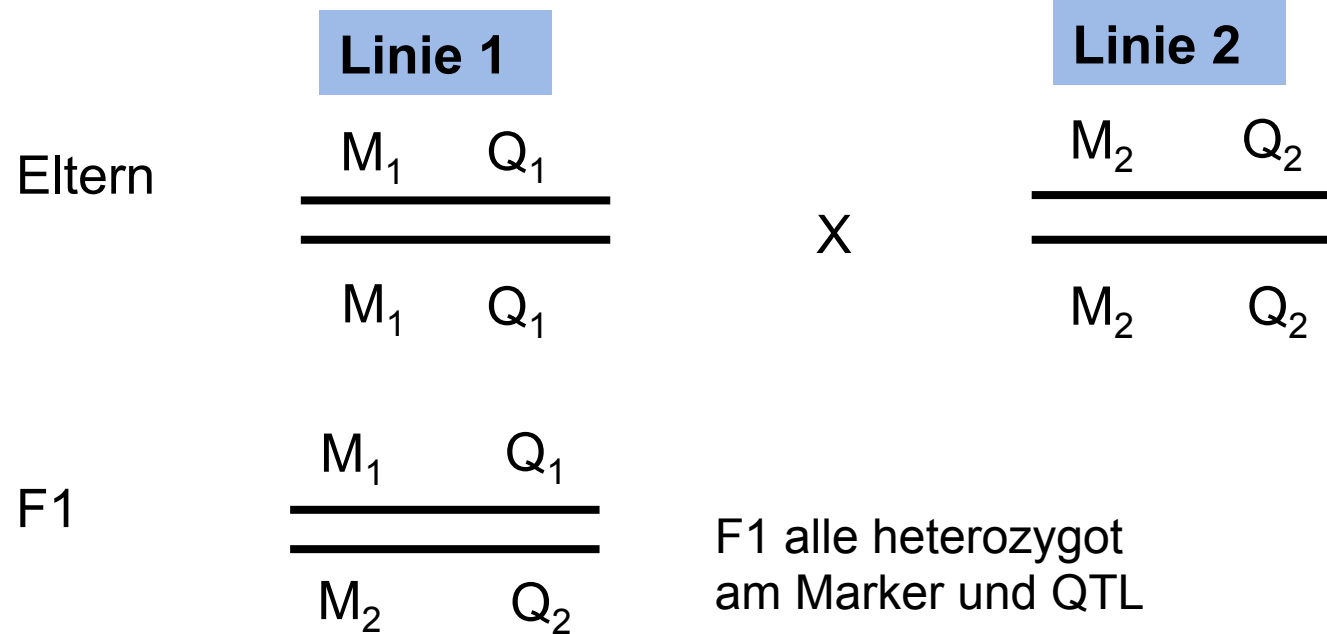
Kreuzungen von Inzuchtlinien

- Inzuchtlinien ...
 - sind genetisch uniform
 - sind fixiert für das alternative QTL-Allel (d.h. QTLs unterscheiden sich zwischen den Inzuchtlinien)
 - Zeigen extrem unterschiedlichen Phänotyp
- Alle F1-Individuen sind heterozygot am Marker
- Kopplungsphase zw. Marker und QTL ist identisch bei allen F1-Individuen → **vollständiges Linkage disequilibrium**

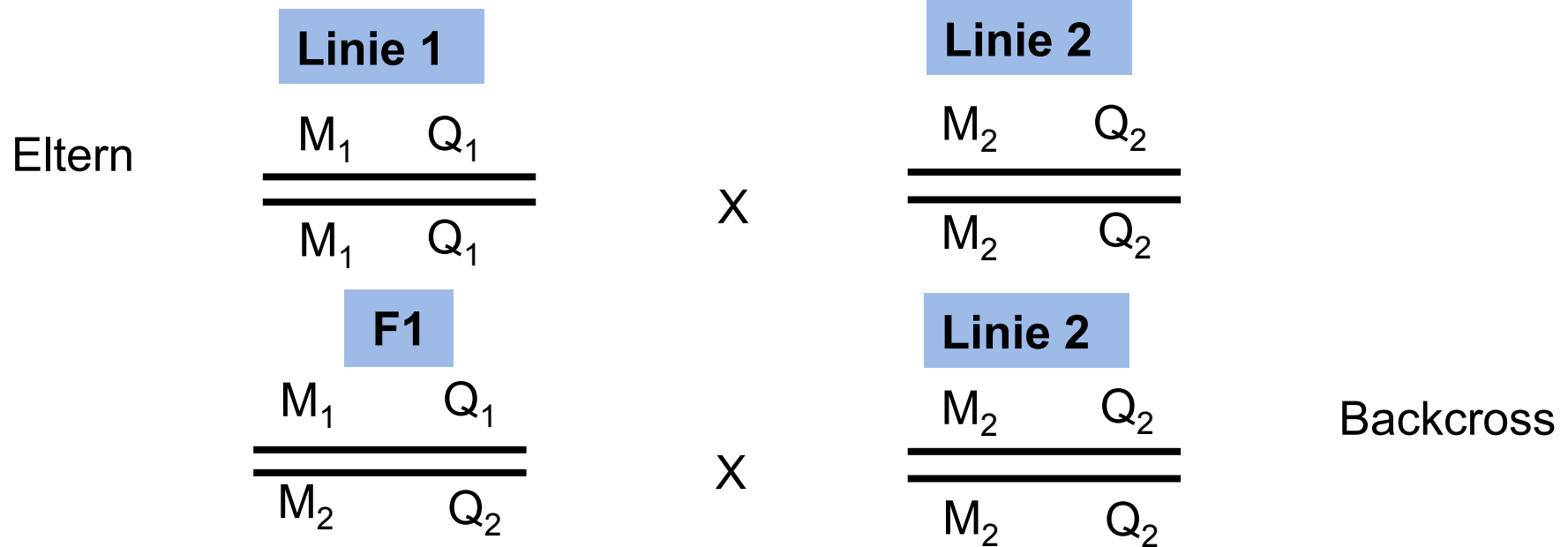
F1 backcross Design



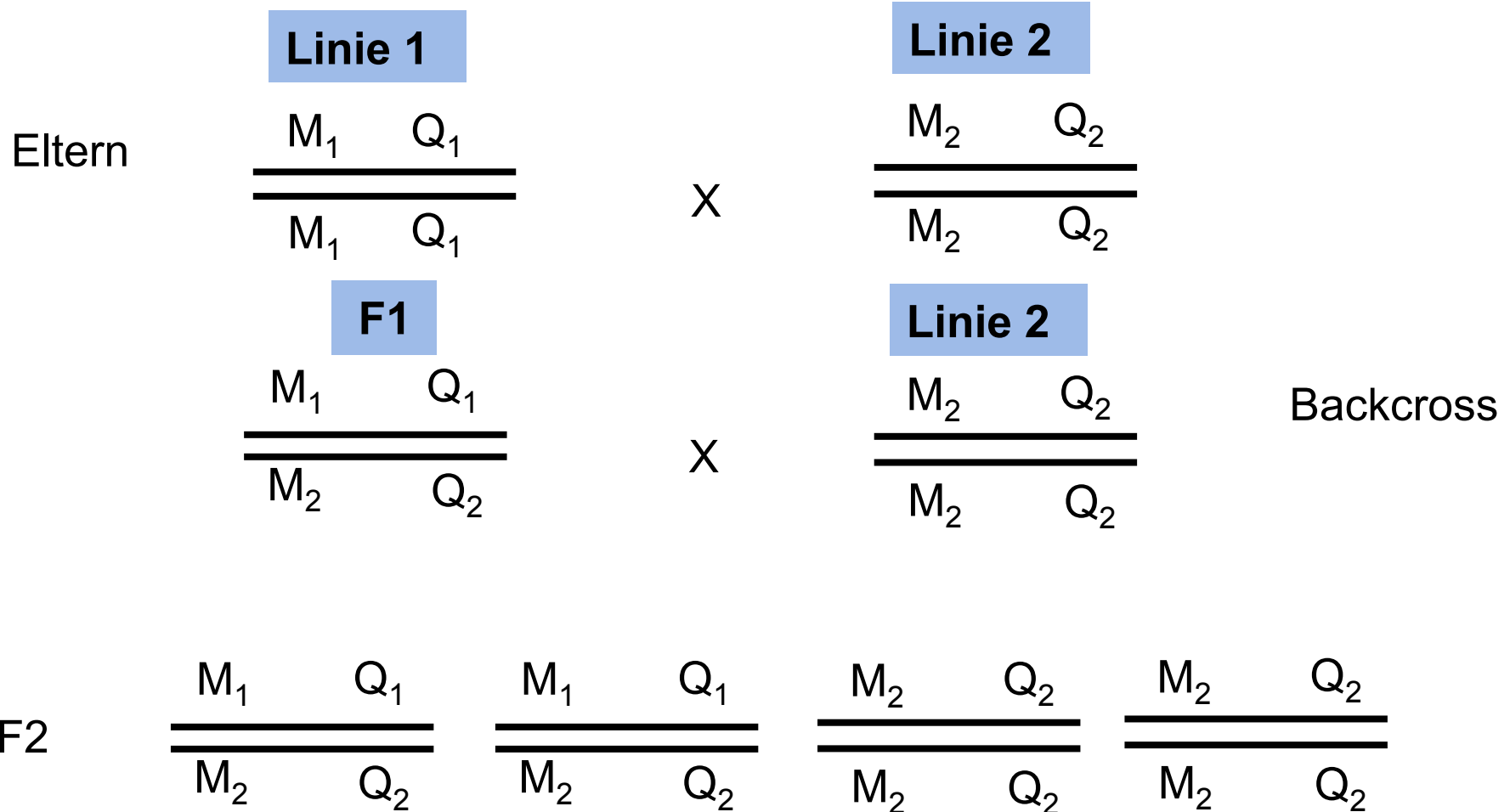
F1 backcross Design



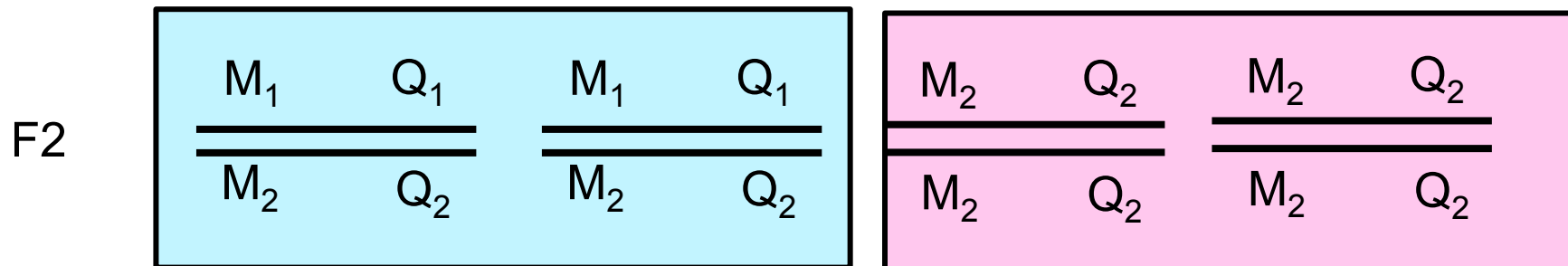
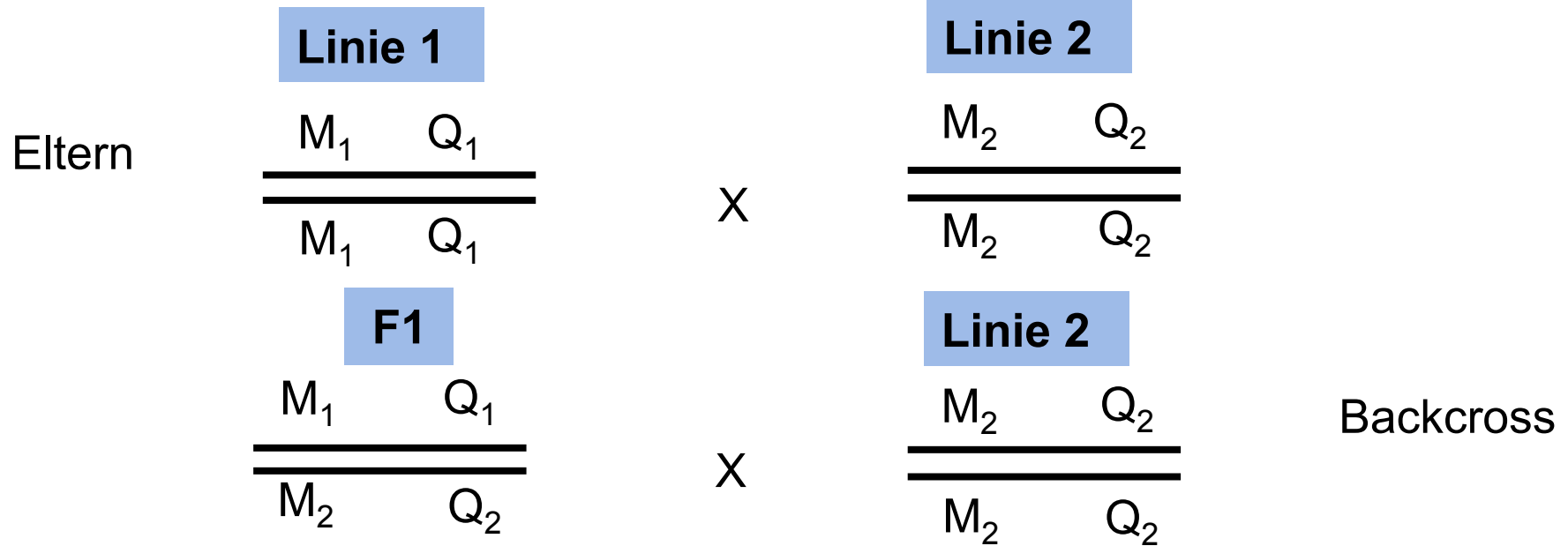
F1 backcross Design



F1 backcross Design



F1 backcross Design



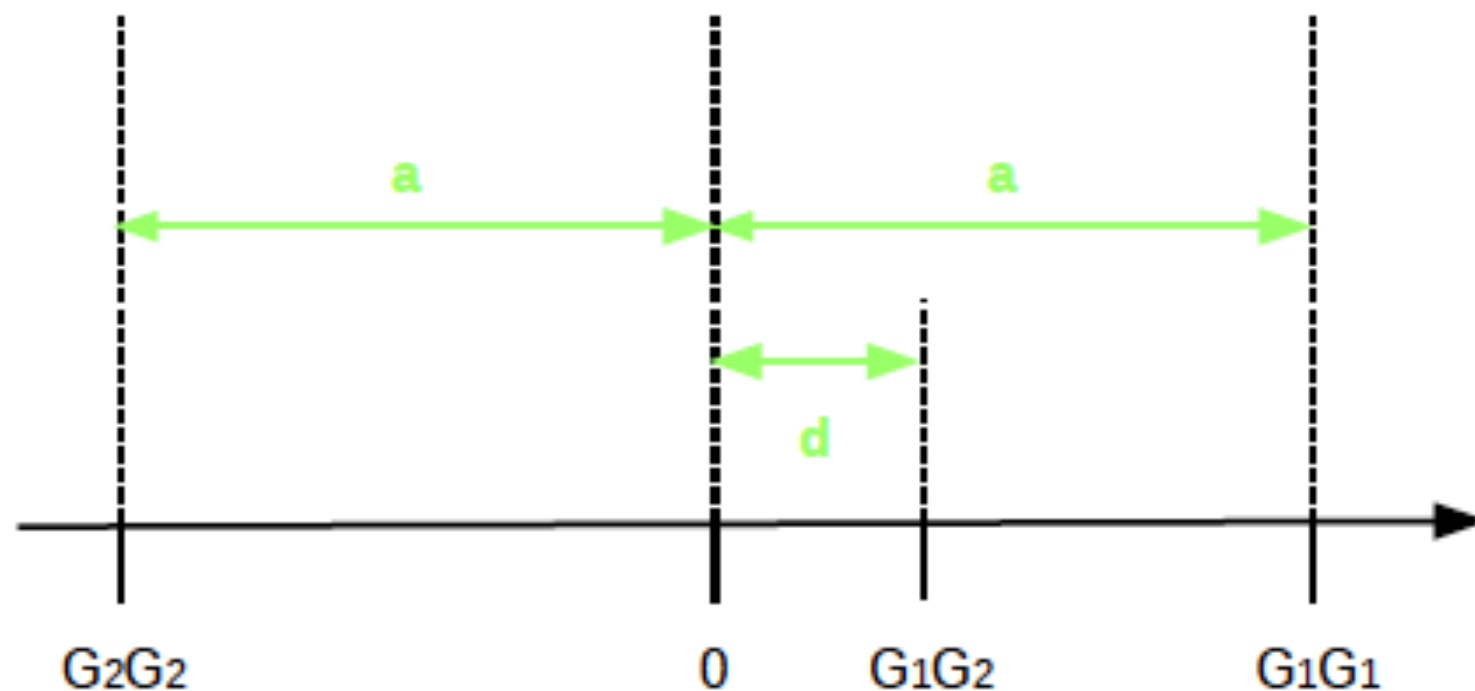
Vergleichen der Mittelwerte der am Marker Heterozygoten und Homozygoten

Wiederholung ...

- Genotypischer Wert (GW) ■
 - GW erfasst den genetisch bedingten Teil des phänotypischen Wertes
 - Annahme: 1 Genort, 2 Allele, Population im Hardy-Weinberg Gleichgewicht
 - Für bestimmten Genotypen $G_i G_j$ ist der genotypische Wert V_{ij} definiert als der mittlere Wert aller Individuen in der gleichen Umwelt mit Genotyp $G_i G_j$

Genotypischer Wert (GW) II

- Nullpunkt der Skala in Mitte zwischen Homozygoten G_1G_1 und G_2G_2
- Wahl des Nullpunkts beliebig, so am einfachsten



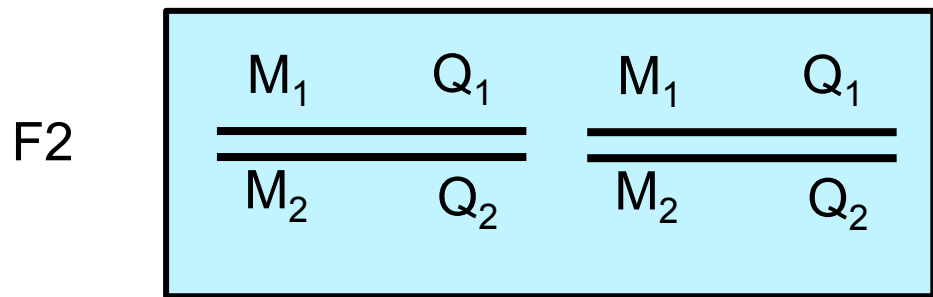
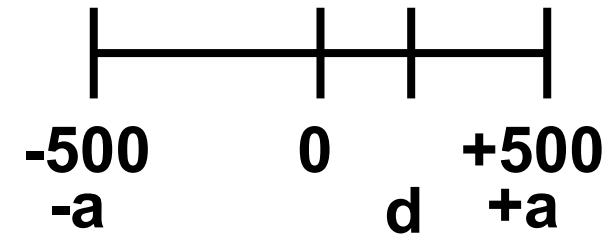
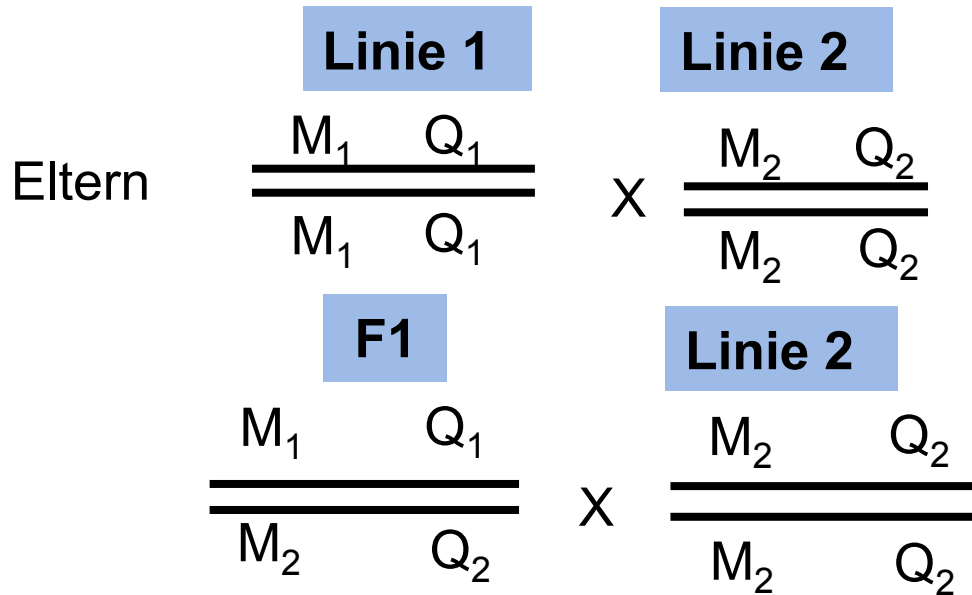
Zusammenfassung Genotypische Werte

Genotyp	genotypischer Wert
$G_1 G_1$	$V_{11} = a$
$G_1 G_2$	$V_{12} = d$
$G_2 G_2$	$V_{22} = -a$

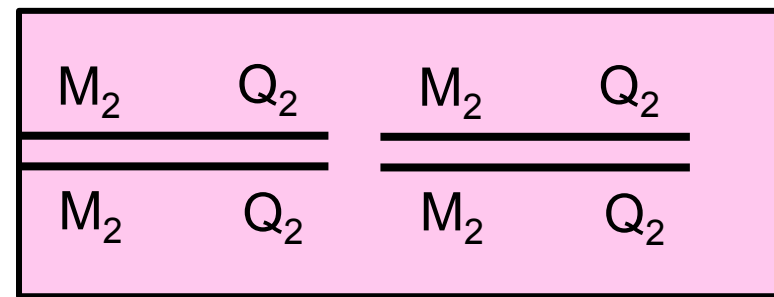
F1 backcross Design

$$Q_1 = +500$$

$$Q_2 = -500$$

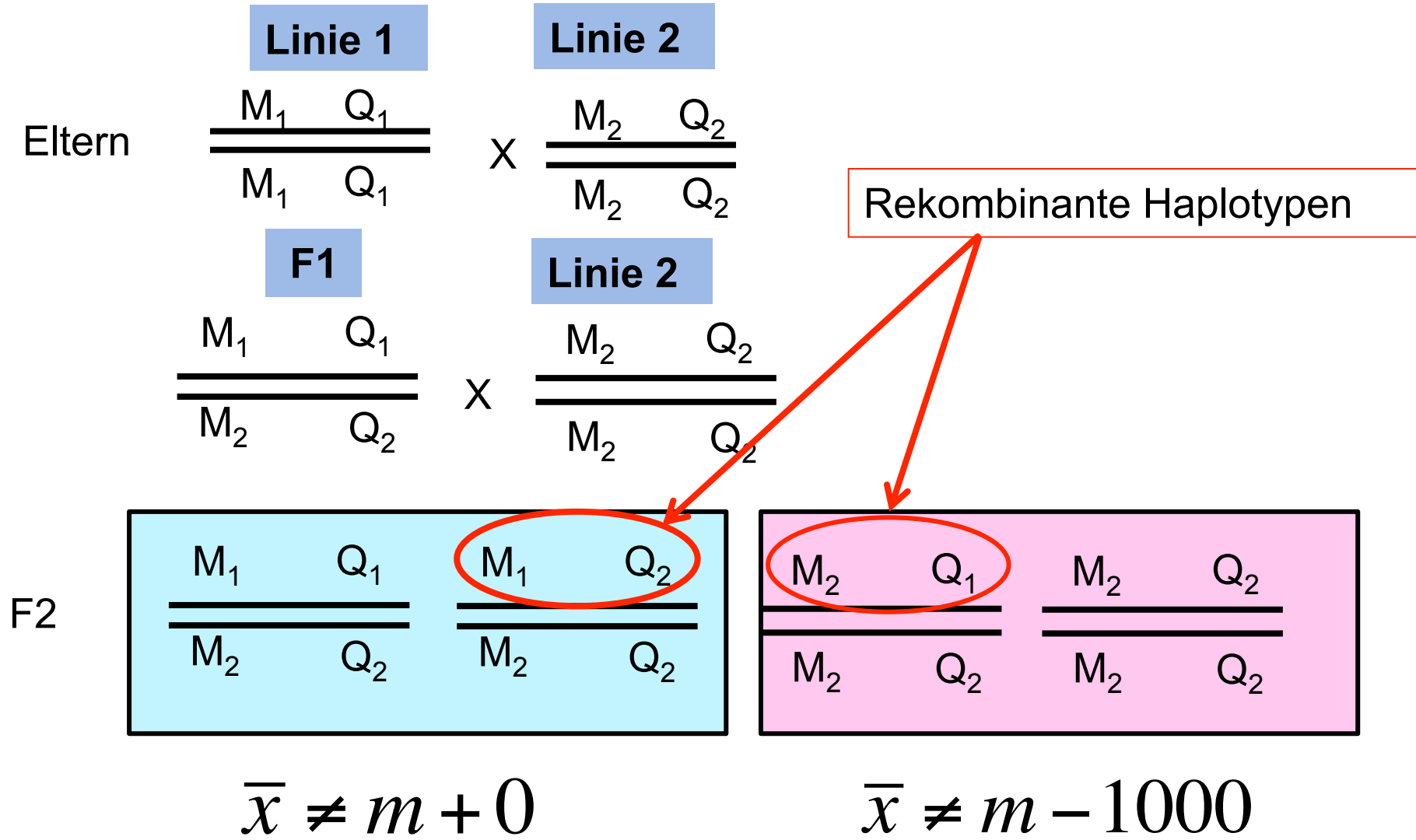


$$\bar{x} = m + 0$$



$$\bar{x} = m - 1000$$

F1 backcross Design



Rekombination

$$\frac{M_1 \quad Q_1}{M_2 \quad Q_2} \times \frac{M_2 \quad Q_2}{M_2 \quad Q_2}$$

$$\frac{M_1 \quad Q_1}{M_2 \quad Q_2}$$

Nicht-Rekombinate
Häufigkeit = 1-r

$$\frac{M_2 \quad Q_2}{M_2 \quad Q_2}$$

$$\frac{M_1 \quad Q_2}{M_2 \quad Q_2}$$

Rekombinate
Häufigkeit = r

$$\frac{M_2 \quad Q_1}{M_2 \quad Q_2}$$

$$r = \frac{\text{Anzahl an rekombinanten Haplotypen}}{\text{Anzahl an rekombinanten und nicht-rekombinanten Haplotypen}}$$

F1 backcross Design

- Folgende Hypothese wird getestet:
- H0: Mittelwerte beider Gruppen (M_1M_2 und M_2M_2) sind gleich (\rightarrow kein QTL)
- H0: Differenz zwischen M_1M_2 und $M_2M_2 = 0$

expected mean difference

$$E(Aa - aa) = (1 - 2r)(d + a)$$

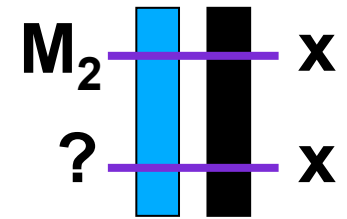
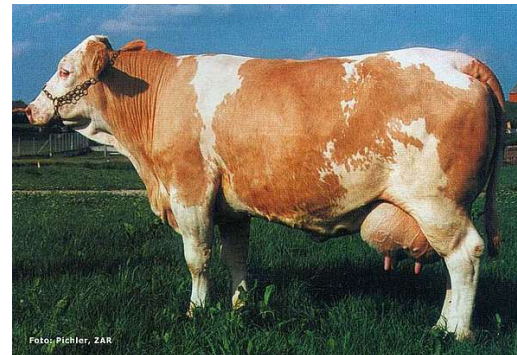
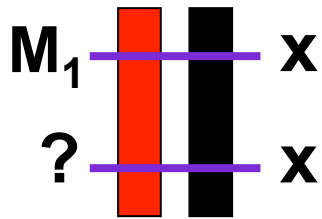
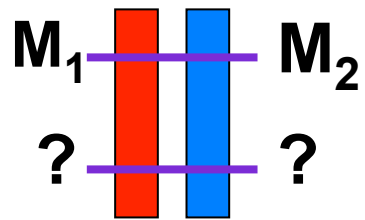
$$\Pr(Q_1Q_2 \mid M_1M_2) = \frac{\Pr(Q_1Q_2M_1M_2)}{\Pr(M_1M_2)}$$

Geno- type	Probabilities				Effect	Expected marker mean
	Marker	Marker + QTL	QTL marker			
M ₁ M ₂ Q ₁ Q ₂	0.5	(1-r)/2	(1-r)	=1	m+d	m + (1-r)d - ra
M ₁ M ₂ Q ₂ Q ₂		r/2	r		m-a	
M ₂ M ₂ Q ₁ Q ₂	0.5	r/2	r	=1	m+d	m + rd - (1-r)a
M ₂ M ₂ Q ₂ Q ₂		(1-r)/2	(1-r)		m-a	
<i>Sum</i>	1.0	1.0	2.0			

Töchterdesign (daughter design)

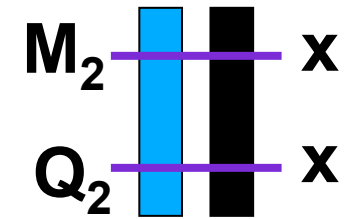
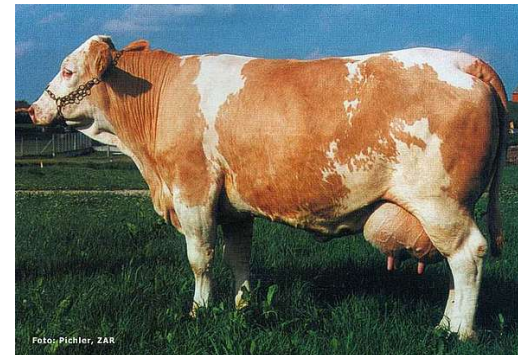
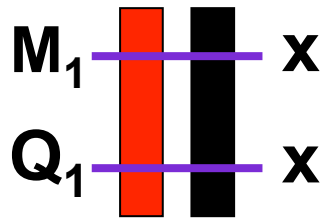
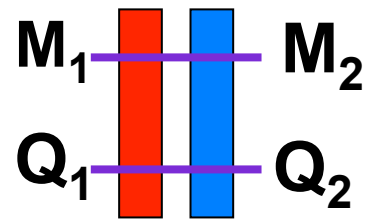
- Markergenotypen eines Stieres und seiner Töchter werden analysiert, um den Zusammenhang zwischen Marker und QTL zu untersuchen
- Töchter werden an Markern genotypisiert und auf Basis der Allele gruppiert und ihre Leistungen gemessen.
- Stier muss am Markerlocus (M_1M_2) heterozygot sein!
- Zum Auffinden QTL-bedingter Leistungsunterschiede muss auch am QTL-Locus (Q_1Q_2) heterozygot sein!

Töchterdesign (daughter design)



Folie: M. Dolezal

Töchterdesign (daughter design)



Folie: M. Dolezal

Töchterdesign (daughter design)

- Sehr grosse Anzahl an genotypisierten Töchtern nötig
- Zuordnung der Töchter in Gruppen problematisch:
 - Rekombination!
 - Töchter haben gleichen Markergenotyp wie der Vater (M_1M_2)
 - → es kann nicht entschieden werden, welches Allel vom Vater kommt und Töchter sind nicht informativ
 - Stier ist nicht heterozygot am QTL (M_1Q_1/M_2Q_1 oder M_1Q_2/M_2Q_2)
 - → Töchter unterscheiden sich nicht hinsichtlich QTL

Markergestützte Selektion

$$P = G + QTL + U$$

- BLUP Zuchtwertschätzung: es wird neben dem additiv-genetischen Effekt (*polygene Effekt*) des Tieres auch der Effekt des QTL-Genotyps (bzw. Marker) berücksichtigt
- → marker-unterstützter BLUP-Zuchtwertschätzung (MABLUP)
- Geschätzter Zuchtwert = polygene + QTL-spezifische Zuchtwertkomponente

**Populationsweite
genomweite
Assoziationsstudien**

***Whole genome-wide association studies
(GWAS)***

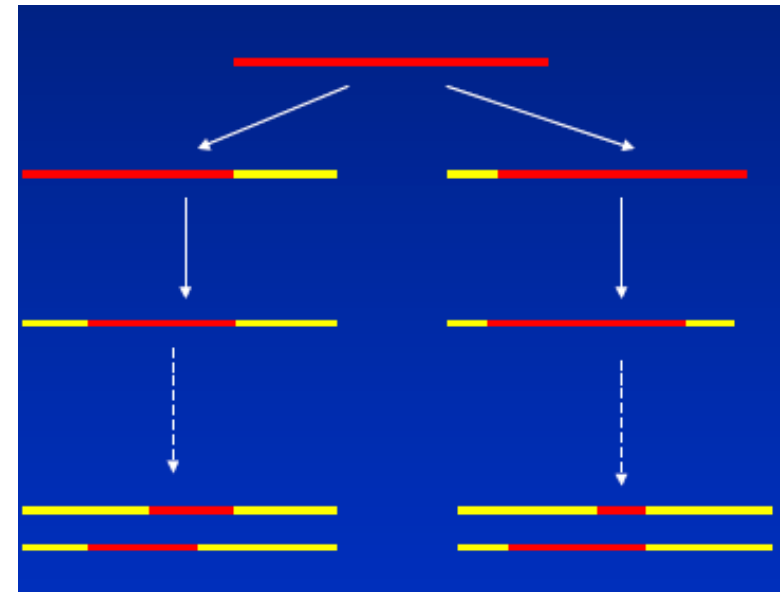
Genomweite Assoziationsstudien

- Bis jetzt oft nur einige wenige Marker getestet
- Innerhalb Familien
- Durch die Entwicklung von **SNP Chips** stehen tausende von SNP Marker für das Testen des Zusammenhanges zwischen Marker (QTL) und Phänotyp zur Verfügung
- **Assoziationen:**
 - **Direkte Assoziation:** Polymorphismus (SNP) ist selbst die ursächliche Variante (d.h. Marker = QTL)
 - **Indirekte Assoziation:** Polymorphismus (SNP) ist teilweise oder in vollständigem LD mit QTLs



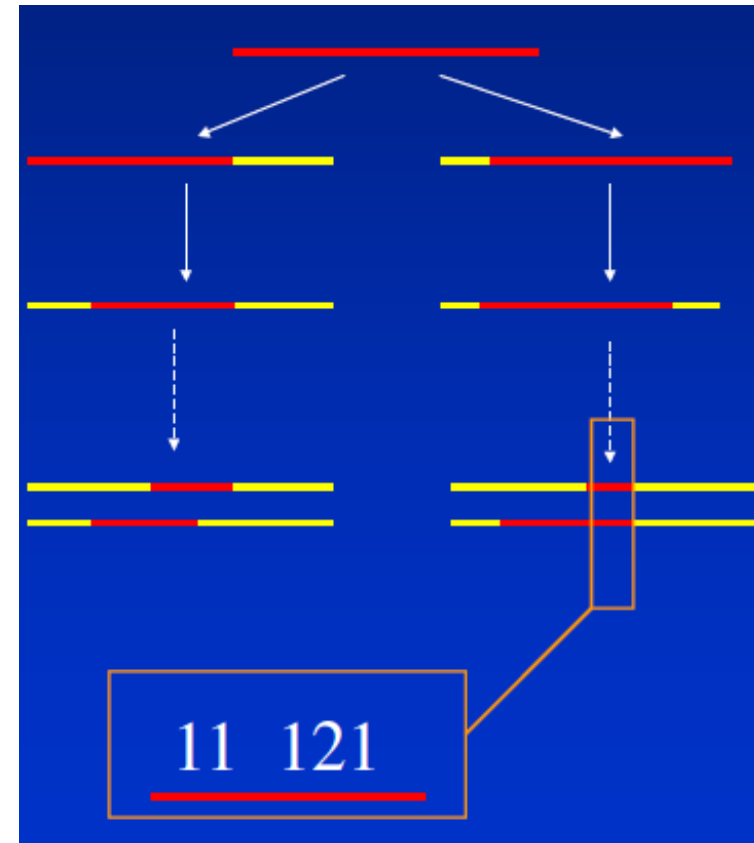
Genomweite Assoziationsstudien

- Voraussetzung für GWAS ist LD zwischen Marker und QTL
- QTL-Mapping basierend auf LD zeigt Assoziationen zwischen Marker und QTL auf Populationsebene:
- Assoziationen treten auf, weil kurze Chromosomenstücke in der gesamten Population vorhanden sind, welche vom gleichen gemeinsamen Ahnen stammen



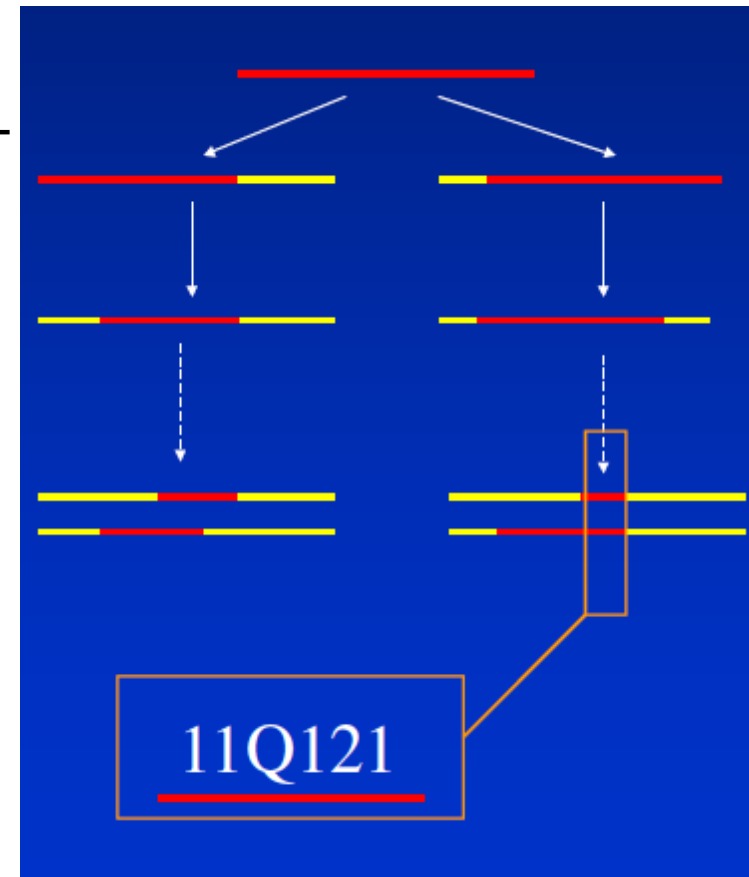
Genomweite Assoziationsstudien

- Voraussetzung für GWAS ist LD
- QTL-Mapping basierend auf LD zeigt Assoziationen zwischen Marker und QTL auf Populationsebene:
- Assoziationen treten auf, weil kurze Chromosomenstücke in der gesamten Population vorhanden sind, welche vom gleichen gemeinsamen Ahnen stammen
- Diese Chromosomenstücke tragen die gleichen Marker Allele (ohne Rekombination!)



Genomweite Assoziationsstudien

- Voraussetzung für GWAS ist LD
- QTL-Mapping basierend auf LD zeigt Assoziationen zwischen Marker und QTL auf Populationsebene:
- Assoziationen treten auf, weil kurze Chromosomenstücke in der gesamten Population vorhanden sind, welche vom gleichen gemeinsamen Ahnen stammen
- Diese Chromosomenstücke tragen die gleichen Marker Allele (ohne Rekombination!)
- Wenn ein QTL innerhalb Chromosomenstück vorkommt, sind auch die QTL-Allele gleich



GWAS mit Single-SNP Regression

$$y = 1_n \mu + Xg + e$$

y = Vektor mit Phänotyp (Merkmalswerte)

1_n = Vektor von „1“, teilt die Phänotypwerte y dem Mittelwert zu

X = Designmatrix, teilt Phänotypwerte den Genotypen (SNP Marker) zu

g = fixer Effekt des SNP (Markers)

e = zufälliger Fehler

GWAS mit Single-SNP Regression

$$y = \mathbf{1}_n \mu + Xg + e$$

y = Vektor mit Phänotyp (Merkmalswerte)

$\mathbf{1}_n$ = Vektor von „1“, teilt die Phänotypwerte y dem Mittelwert zu

X = Designmatrix, teilt Phänotypwerte den Genotypen (SNP Marker) zu

g = fixer Effekt des SNP (Markers)

e = zufälliger Fehler

Annahme: Marker beeinflusst den Phänotyp nur, wenn er im LD mit einem nicht beobachteten QTL ist.

Ein Beispiel

Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	2
4	4.871137	2	2
5	3.407128	1	2
6	2.335734	1	1
7	2.646192	1	1
8	3.762855	1	2
9	3.689349	1	2
10	3.685757	1	2

Ein Beispiel

Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	2
4	4.871137	2	2
5	3.407128	1	2
6	2.335734	1	1
7	2.646192		
8	3.762855		
9	3.689349	1	2
10	3.685757	1	2

Schätzen des Mittelwerts
und des SNP Effektes

Ein Beispiel

Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	1
4	4.871137	2	2
5	3.407128	1	2
6	2.335724	1	1
7	2.6461	1	1
8	3.7628	1	1
9	3.6893	1	1
10	3.6857	1	1

Schätzen des Mittelwerts
und des SNP Effektes

Hypothesentestung:

H0: Der SNP Marker hat keinen Einfluss
auf das Merkmal

H1: Der SNP Marker hat einen Einfluss auf
das Merkmal (weil er in LD mit QTL ist)

GWAS mit Single-SNP Regression

$$y = \mathbf{1}_n \mu + Xg + e$$

Lösen des Gleichungssystems

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' X \\ X' \mathbf{1}_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}_n' y \\ X' y \end{bmatrix}$$

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'_n \mathbf{1}_n & \mathbf{1}'_n X \\ X' \mathbf{1}_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'_n y \\ X' y \end{bmatrix}$$

Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	2
4	4.871137	2	2
5	3.407128	1	2
6	2.335734	1	1
7	2.646192	1	1
8	3.762855	1	2
9	3.689349	1	2
10	3.685757	1	2

$$\mathbf{1}_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'_n \mathbf{1}_n & \mathbf{1}'_n X \\ X' \mathbf{1}_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'_n y \\ X' y \end{bmatrix}$$

Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	2
4	4.871137	2	2
5	3.407128	1	2
6	2.335734	1	1
7	2.646192	1	1
8	3.762855	1	2
9	3.689349	1	2
10	3.685757	1	2

$$y = \begin{bmatrix} 2.030502 \\ 3.542274 \\ 3.834241 \\ 4.871137 \\ 3.407128 \\ 2.335734 \\ 2.646192 \\ 3.762855 \\ 3.689349 \\ 3.685757 \end{bmatrix}$$

Ein Single-SNP Regression - ein Beispiel

X ordnet Phänotypen den Genotypen zu

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$


Tier	Phänotyp	SNP Allel 1	SNP Allel 2
1	2.030502	1	1
2	3.542274	1	2
3	3.834241	1	2
4	4.871137	2	2
5	3.407128	1	2
6	2.335734	1	1
7	2.646192	1	1
8	3.762855	1	2
9	3.689349	1	2
10	3.685757	1	2

- SNP = biallelisch
- Allele werden als 1 und 2 kodiert
- X zählt die Anzahl an Kopien von Allel „2“
- 0 = homozygot für Allel 1
- 1 = heterozygot
- 2 = homozygot für Allel 2

$$X = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Single-SNP Regression - ein Beispiel

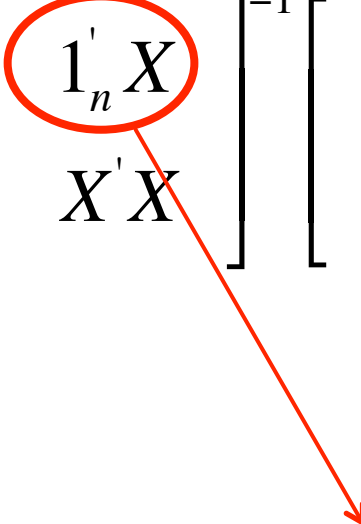
$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'_n \mathbf{1}_n & \mathbf{1}'_n X \\ X' \mathbf{1}_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'_n y \\ X' y \end{bmatrix}$$



$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 10$$

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$

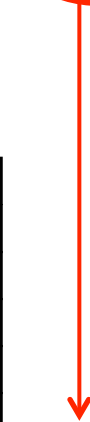


$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 8$$

[11111111111]

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 1 & 2 & 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 10$$


Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 33.8 \\ 31.7 \end{bmatrix}$$

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 33.8 \\ 31.7 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 0.28 & -0.22 \\ -0.22 & 0.28 \end{bmatrix} \begin{bmatrix} 33.8 \\ 31.7 \end{bmatrix}$$

Single-SNP Regression - ein Beispiel

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 1_n' 1_n & 1_n' X \\ X' 1_n & X' X \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ X' y \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 0.28 & -0.22 \\ -0.22 & 0.28 \end{bmatrix} \begin{bmatrix} 33.8 \\ 31.7 \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} 2.35 \\ 1.28 \end{bmatrix}$$

Ist der Effekt des SNP Markers signifikant?

- **Hypothesentestung - Signifikanztest**
- Das **Grundprinzip** der Hypothesentestung liegt in der Ermittlung des **Übereinstimmungsgrades** zwischen einer über einen Sachverhalt aufgestellten **Hypothese** und einem hierzu beobachtbaren **Stichprobenereignis**

Hypothesentestung - Signifikanztest

- **Frage:**
- Hat der SNP Genotyp einen Einfluss auf den Phänotyp?
- **Nullhypothese H_0 :** Genotyp hat **keinen** Einfluss auf den Phänotyp
- **Alternativhypothese H_1 :** Genotyp **hat** einen **Einfluss** auf den Phänotyp

Hypothesentestung - Signifikanztest

- **Zwei Arten von Fehlern:**
- **Fehler 1. Art** („falsch positive“)
 - Wir schliessen, dass der SNP Genotyp einen Einfluss hat, aber in Wirklichkeit hat er keinen Einfluss (H_0 wird abgelehnt)
- **Fehler 2. Art** („falsch negative“)
 - SNP Genotyp hat tatsächlich einen Effekt, wird aber nicht entdeckt (H_0 wird angenommen)

Hypothesentestung - Signifikanztest

Zwei Arten von Fehlern
H0: Es gibt keinen Effekt

Wirklichkeit

Testergebnis	Kein Effekt	Effekt
H0 abgelehnt	Fehler 1. Art	Korrekt 😊
H0 angenommen	Korrekt 😊	Fehler 2. Art

Hypothesentestung - Signifikanztest

- **P-Wert**
- Gibt die Wahrscheinlichkeit an, dass die beobachteten Effekte aus dem Testergebnis rein durch Zufall zustande kommen – in Wirklichkeit existiert kein Effekt (H_0 ist wahr)
- Signifikanzniveau/Irrtumswahrscheinlichkeit α :
 - Die Wahrscheinlichkeit, dass wir einen Fehler 1. Art begehen (H_0 verworfen obwohl tatsächlich kein Effekt) darf maximal 5%, 1%, 0.1% ... sein
- Wenn der P-Wert kleiner ist als die festgelegte Irrtumswahrscheinlichkeit (Fehler 1. Art) α , dann lehnen wir die Nullhypothese ab und glauben an einen Effekt