

# Funktionelle Genomanalysen 2023

## Asynchrone Übung zu SNP-Clusterplots

Dr. Janne Pott

09.-11. Juni 2023

# Allgemeines

- Array-abhängig werden die verschiedenen Allele mit Labels versehen (farbcodiert)
- Pro Sample und SNP werden die Intensitäten beider Labels bestimmt
- Jeder Plot ist SNP-spezifisch
- Jeder Punkt im Plot entspricht einem Sample
- X-Achse ist die log-transformierte Ratio der Intensitäten:  $\log_2(A/B)$
- Y-Achse ist das log-transformierte Produkt der Intensitäten:  
 $0.5 \cdot \log_2(AB)$
- AB Cluster sollte über den anderen beiden liegen, weil hier doppelte Intensität gemessen

# Clusterdefinition im Idealfall

- Genotyp AA
  - ▶ hohe A Intensität, niedrige B Intensität
  - ▶ Ratio ist größer als 1
  - ▶ Log-Transformierte Ratio ist größer als 0
  - ▶ **blaues Cluster**
- Genotyp BB
  - ▶ hohe B Intensität, niedrige A Intensität
  - ▶ Ratio ist kleiner als 1
  - ▶ Log-Transformierte Ratio ist kleiner als 0
  - ▶ **grünes Cluster**
- Genotyp AB
  - ▶ mittlere A Intensität, mittlere B Intensität
  - ▶ Ratio ist in etwa 1
  - ▶ Log-Transformierte Ratio ist in etwa 0
  - ▶ **gelbes Cluster**

# SNP Kriterien

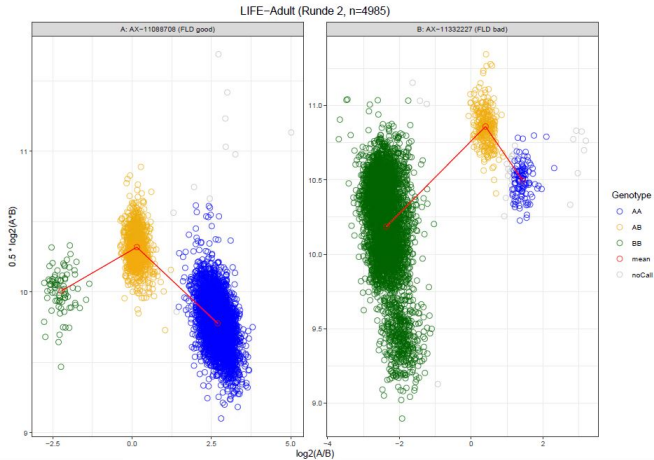
Kriterium	Bedeutung	Threshold
Call Rate	1 – Anteil missings	$< 97\%$
p(HWE)	p-Wert des exakten Fisher Test	$< 1 \times 10^{-6}$
p(PA)	p-Wert des Chi-Quadrat Test der AF pro Platte	$< 1 \times 10^{-7}$
MAC	Minor Allele Count	$< 2$
FLD	Minimaler Abstand zwischen den Cluster (bzgl. X-Achse)	$< 3.6$
HetSO	Abstand des AB-Clusters zu AA bzw. BB (bzgl. Y-Achse)	$< -0.1$
HomRO	Verteilung der Cluster (bzgl. 0 der X-Achse)	3 Cluster: $< -0.9$ 2 Cluster: $< 0.3$ 1 Cluster: $< 0.6$

**Tabelle:** Übersicht von relevanten SNP Kriterien. Die ersten vier sind Array-unspezifisch, die letzten drei sind spezifisch für Affymetrix Arrays. Es gilt: SNP filtern, wenn Wert kleiner als Threshold ist

# Plattform-unspezifische Kriterien

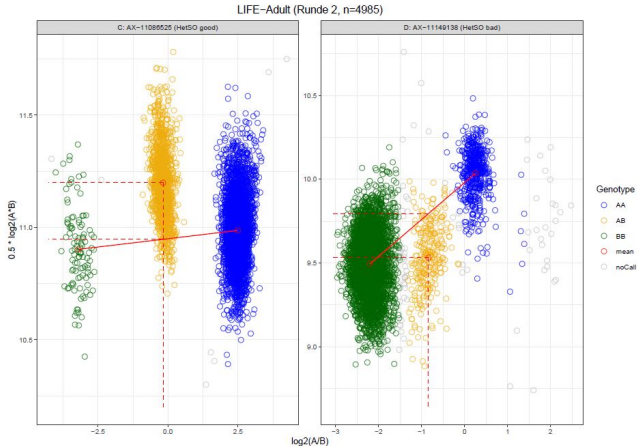
- **Call Rate:** s. HWE Aufgabe in erster Übung - es konnten nicht für alle Samples die Genotypen bestimmt werden
- **HWE:** s. HWE Aufgabe in erster Übung - in der Übung zu  $\alpha = 0.5$  getestet, jetzt auf multiples Testen korrigiert
- **PA:** Plattenassoziation, Test ob es Batcheffekte pro Platte gibt. Ebenfalls auf multiples Testen korrigiert
- **MAC:** wenn das selterne Allel nicht mindestens 2 mal vorkommt, dann ist das quasi kein SNP mehr, weil fast alle homozygot sind und nur eine Person heterozygot. Damit kann man nicht vernünftig rechnen

# FLD: Fisher's Linear Discriminant



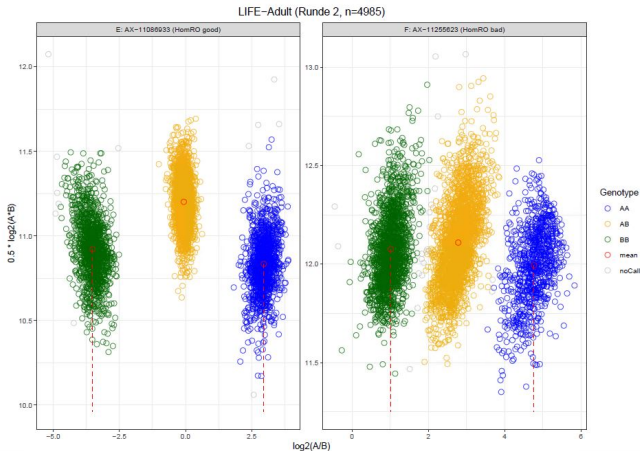
**Abbildung:** Links FLD ok, rechts FLD kleiner als 3.6 (zwischen AB und AA) → SNP muss gefiltert werden

# HetSO: Heterozygous Strength Offset



**Abbildung:** Links HetSO ok, rechts HetSO kleiner als -0.1 (Mittelpunkt AB deutlich unter der Verbindungslinie zwischen AA und BB) → SNP muss gefiltert werden

# HomRO: Homozygote Ratio Offset



**Abbildung:** Links HomRO ok, rechts HomRO kleiner als -0.9 (BB Cluster deutlich im positiven Bereich der x-Achse) → SNP muss gefiltert werden