Analyses of data from stratified populations: a methods comparison

LUPA Workshop in Statistical Methods for GWAS studies

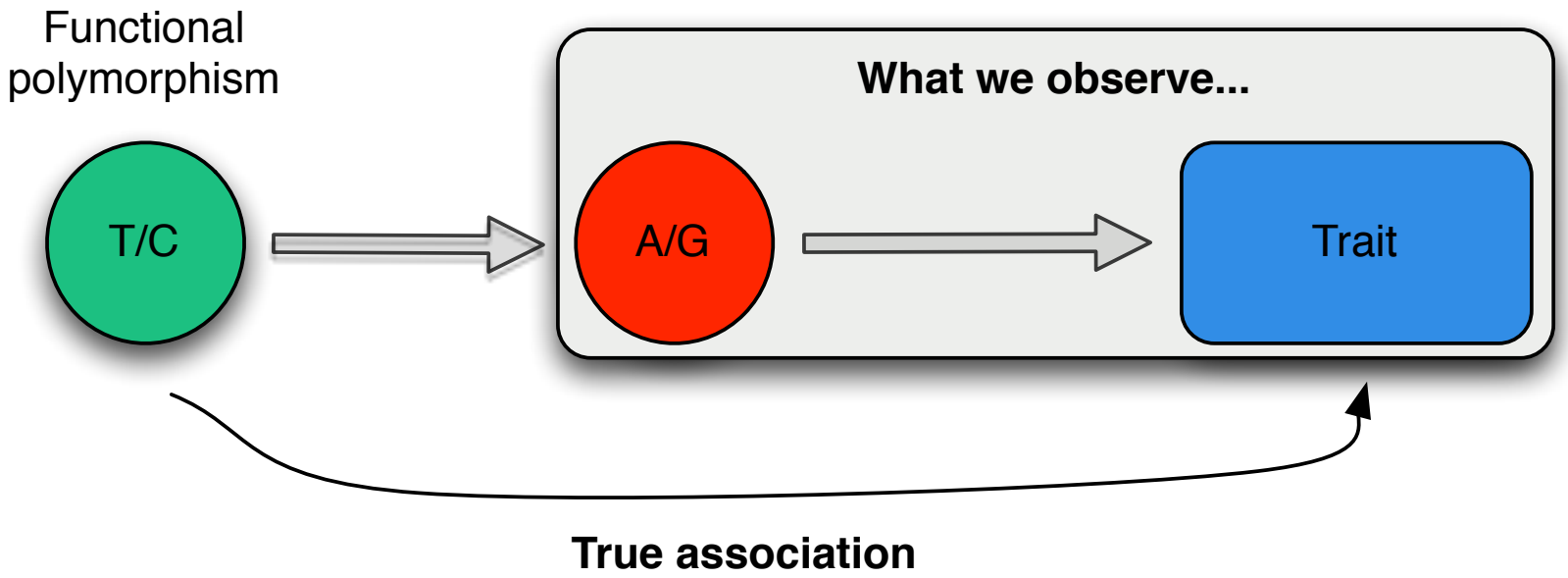
Marcin Kierczak*

*Computational Genetics Group
Faculty of Veterinary and Animal Breeding
Swedish University of Agricultural Sciences
Uppsala, SWEDEN

04-05 May 2011, Uppsala

Observed associations

Most of the observed associations are due to confounders!



Definition 1 – population (after: ABEL tutorial)

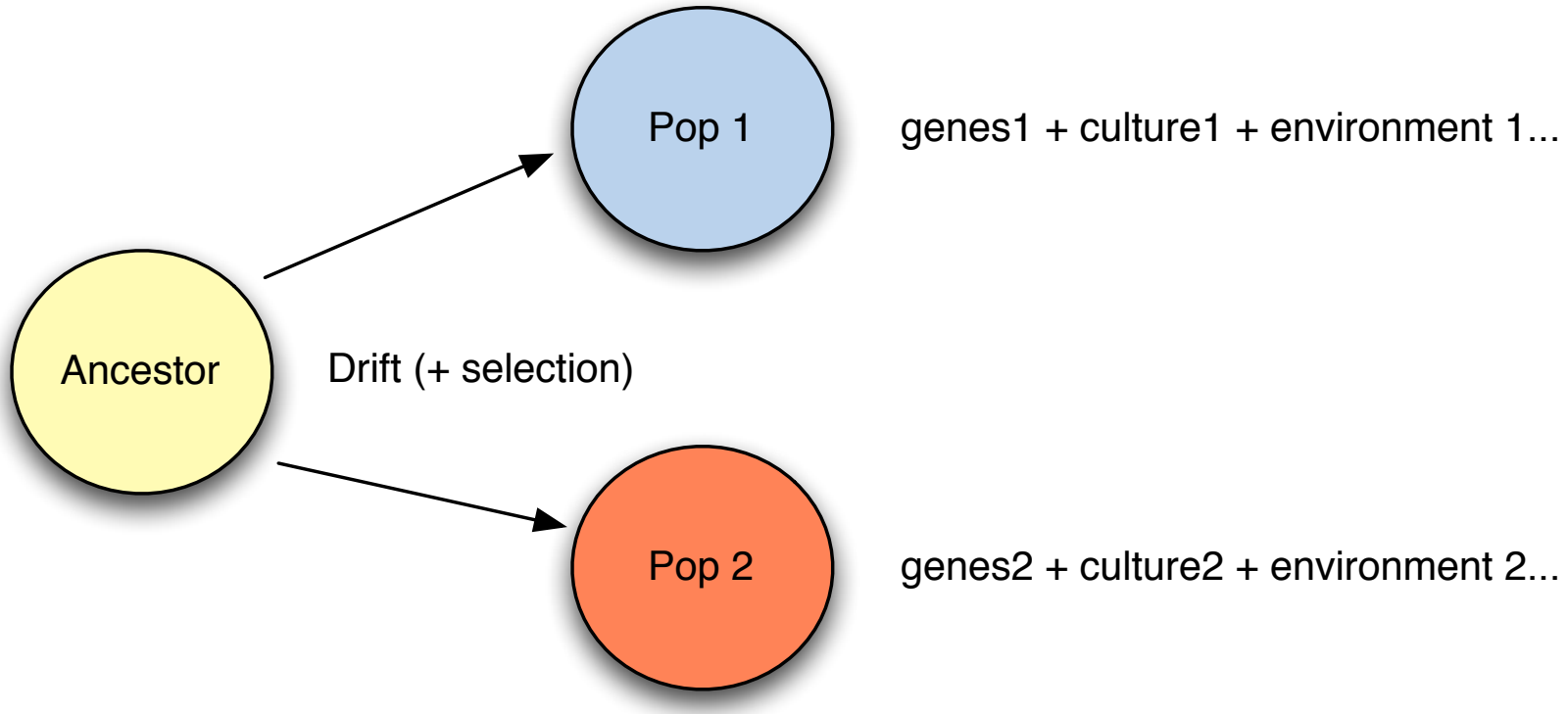
Individuals I_1 and I_2 belong to the same population P_1 and I_3 belongs to some other population iff:
$kin(I_1, I_2) > 0$ &
$\mathrm{kin}(\mathrm{I}_1, I_2) << kin(I_1, I_3)$ &
$\mathrm{kin}(\mathrm{I}_1, I_2) << kin(I_2, I_3)$

Definition 2 – population stratification (after: Price et al., 2006)

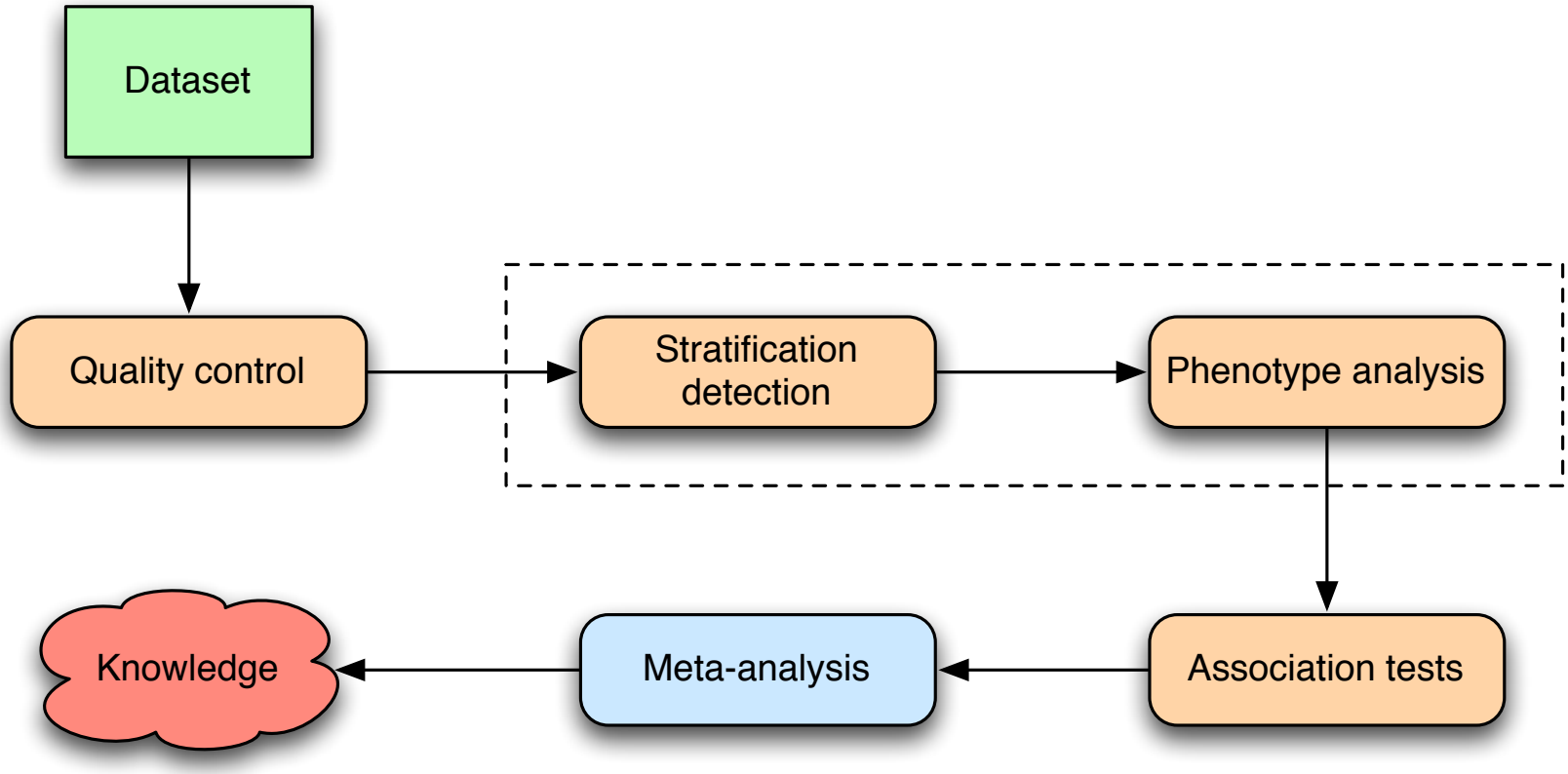
Population stratification – differences in allele frequency in cases and controls due to systematic ancestry differences.

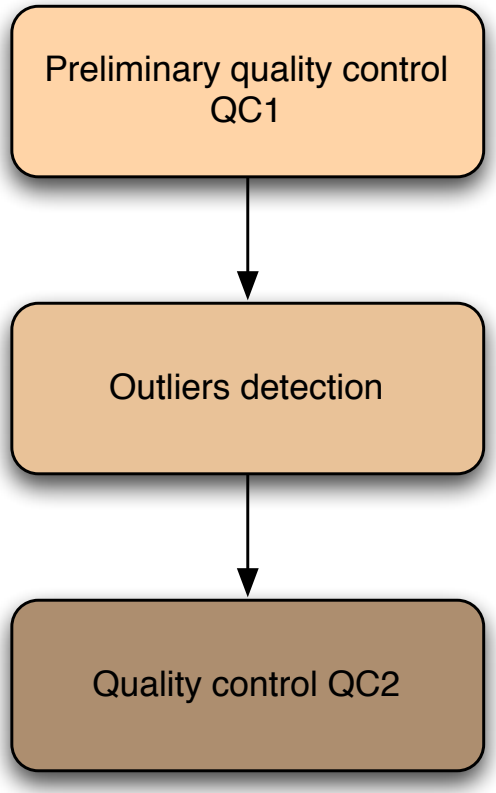
Consequence

Possible spurious associations in GWAS!


Pop 1
genes1 + culture1 + environment 1...
Ancestor
Drift (+ selection)
Pop 2
genes2 + culture2 + environment 2...

Workflow
Basic workflow in GWAS

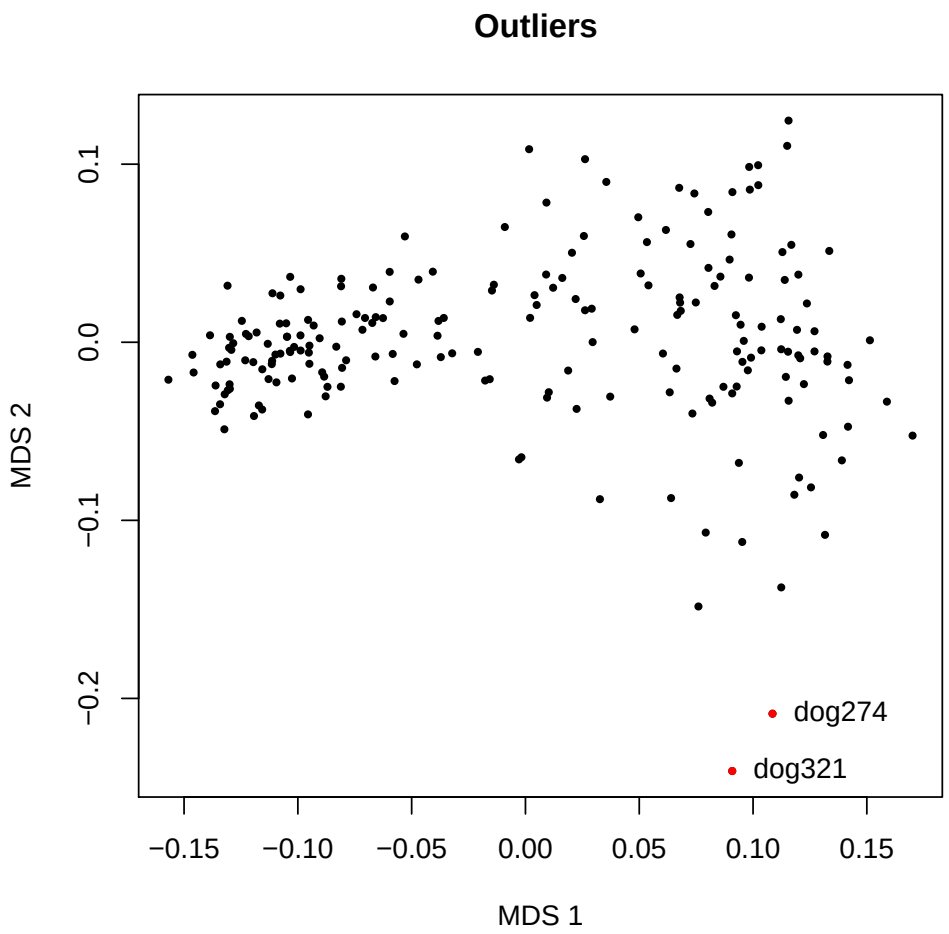




Three phases of QC

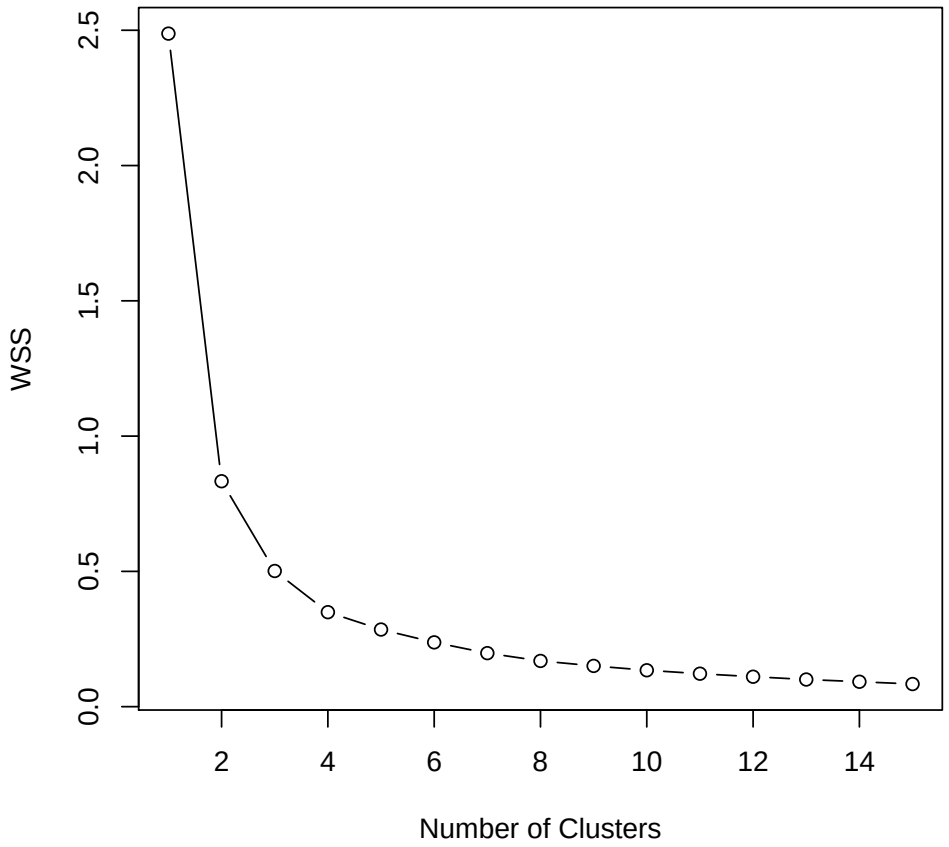
- **Preliminary (QC1)** – call rate, too close kinship, only strong departures from HWE.
- **Outliers detection** – removal of outliers using genomic kinship-based distances.
- **Final (QC2)** – departures from HWE, esp. in controls.

Qc1 and QC2 are iterative!



Outliers detection

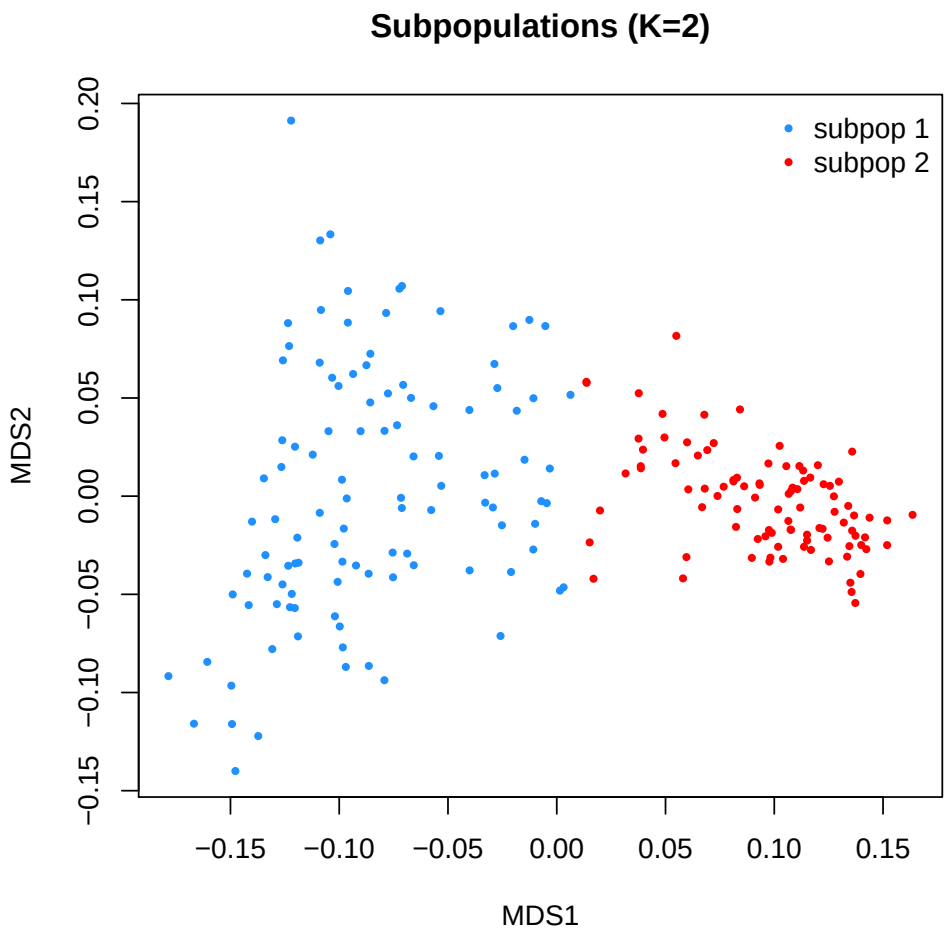
- Compute IBS matrix using autosomal markers (subset).
- Transform IBS matrix to dist. matrix.
- Run MDS (=>2D)
- Visualize and identify outliers.
- Remove outliers (avg. dist. from cluster centre?).



How many subpops?

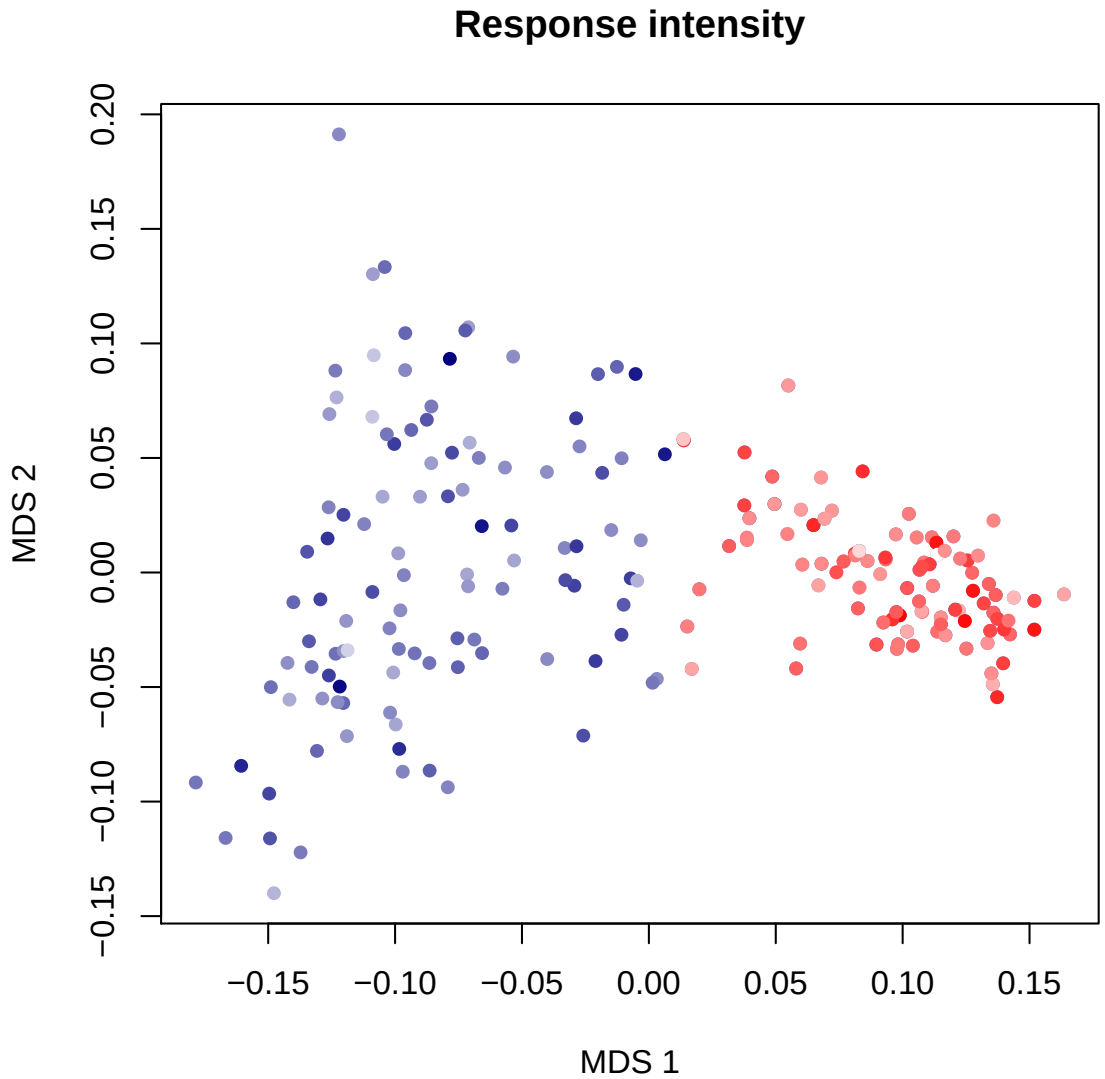
- Visual assessment.
- Clustering e.g., K-means.

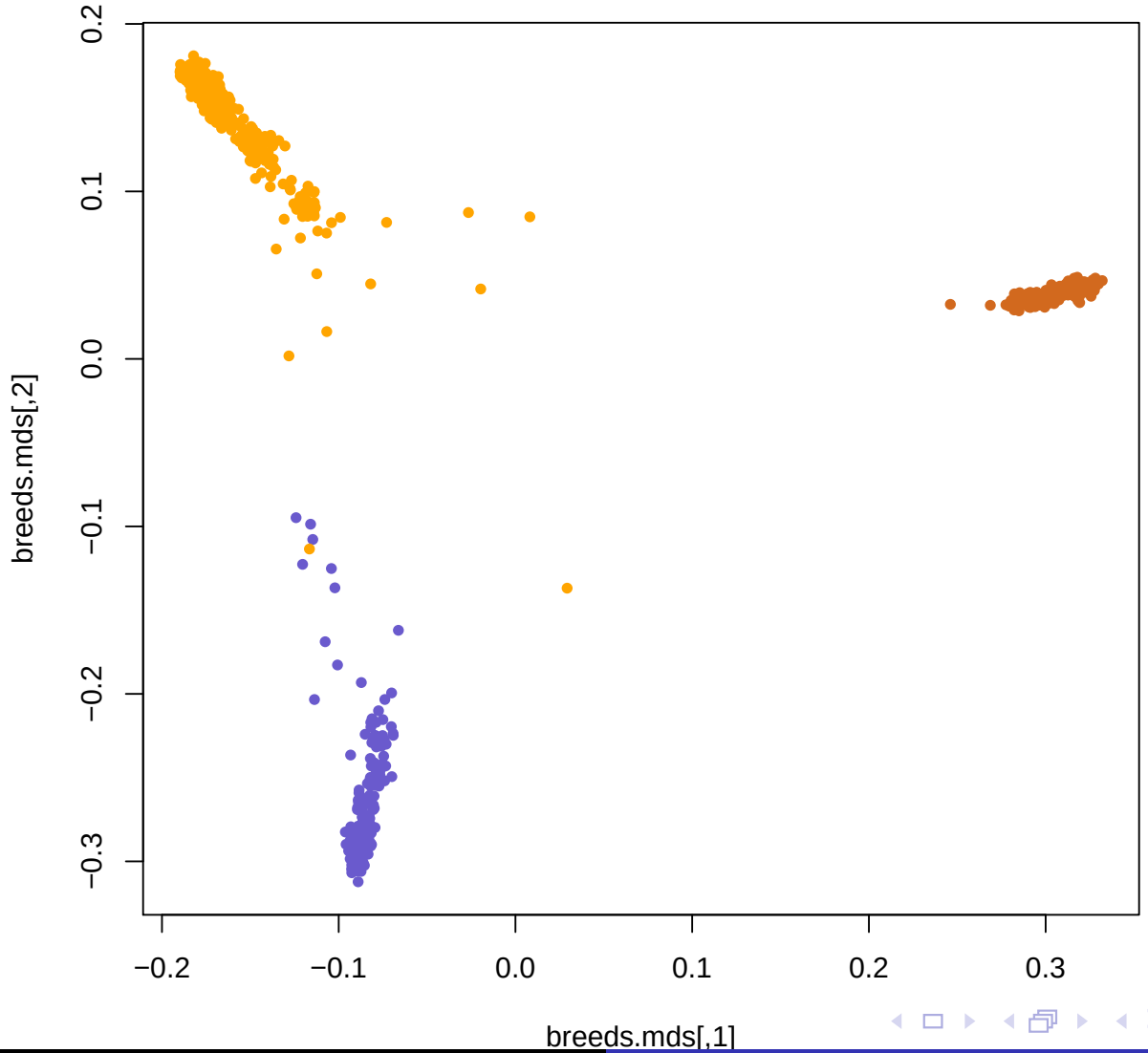
In K-means clustering, plot WCSS for different K and look for a bend. Similar to a scree test.



What else can be done?

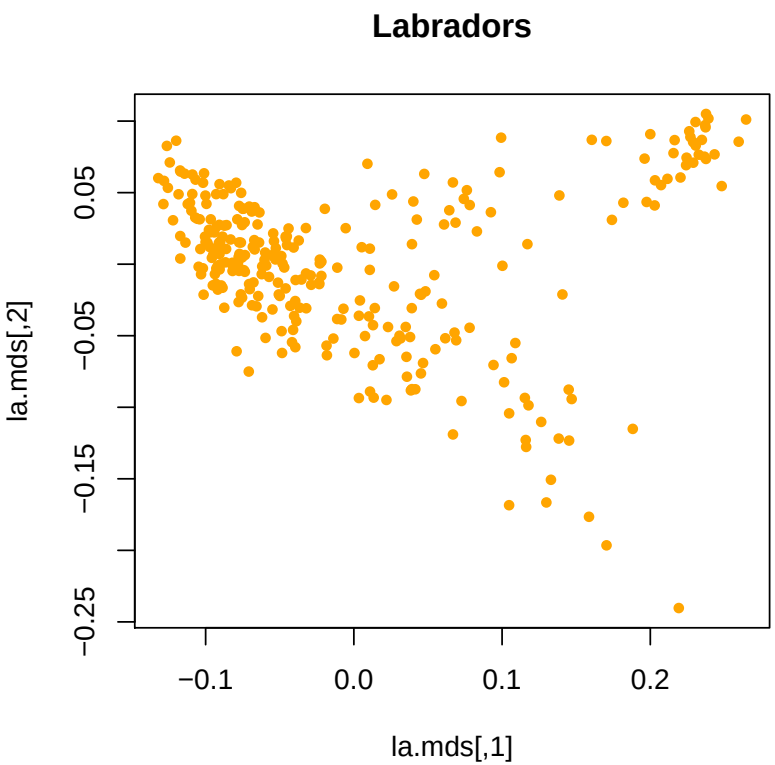
- Check which individuals fall into the clusters.
- Compare sex distribution in the clusters.
- Compare distribution of cases/controls in the clusters.
- More strata?

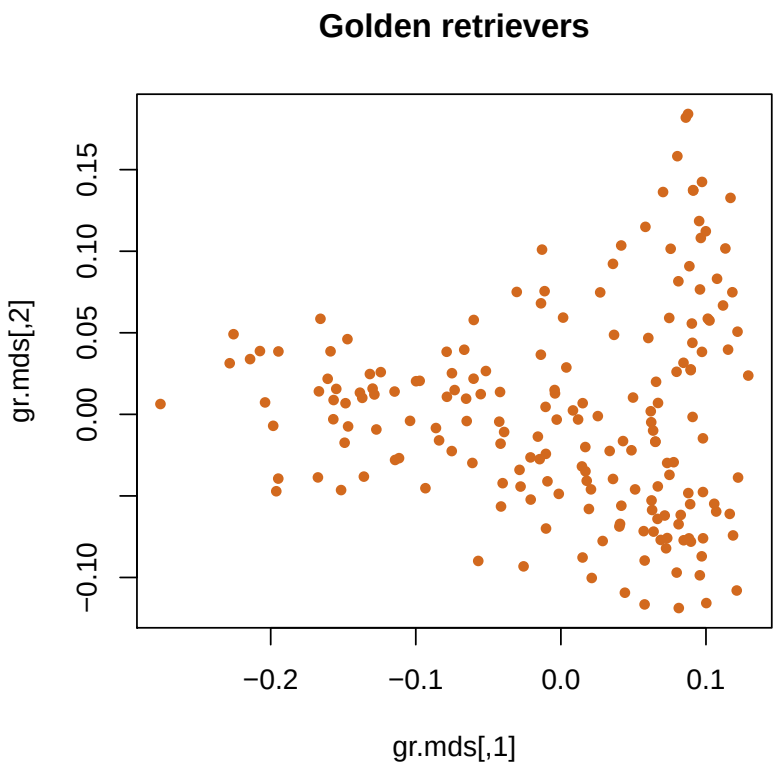

Response intensity
MDS 2
MDS 1
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15


breeds.mds[,2]
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
breeds.mds[,1]

Workflow

Population structure – separately for each breed





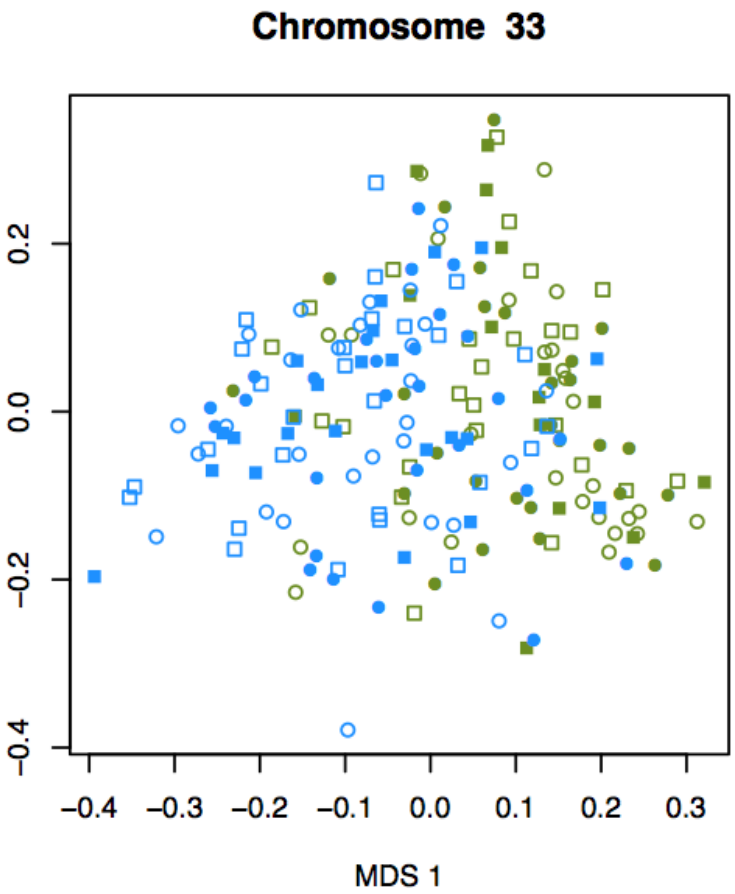
It is always worth to include all available information and to make comparisons between clusters.

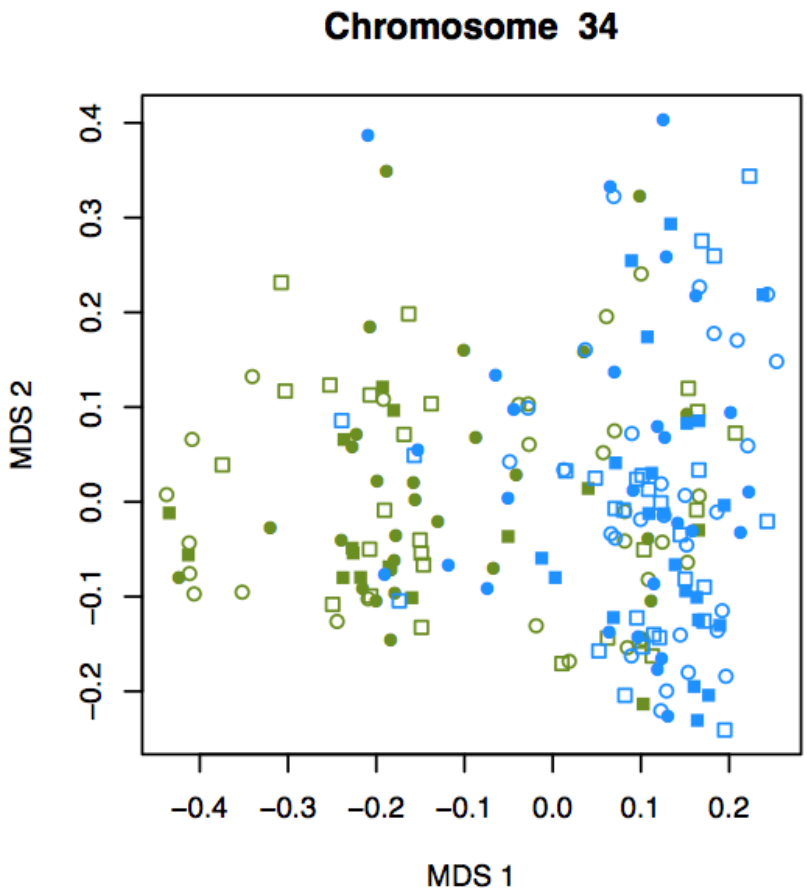
Consider

- Geographic location.
- Kennel club.
- Age (generation).
- Type of dog, e.g.: working, show.
- Import due to, e.g.: fashion.

Workflow

Population structure – separately for each chromosome

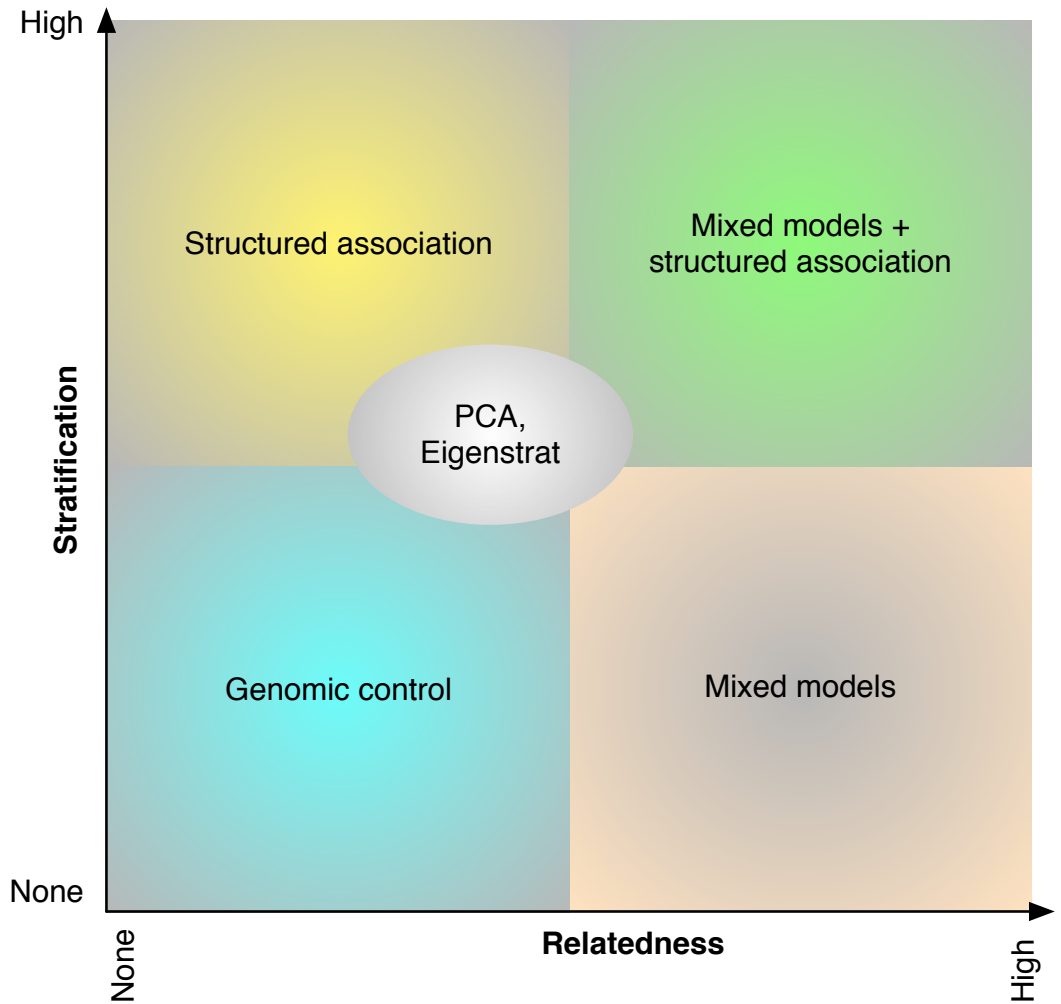




Different approaches exist to correct for stratification:

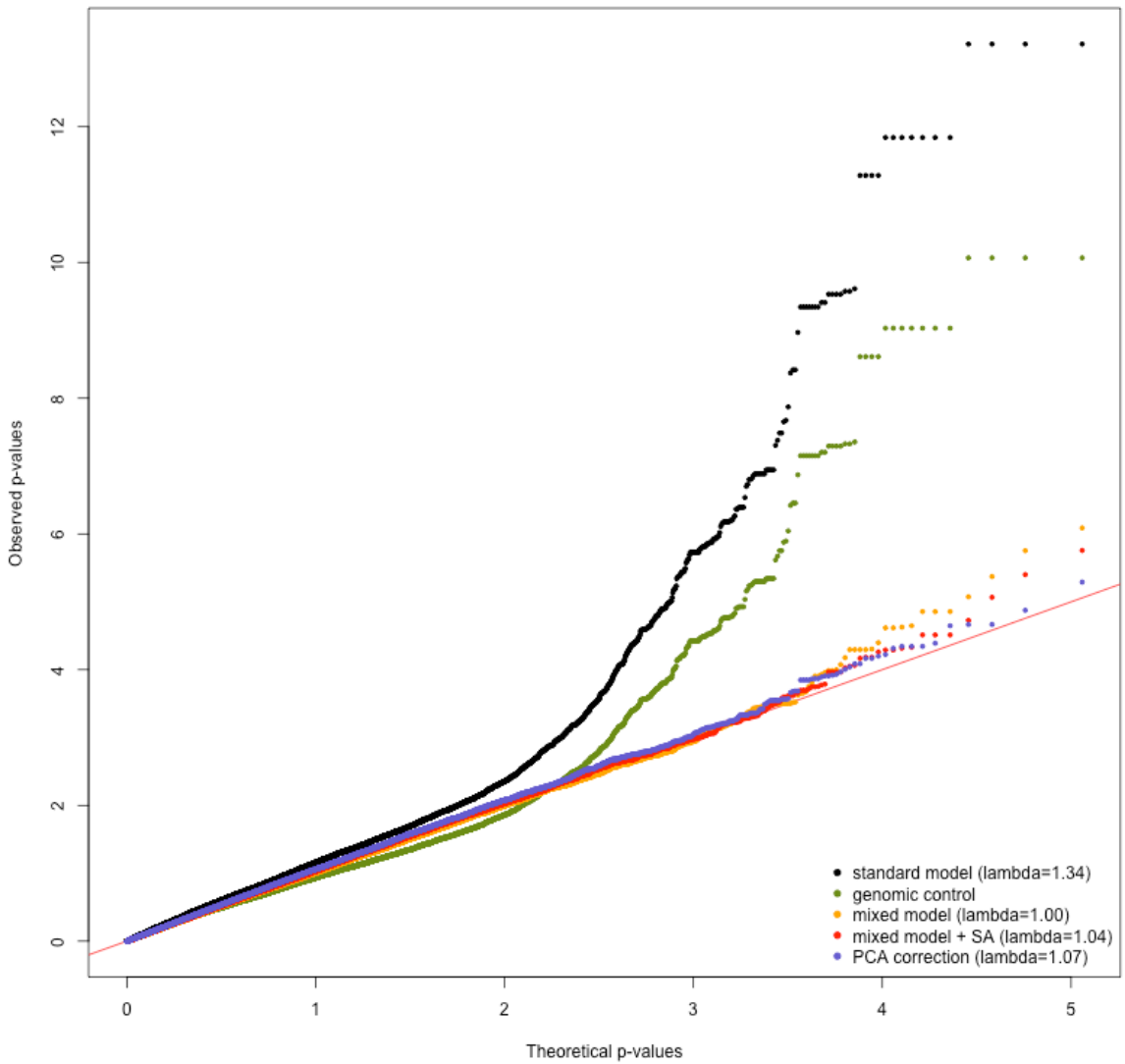
- Simple association tests.
- Genomic control.
- PCA-based correction – Eigenstrat.
- Mixed models.
- Structured association.
- Combinations of the above!
- Use of co-variate, e.g.: breed, address, etc.

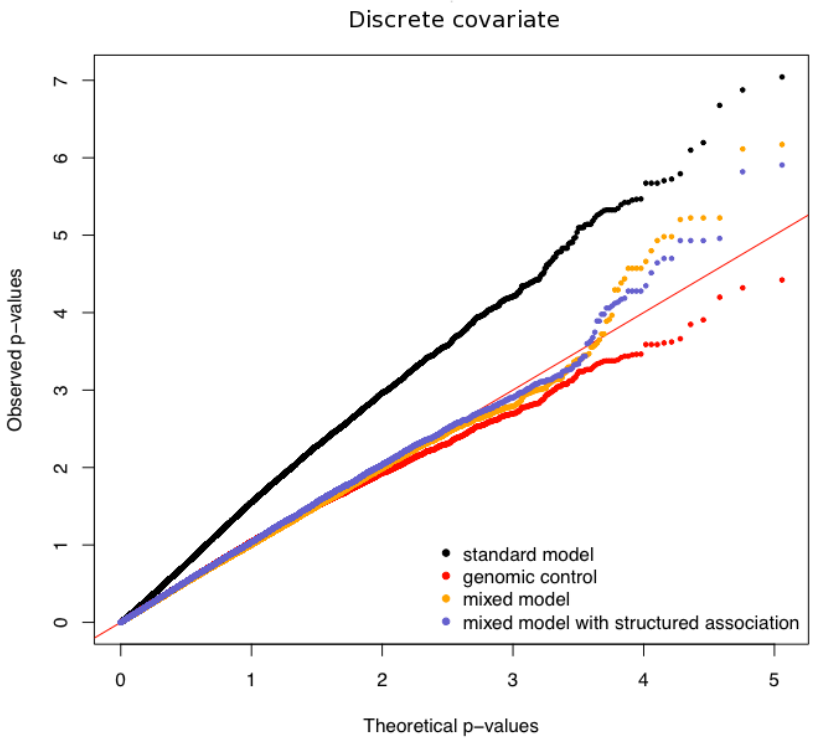
Which method to use?

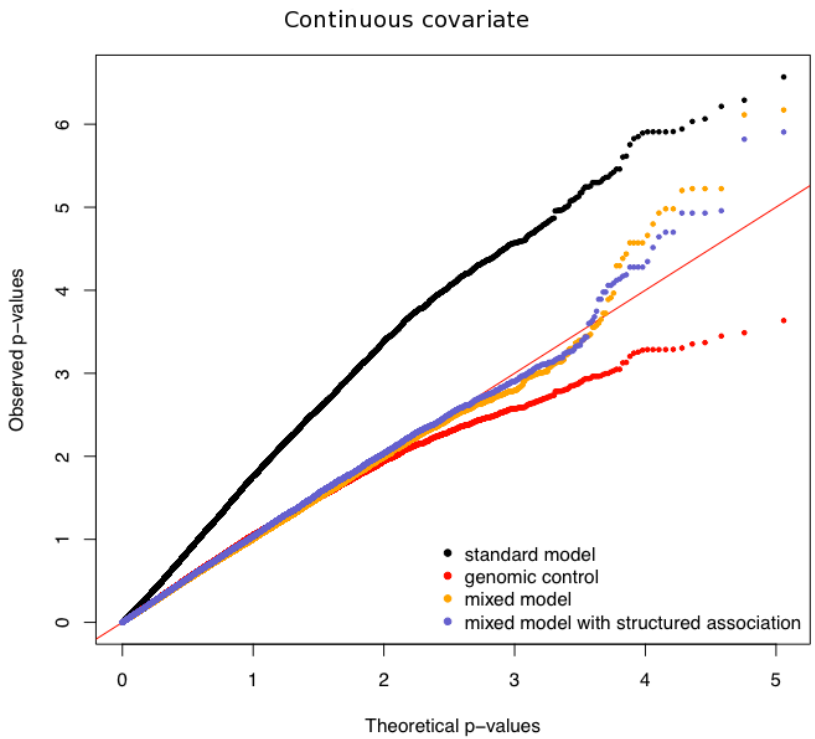


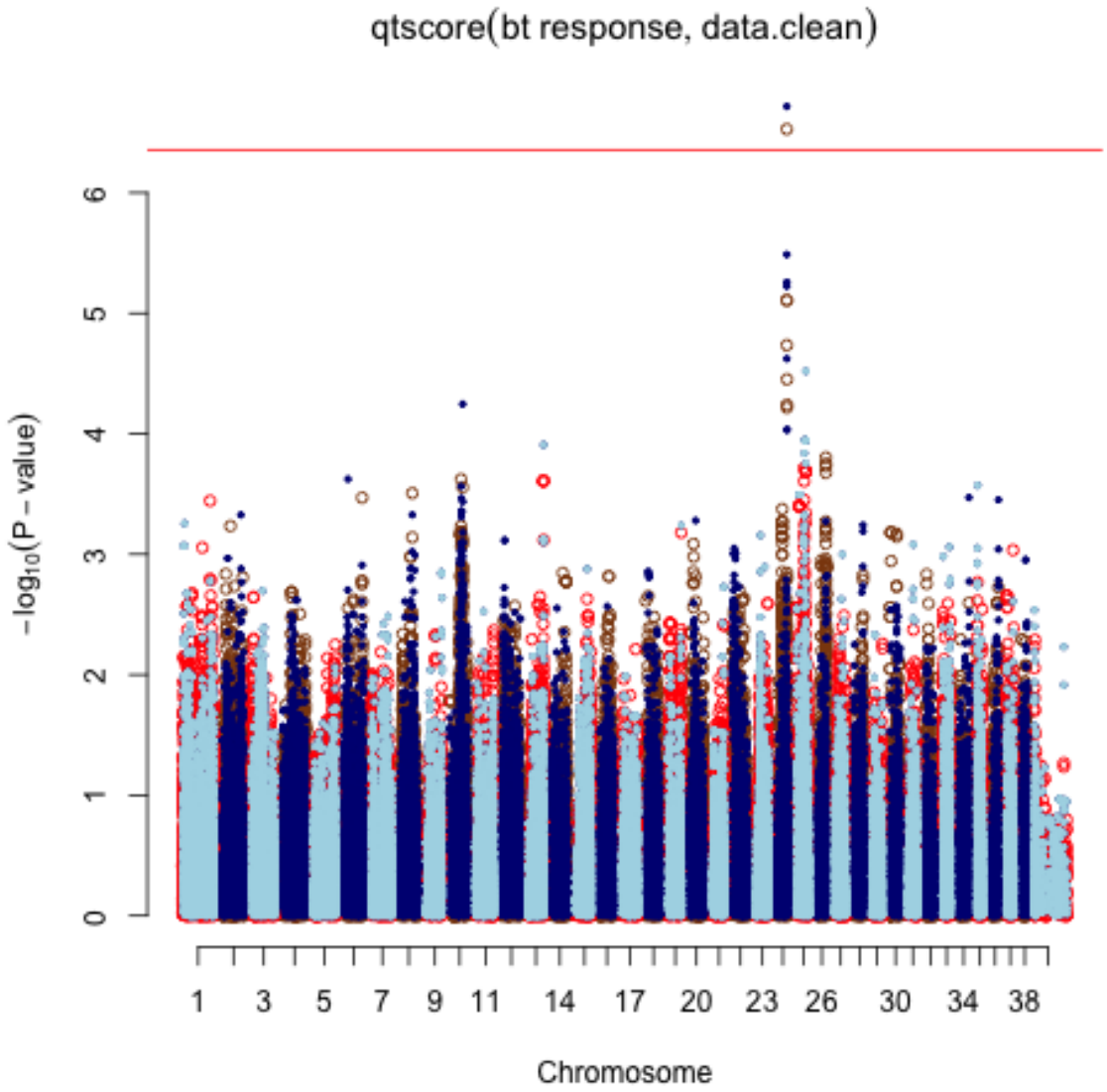
after: Aulchenko et. al., 2010 (ABEL Tutorial)

Workflow
Comparing methods...




Discrete covariate
Observed p−values
Theoretical p−values
standard model
genomic control
mixed model
mixed model with structured association


Continuous covariate
Observed p−values
Theoretical p−values
standard model
genomic control
mixed model
mixed model with structured association


qtscore(bt response, data.clean)
−log10(P − value)
0
1
2
3
4
5
6
1
3
5
7
9
11
14
17
20
23
26
30
34
38
Chromosome

Conclusion

There is no single universal approach. Each dataset is specific and has to be examined in detail! Look at your data and experiment with different approaches!

Occam's razor

The simplest approach that gives good correction should always be chosen! BUT: do not fall into a trap of choosing a method that corrects "by chance". Do not use GC when you know your data is stratified!

The END

Thank You! and:

- Leif Andersson
- Yurii Aulchenko
- Örjan Carlborg
- Hille Fieten
- Dirk Jan de Koonig
- Kerstin Lindblad-Toh
- Xia Shen
- Katarina Tengvall