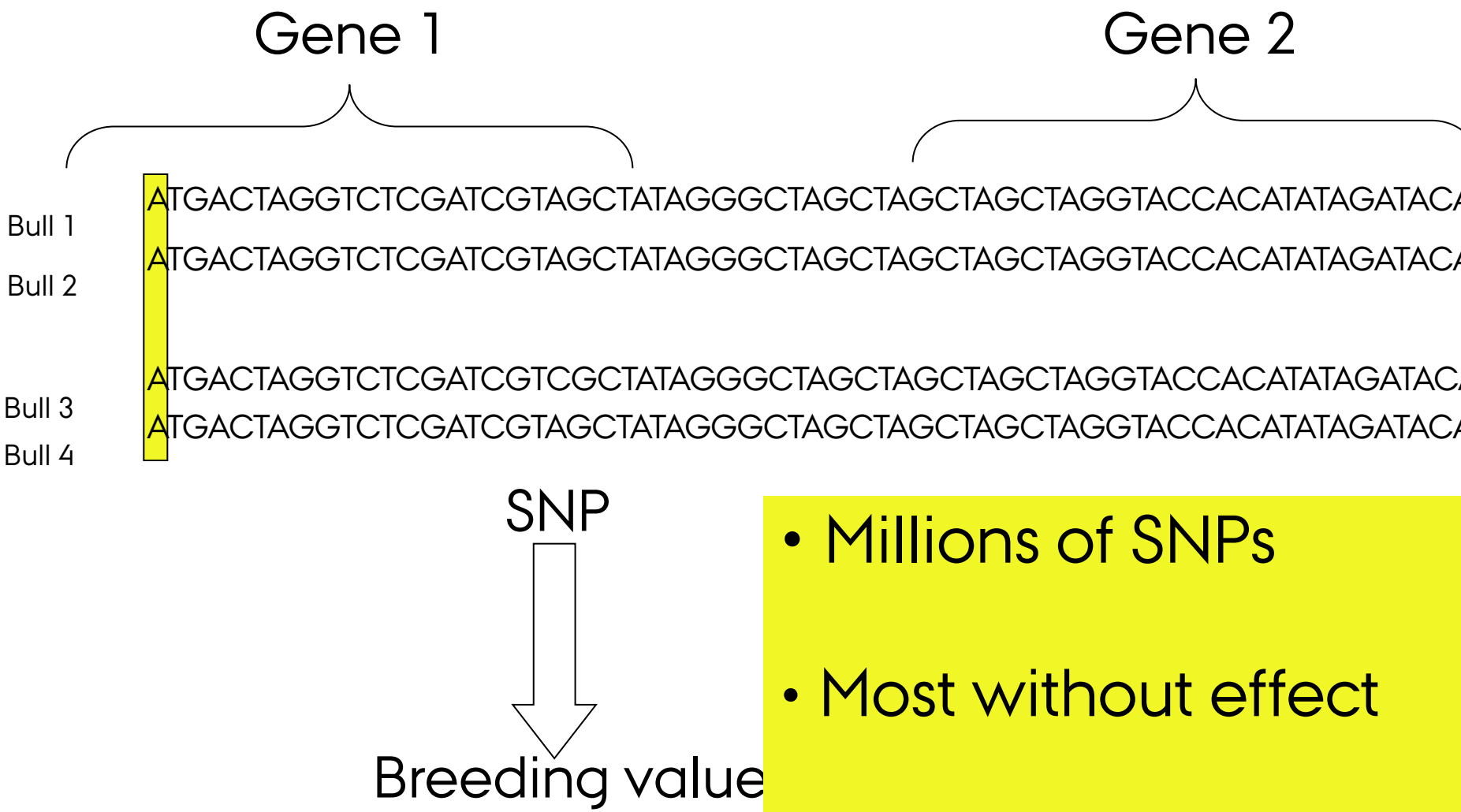
Sequencing for better genomic predictions

Mogens Sandø Lund

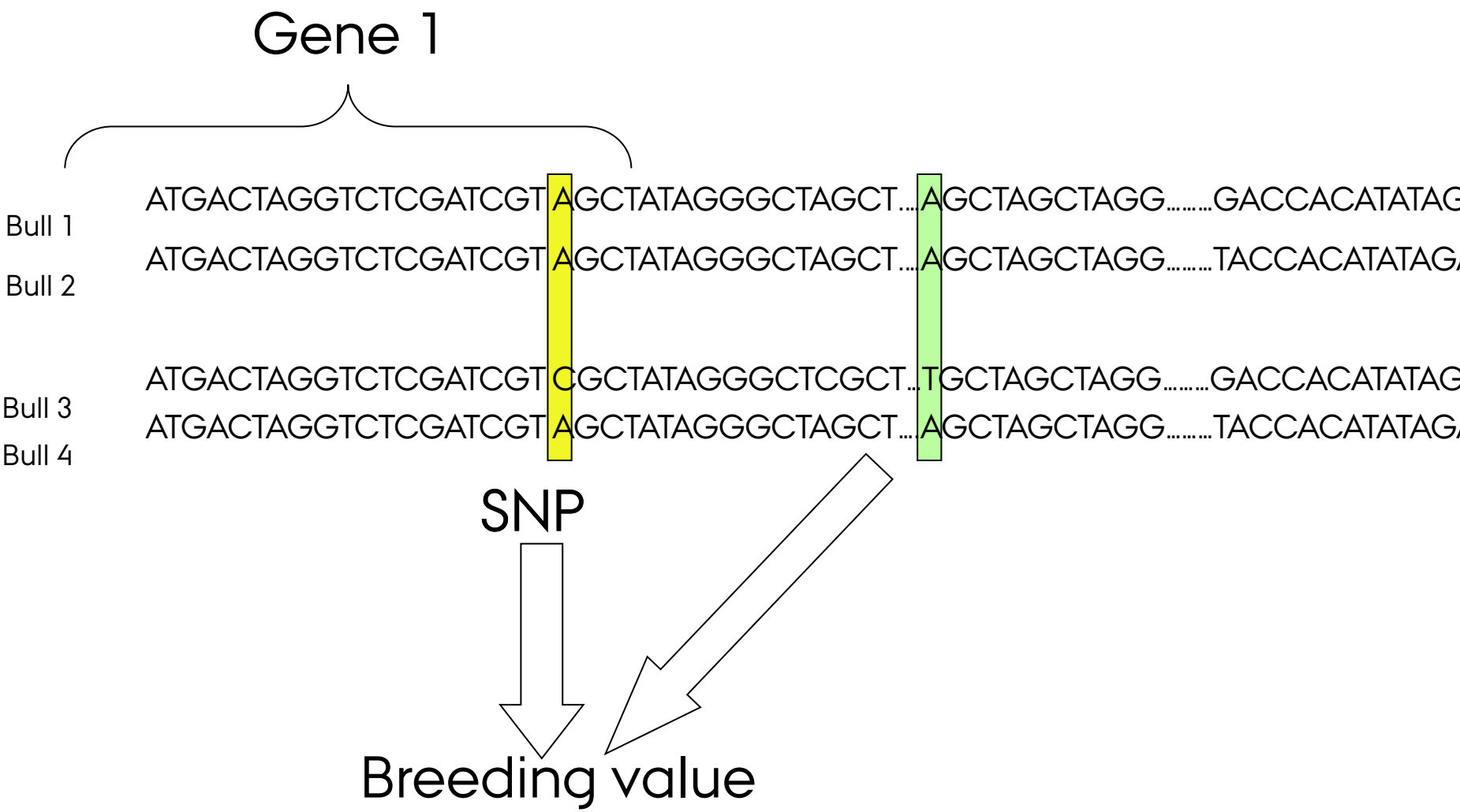
Quantitative Genetics and Genomics

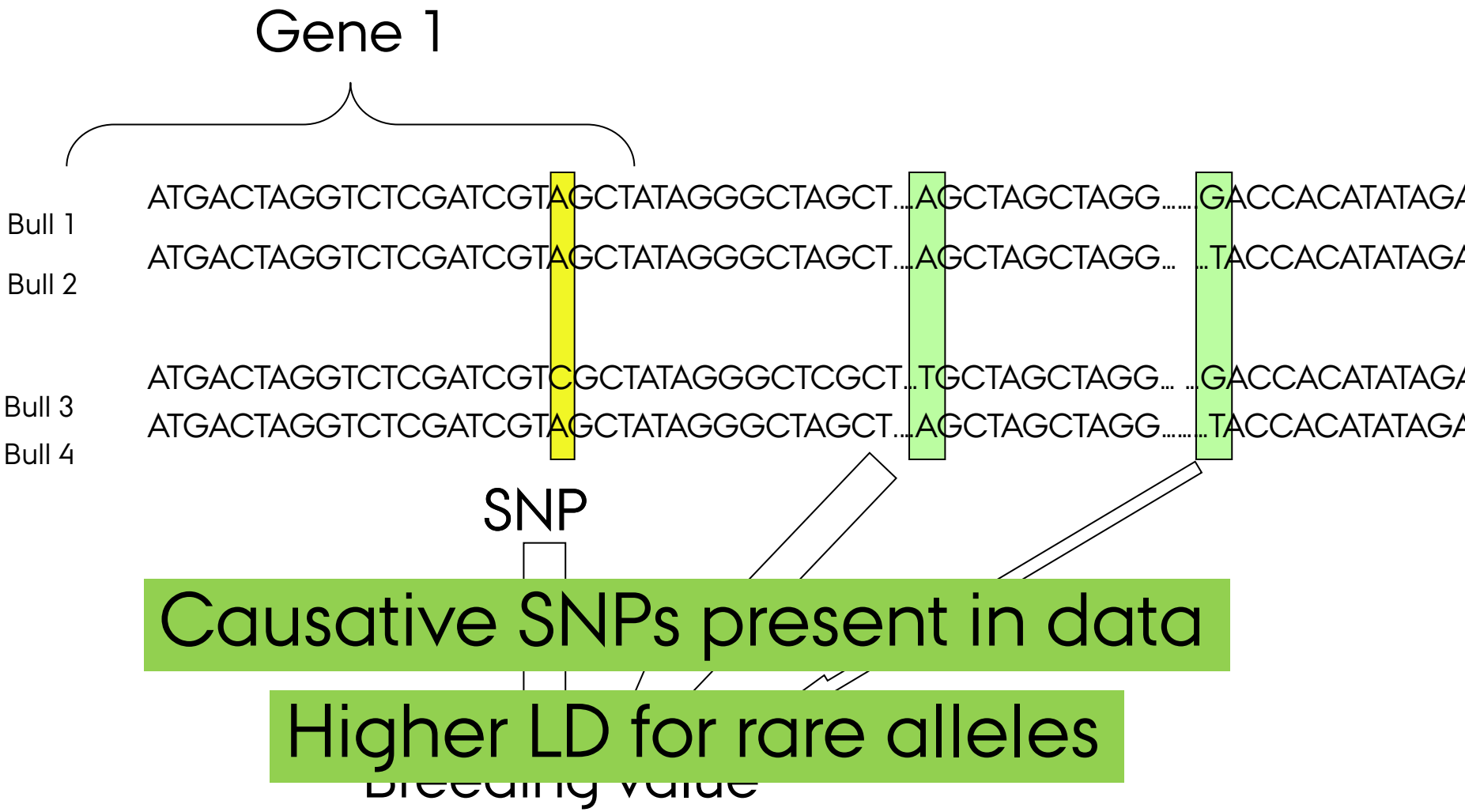



Gene = a sequence that affect a phenotype



- Millions of SNPs
- Most without effect
- We don't know which


Gene 1
Bull 1
ATGACTAGGTCTCGATCGT AGCTATAGGGCTAGCT.... AGCTAGCTAGG.....GACCACATATAG
Bull 2
ATGACTAGGTCTCGATCGT AGCTATAGGGCTAGCT.... AGCTAGCTAGG.....TACCACATATAG
Bull 3
ATGACTAGGTCTCGATCGT CGCTATAGGGCTCGCT... TGCTAGCTAGG.....GACCACATATAG
Bull 4
ATGACTAGGTCTCGATCGT AGCTATAGGGCTAGCT.... AGCTAGCTAGG.....TACCACATATAG
SNP
Breeding value

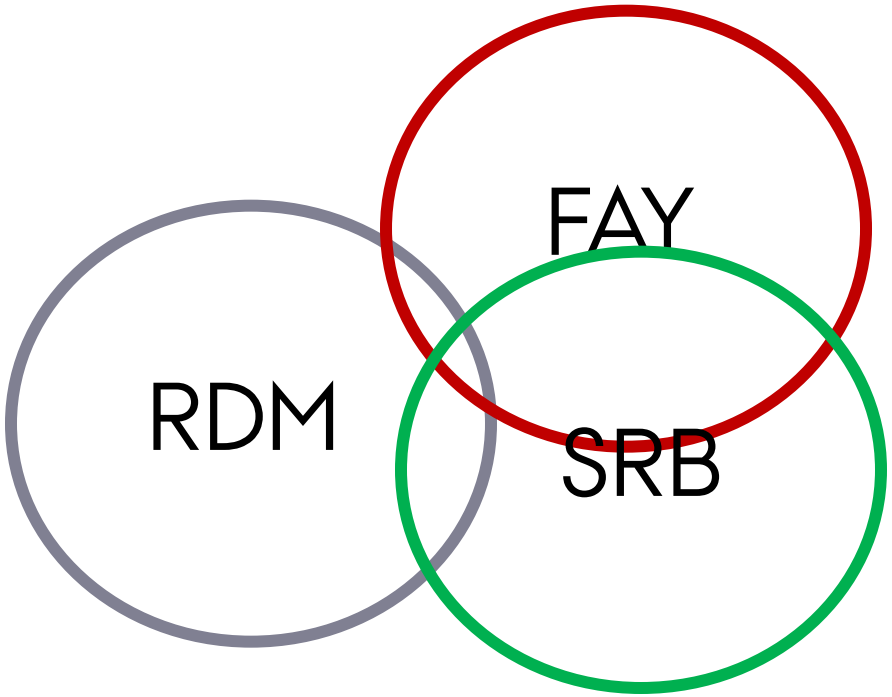

Gene 1
Bull 1
Bull 2
Bull 3
Bull 4
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT....AGCTAGCTAGG.....GACCACATATAGA
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT....AGCTAGCTAGG... ...TACCACATATAGA
ATGACTAGGTCTCGATCGTCGCTATAGGGCTCGCT...TGCTAGCTAGG... ...GACCACATATAGA
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT....AGCTAGCTAGG.....TACCACATATAGA
SNP
Causative SNPs present in data
Higher LD for rare alleles
Breeding value

The challenge

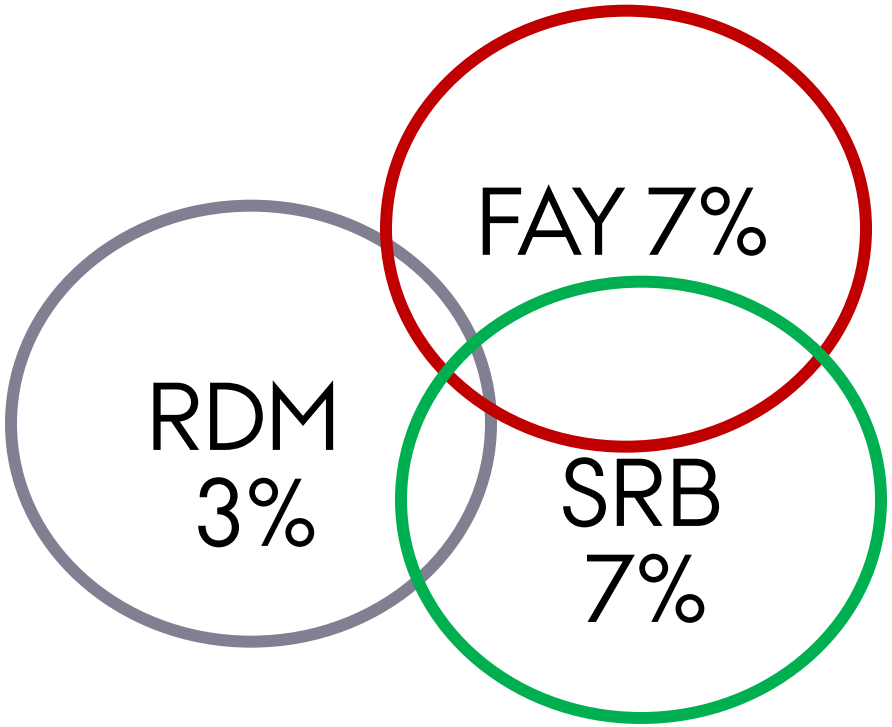
- Lower reliabilities for small breeds

- Increase reference population
  - Collaborate with other populations of same breed
  - Include thousands of genotyped cows
  - **Across breed predictions**

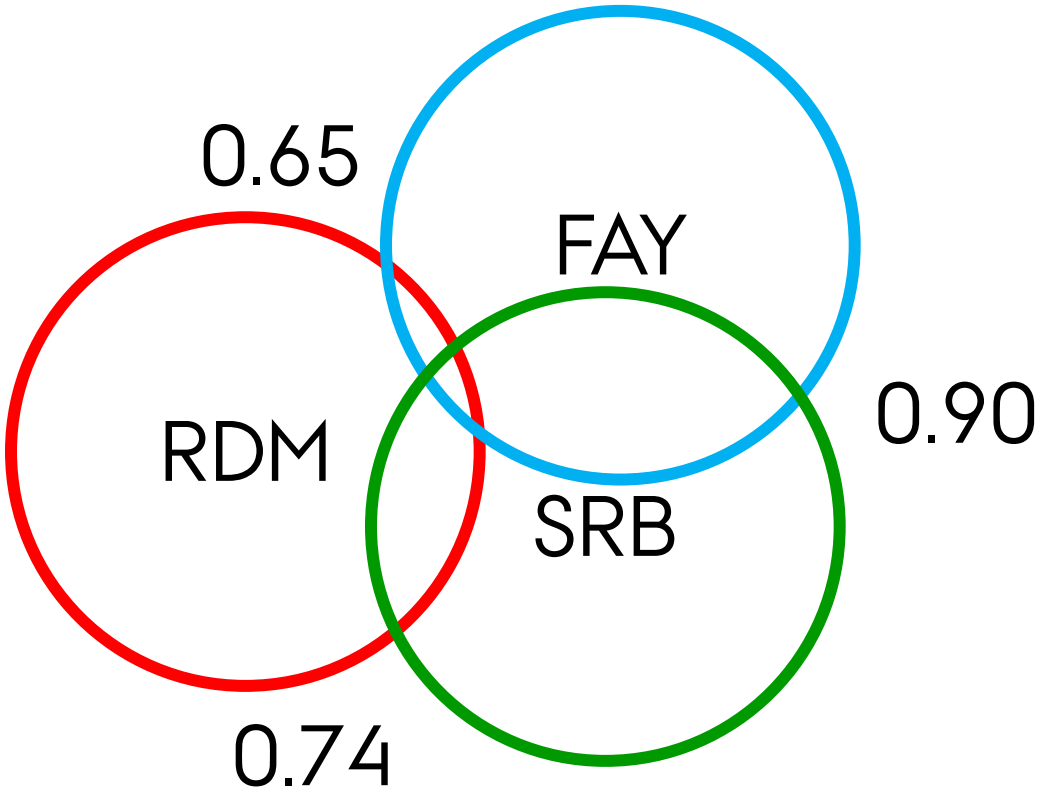
Across breeds predictions



Across breeds predictions



Across breeds predictions



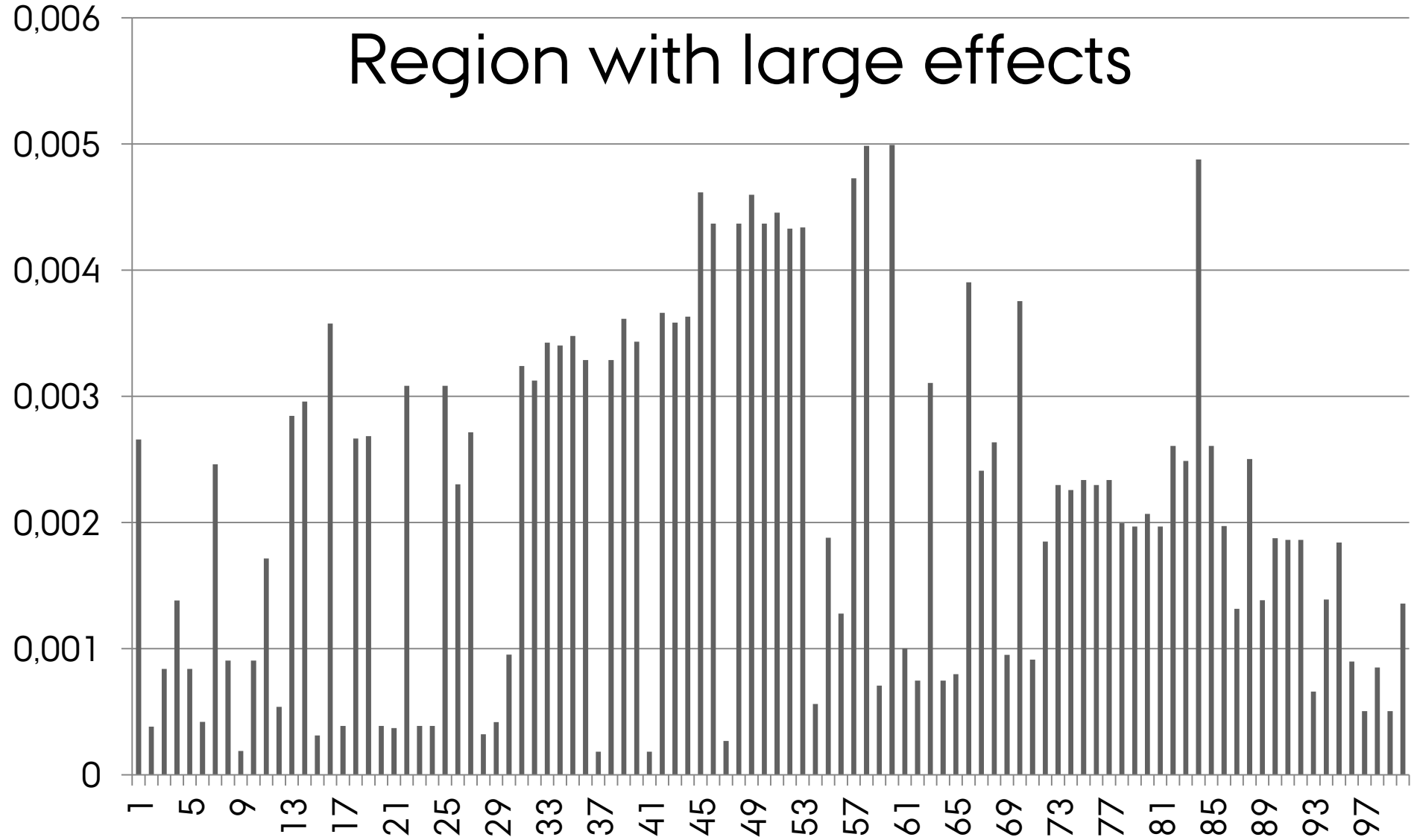
More markers → Higher LD → Higher accuracy

50K → 700K → Higher LD → 1.2% higher accuracy

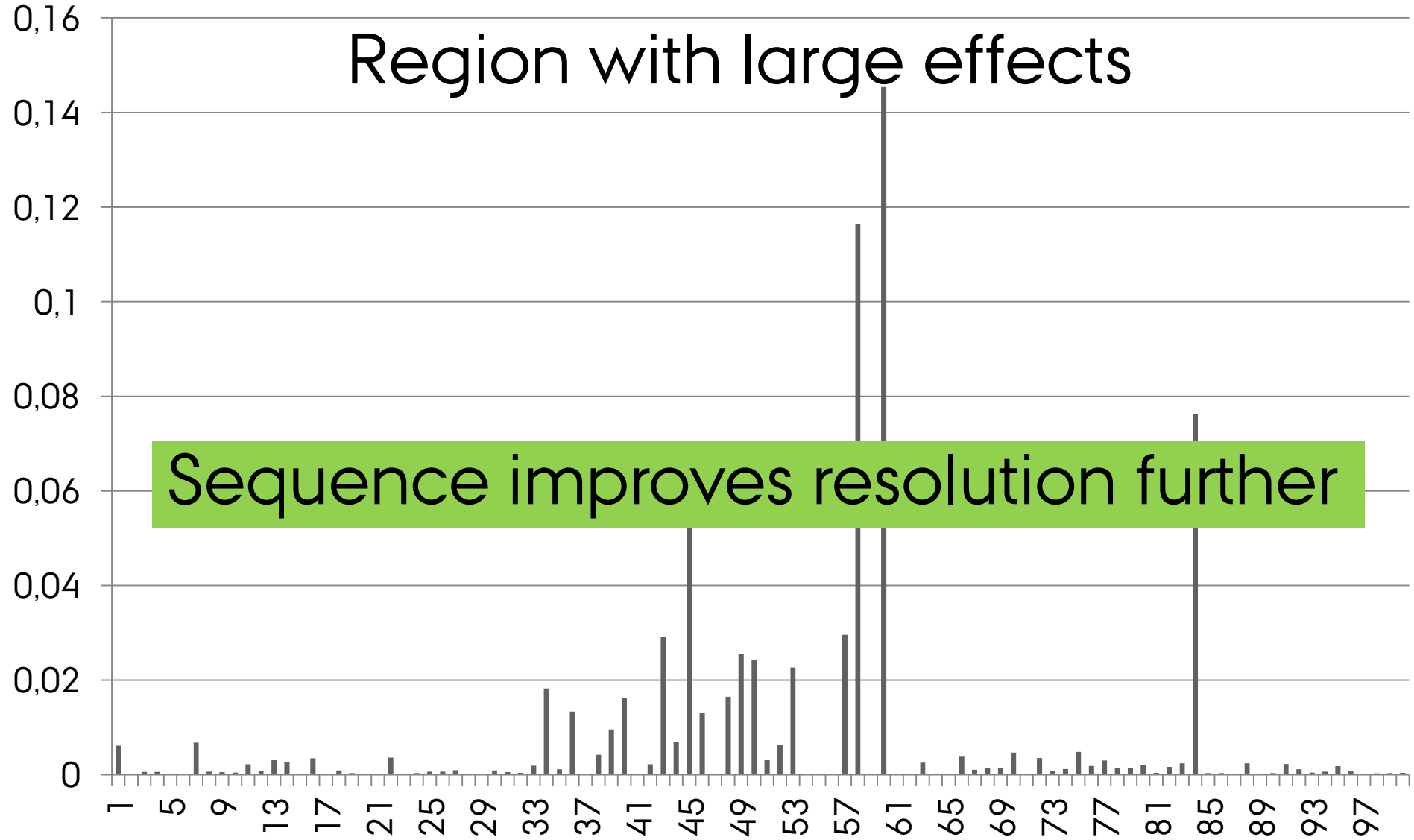
Small difference in models

- GBLUP
  - All SNPs assumed to explain equal amount of variance
- Bayesian variable selection models
  - Some SNP large effects
  - Most SNP small or no effect
- Models fit genomic values equally well
- Thousands of genes affecting mastitis??

GBLUP


Region with large effects
0,006
0,005
0,004
0,003
0,002
0,001
0
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
77
81
85
89
93
97

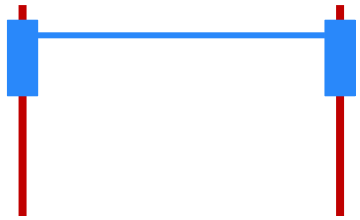
Bayesian variable selection model



Genomic selection across breeds

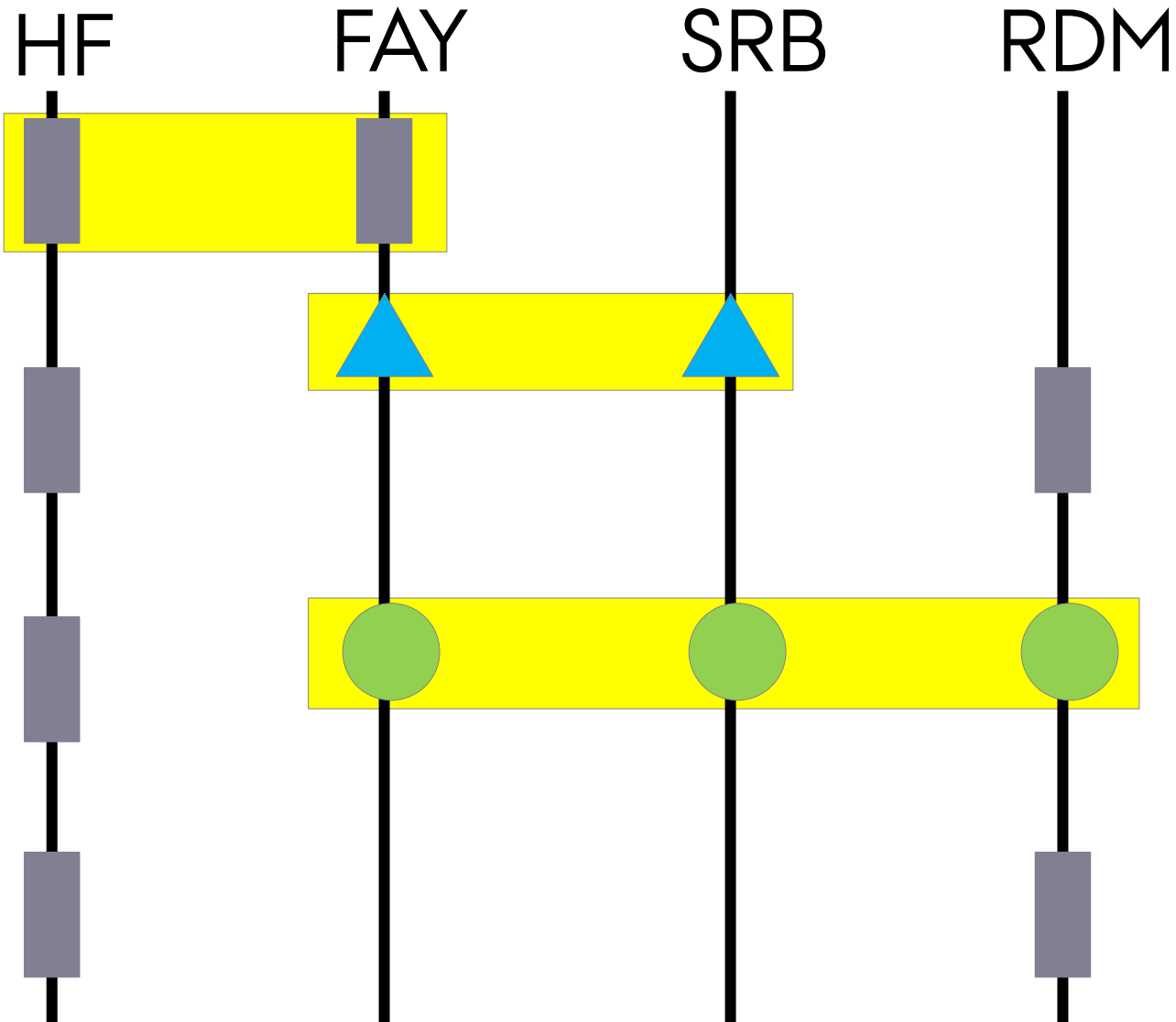
Breed 1 **Breed 2**

Genes are IBD



→ estimate effects across breeds

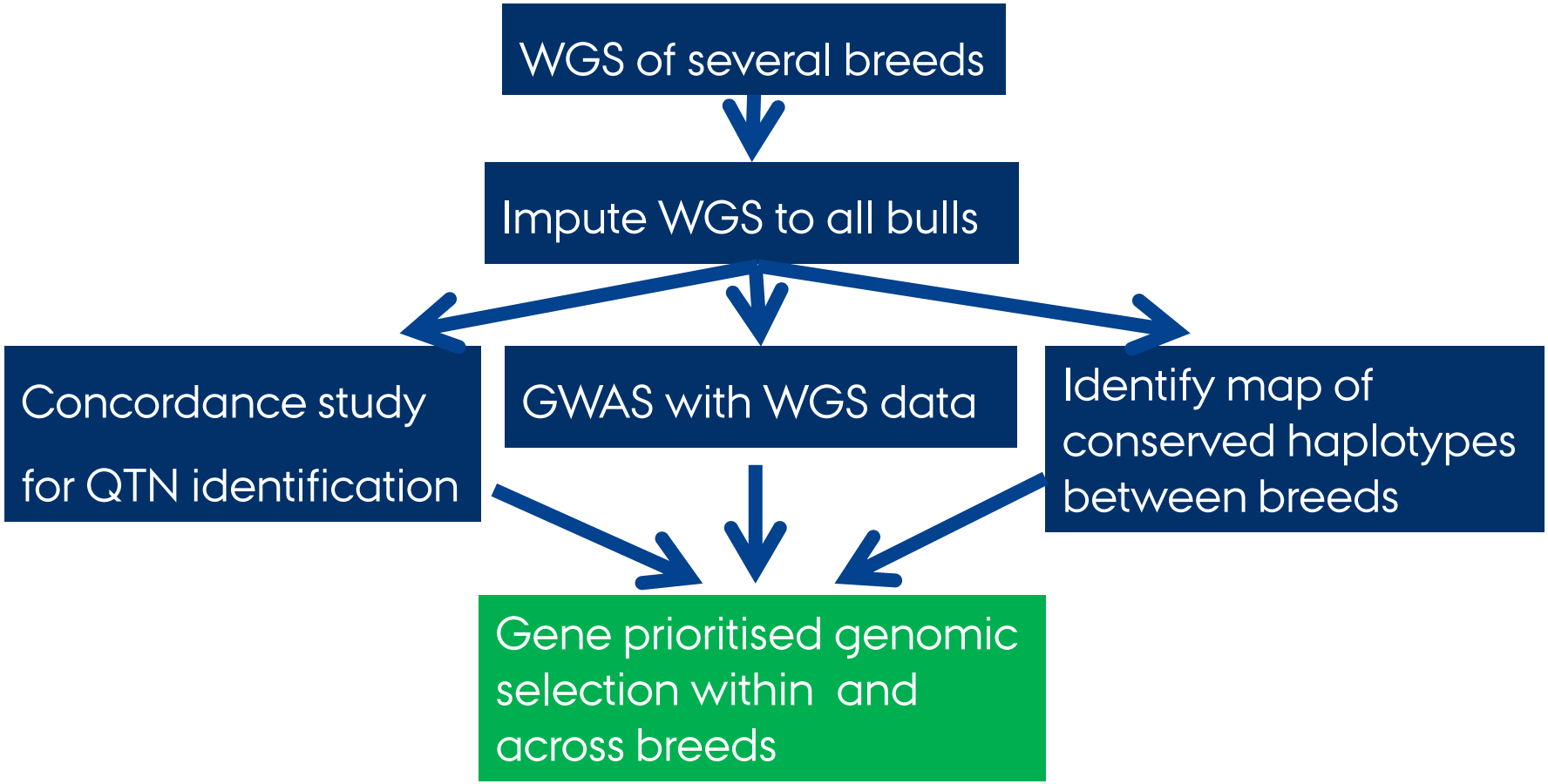
Across breeds predictions



Summary

- Need to use across breed information smarter
- Too many genes to understand traits in full
- Identify most important genes
  - Conserved haplotypes for across breed predictions
  - Prioritise regions in WGS genomic models
- Most efficient with sequence data

What is next?


WGS of several breeds
Impute WGS to all bulls
Concordance study for QTN identification
GWAS with WGS data
Identify map of conserved haplotypes between breeds
Gene prioritised genomic selection within and across breeds

Nordic Red WGS project

- Sequences combined from Sweden, Finland, Denmark

- Bioinformatic platform
  - Sequence allignment
  - SNP calling
  - WGS imputation

- Common research
  - Genetic architechture
  - Prediction models