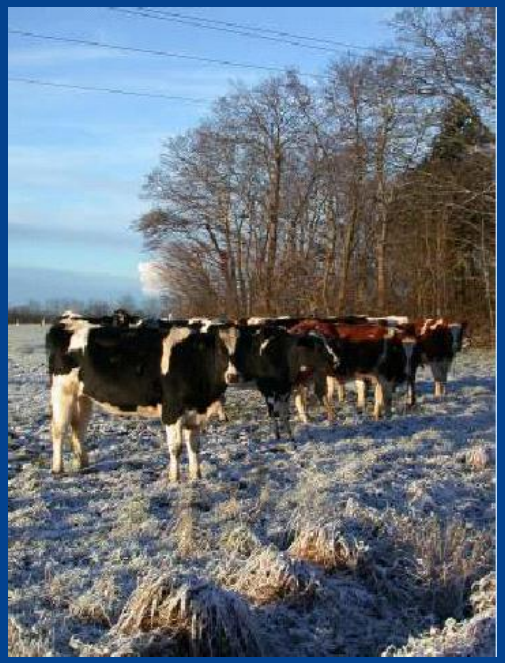
DET JORDBRUGSVIDENSKABELIGE FAKULTET

AARHUS UNIVERSITET

Genomic predictions using 3K, 50K, 800K or sequence data within and across populations – results and speculations

Mogens Sandø Lund
Rasmus Froberg Brøndum
Guosheng Su

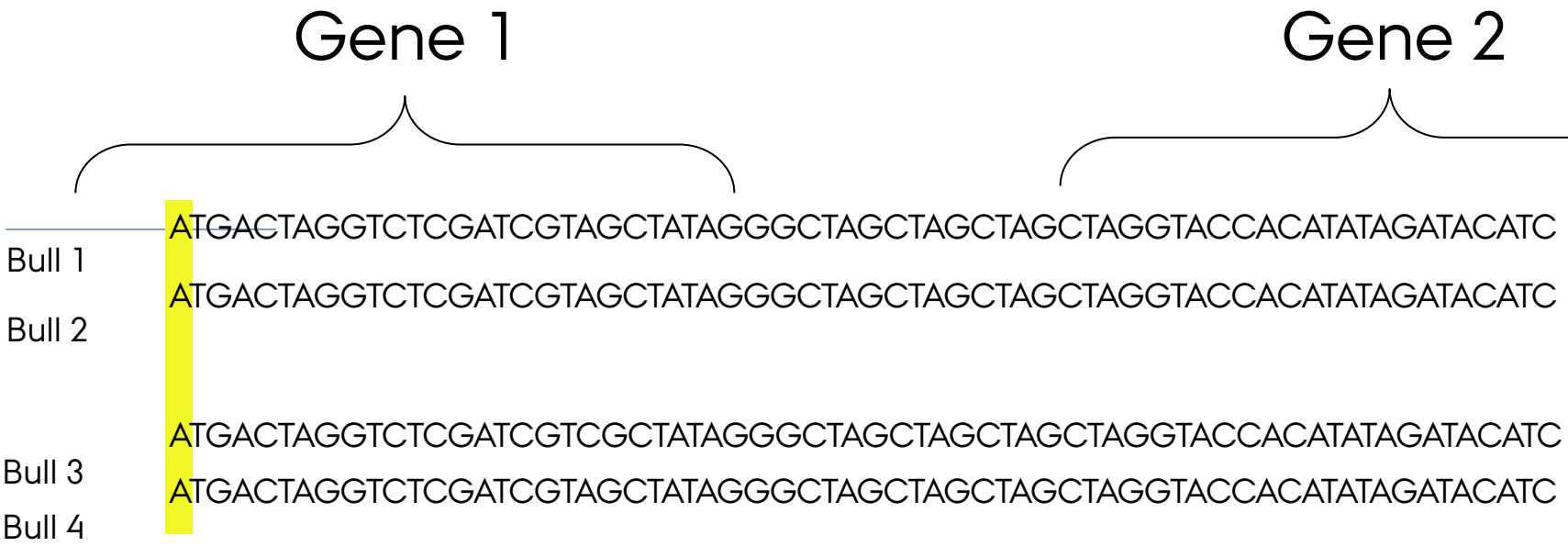


Genomic Selection

- Using dense SNPset på predict breeding values
  - Markers/haplotypes in close linkage disequilibrium with QTL
  - Many ”phenotypes” for each marker/haplotype to estimate effects accurately
- Critical factors
  - SNP density
  - Size of reference population (and effective population size)
  - Interplay between these factors in our breeds?
- Questions
  - Should we use 3K tests for candidates ?
  - should we double genotype bulls with 800K chip ?
  - Should we resequence bulls ?

Gene = a sequence (WORD) that affect a phenotype

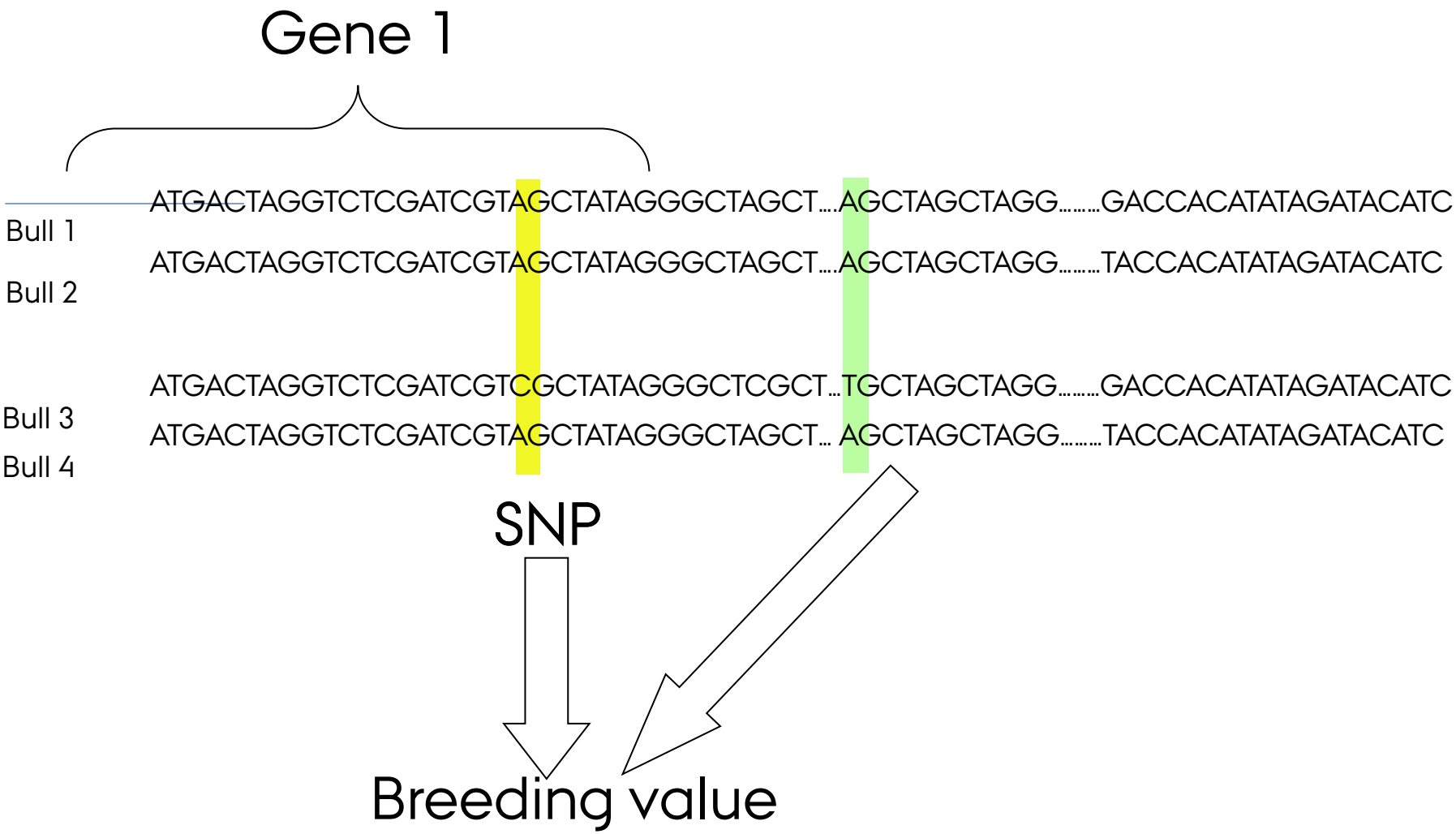
Aprox. 20.000 genes and 95% ”junk” DNA

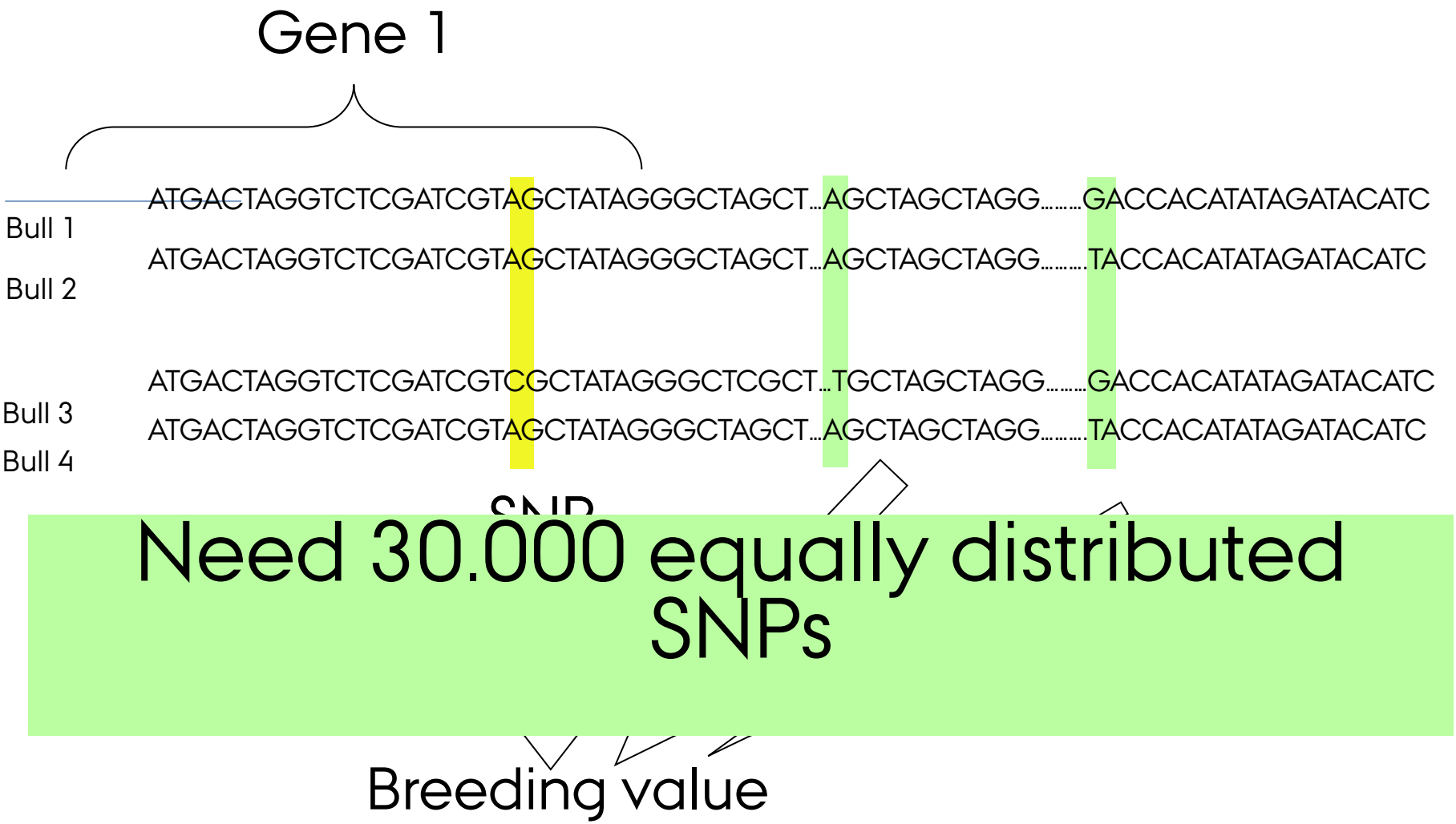


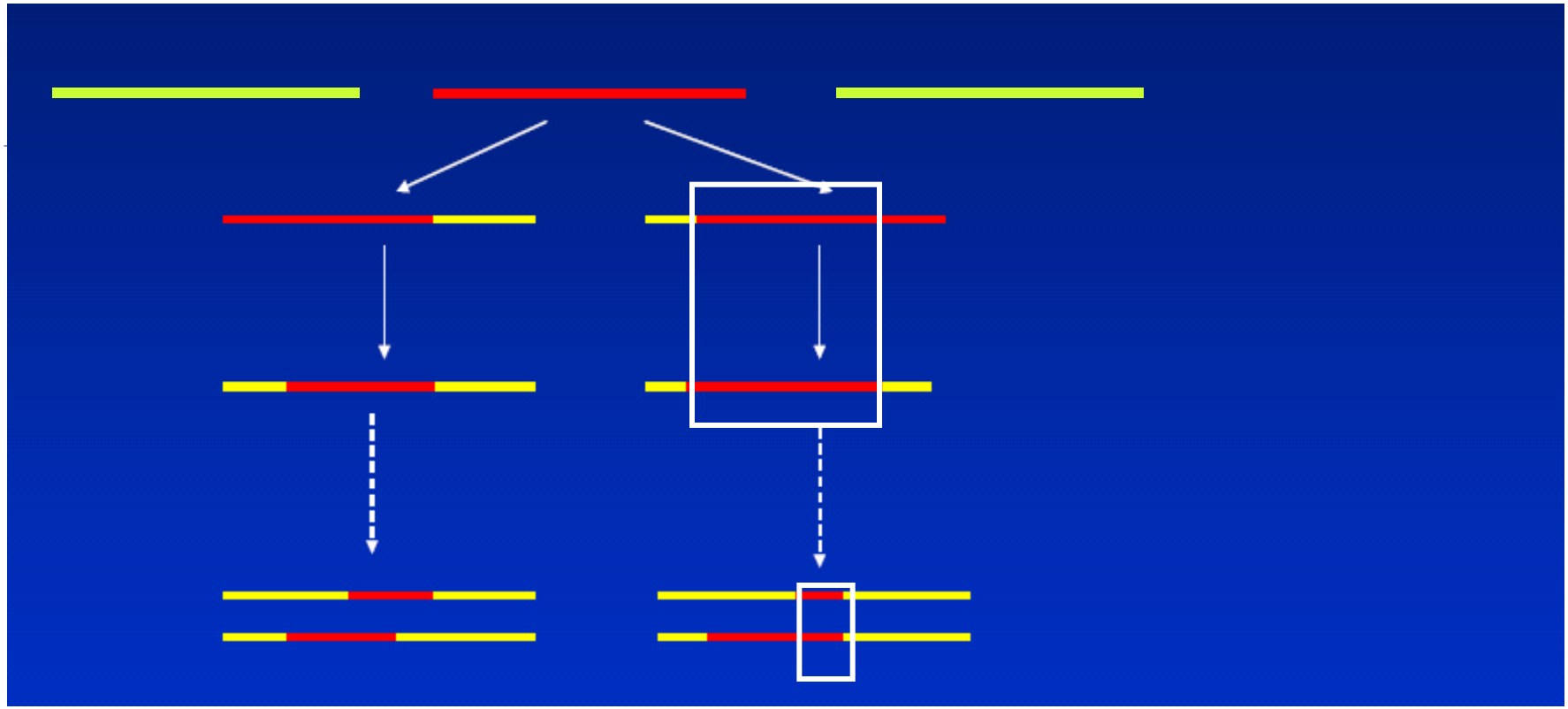
SNP

Breeding value

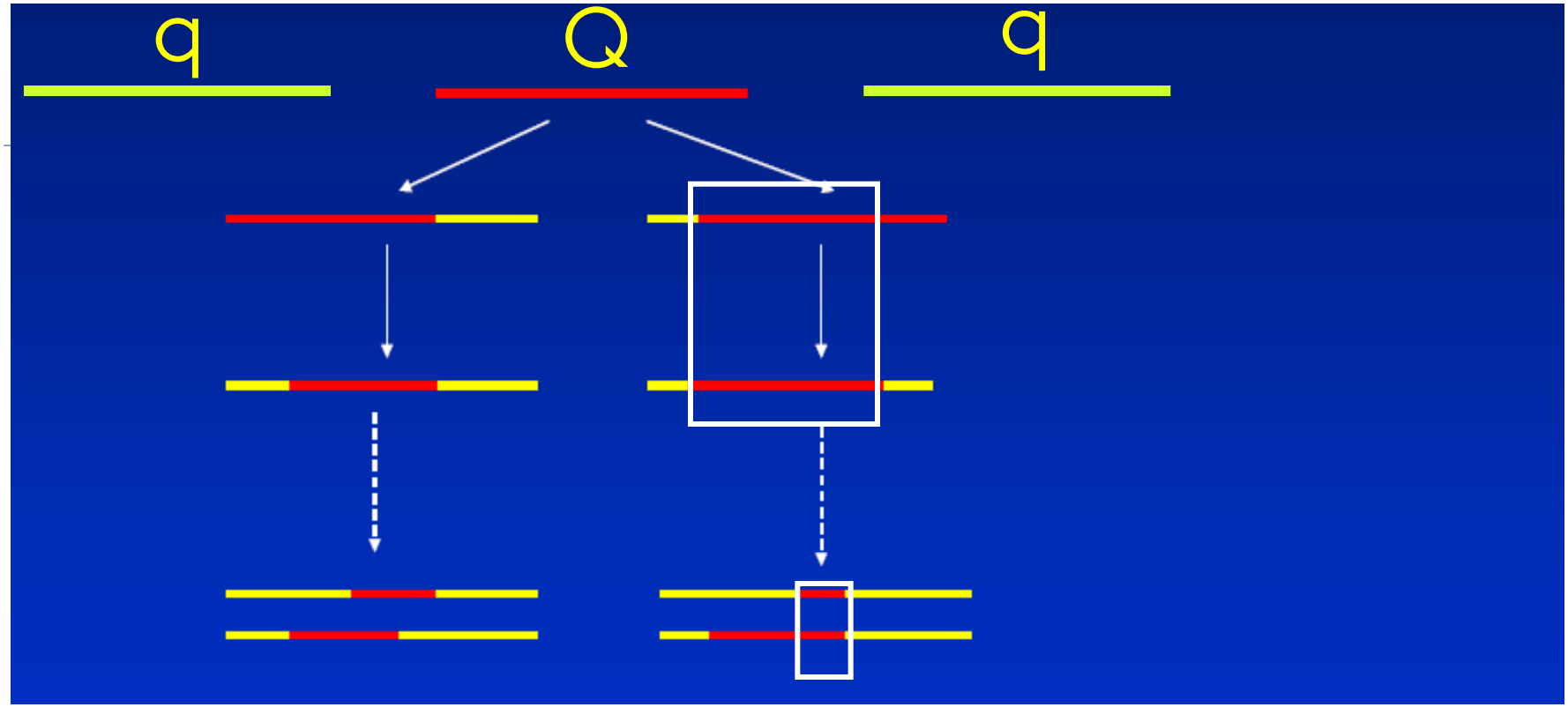
- Millions of SNPs
- Most without effect
- We don't know whi


Gene 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT….AGCTAGCTAGG……GACCACATATAGATACATC
Bull 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT….AGCTAGCTAGG……TACCACATATAGATACATC
Bull 2
ATGACTAGGTCTCGATCGTCGCTATAGGGCTCGCT…TGCTAGCTAGG……GACCACATATAGATACATC
Bull 3
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT… AGCTAGCTAGG……TACCACATATAGATACATC
Bull 4
SNP
Breeding value


Gene 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT…AGCTAGCTAGG……GACCACATATAGATACATC
Bull 1
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT…AGCTAGCTAGG……TACCACATATAGATACATC
Bull 2
ATGACTAGGTCTCGATCGTCGCTATAGGGCTCGCT…TGCTAGCTAGG……GACCACATATAGATACATC
Bull 3
ATGACTAGGTCTCGATCGTAGCTATAGGGCTAGCT…AGCTAGCTAGG……TACCACATATAGATACATC
Bull 4
SNP
Need 30.000 equally distributed SNPs
Breeding value

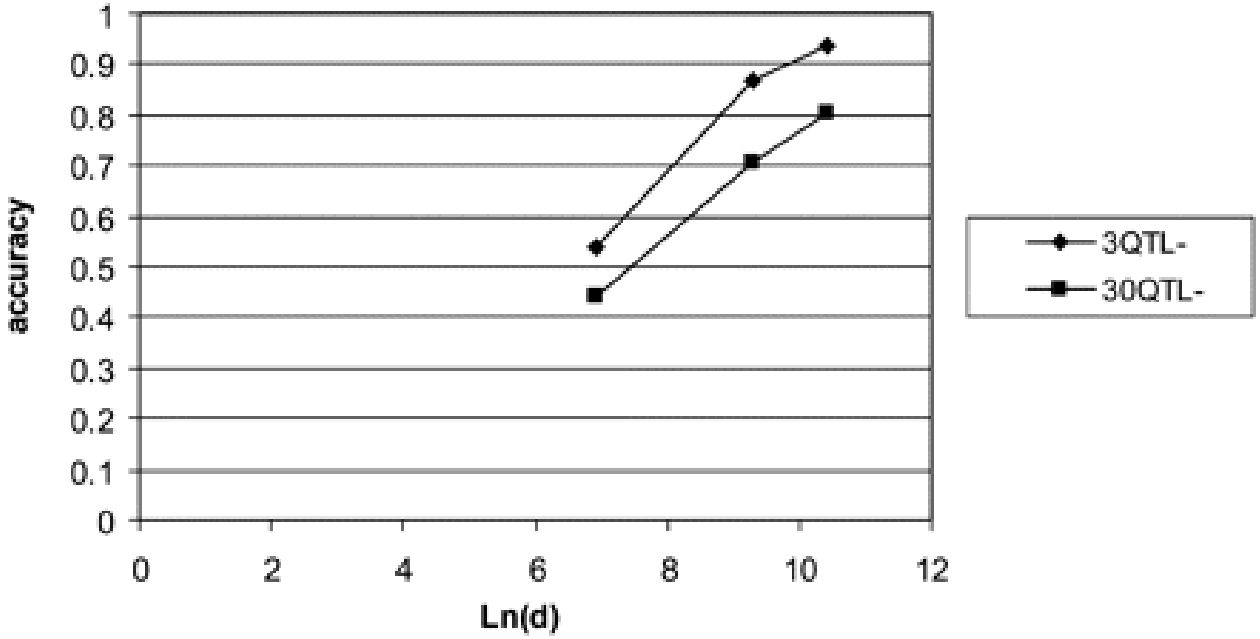


- Chunks of ancestral chromosome are conserved in the current population
- Size of chunks depend on number of generations to the common ancestor



- QTL alleles are in linkage disequilibrium with marker alleles
- Over many generations LD span a short region
- Denser markermap is needed

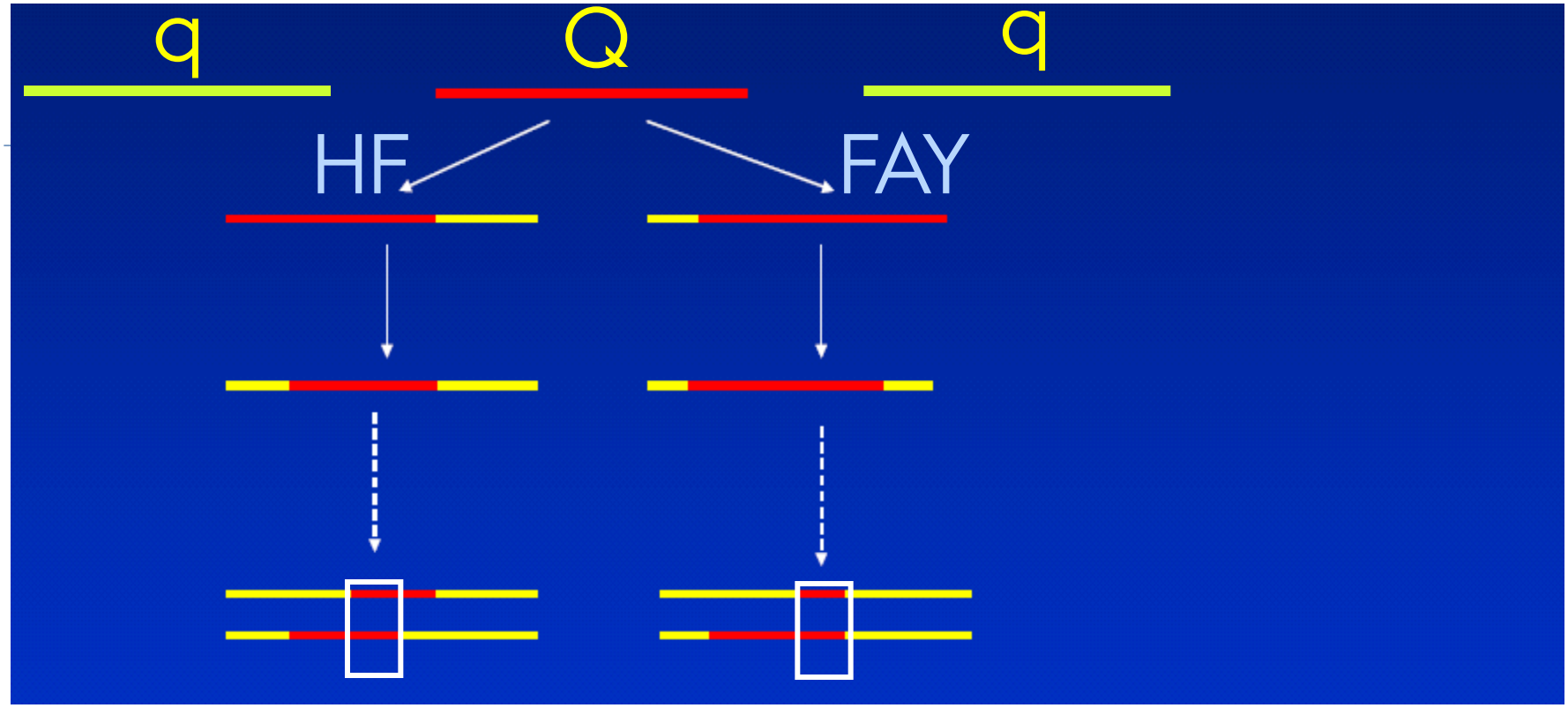
Accuracy of DGV depend on SNP-density



Theo Meuwissen, Genetics 2010

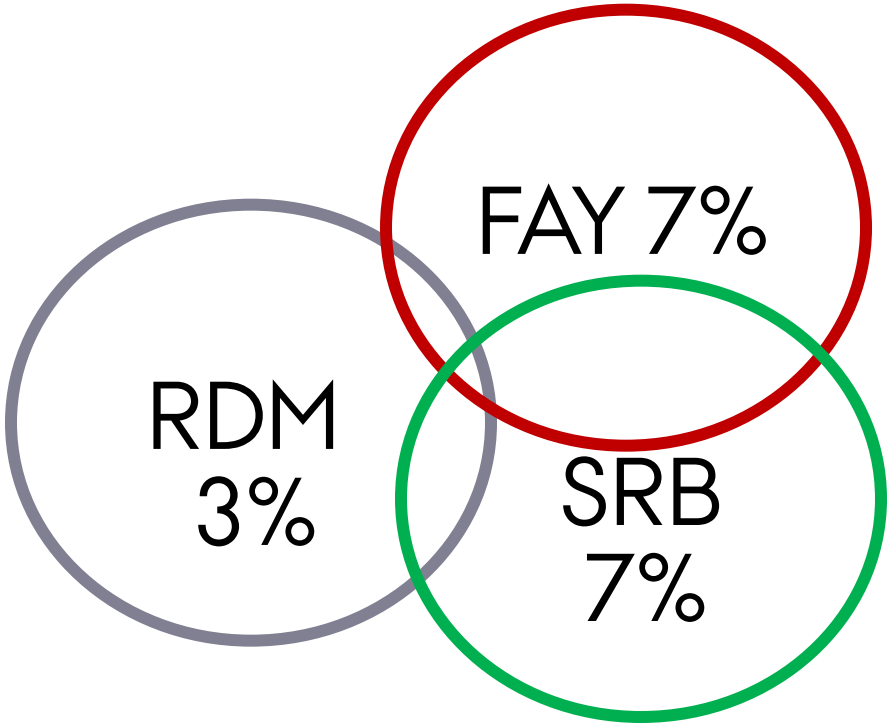
SNP density

- Accuracy increase dramatically with SNP density
- Why we do not expect as large increase for HF
  - Higher effective population size in simulations (denser map needed)
  - More QTL in real data with smaller effects
  - Larger reference population in HF (higher starting point)

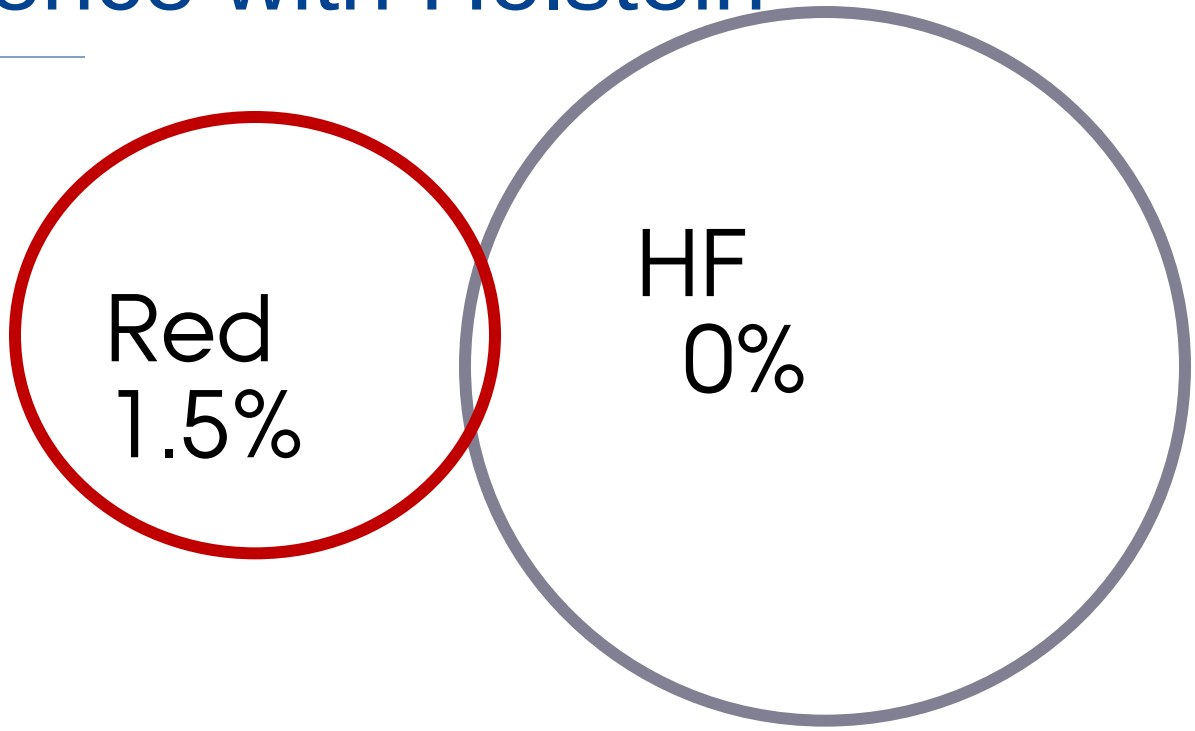


- Assuming the same QTL
- Assuming additive effects
- Prediction across breeds but very dense markermaps are needed

Increase in accuracy by common reference (SLU,AU, MTT)



Increase in accuracy by common reference with Holstein



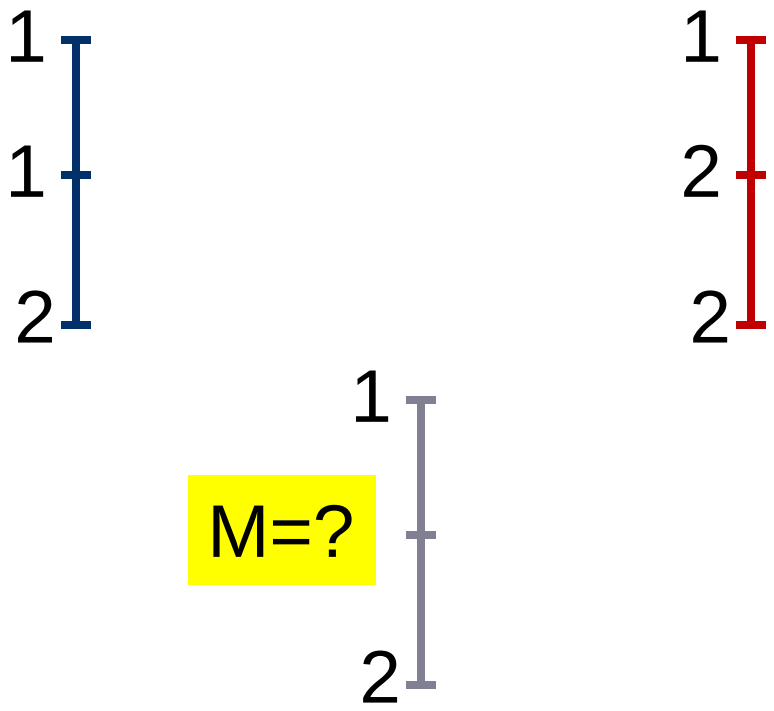
Summary

- Accuracy of genomic predictions depend on marker density
- Some increase is expected for Holstein with HD or sequence for all bulls
- With HD or sequence prediction across breeds is possible
- Huge potential by using large HF-reference in other breeds
- However – regenotyping and sequencing is expensive

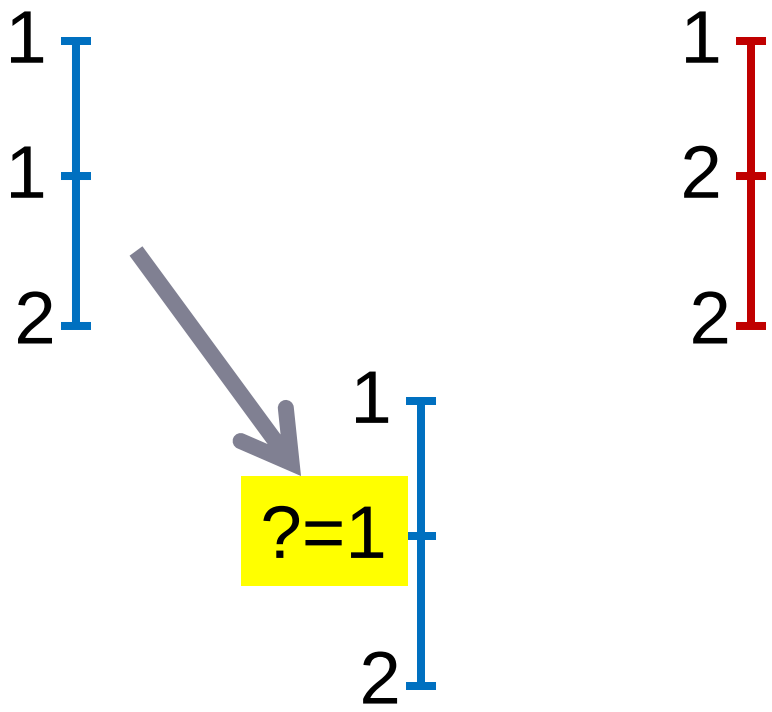
Imputation – ’poor mans genotyping’

- Predict unobserved SNPs
- Uses LD and LA information

Imputation using LD and LA information



Imputation using LD and LA information



Two imputation studies (Rasmus)

Holstein – prediction of candidates using 3K chip

RDC – Can we impute HD-chip from 50K?

Holstein – prediction of candidates using 3K chip

- Reference and test populations same as Eurogenomics study
- Impute for test population 3K to 50K (Illumina 3K panel)
- Investigate error rate
- Investigate impact on genomic predictions
- Investigate impact of Eurogenomics data

Methods - Holstein study

- Two imputation methods
  - LD Model (Beagle 2.1.3)
  - Combination of Linkage and LD information
- Error rate: percentage of falsely predicted alleles
- Genomic predictions done using SNP-BLUP with DRP as response

Holstein: data

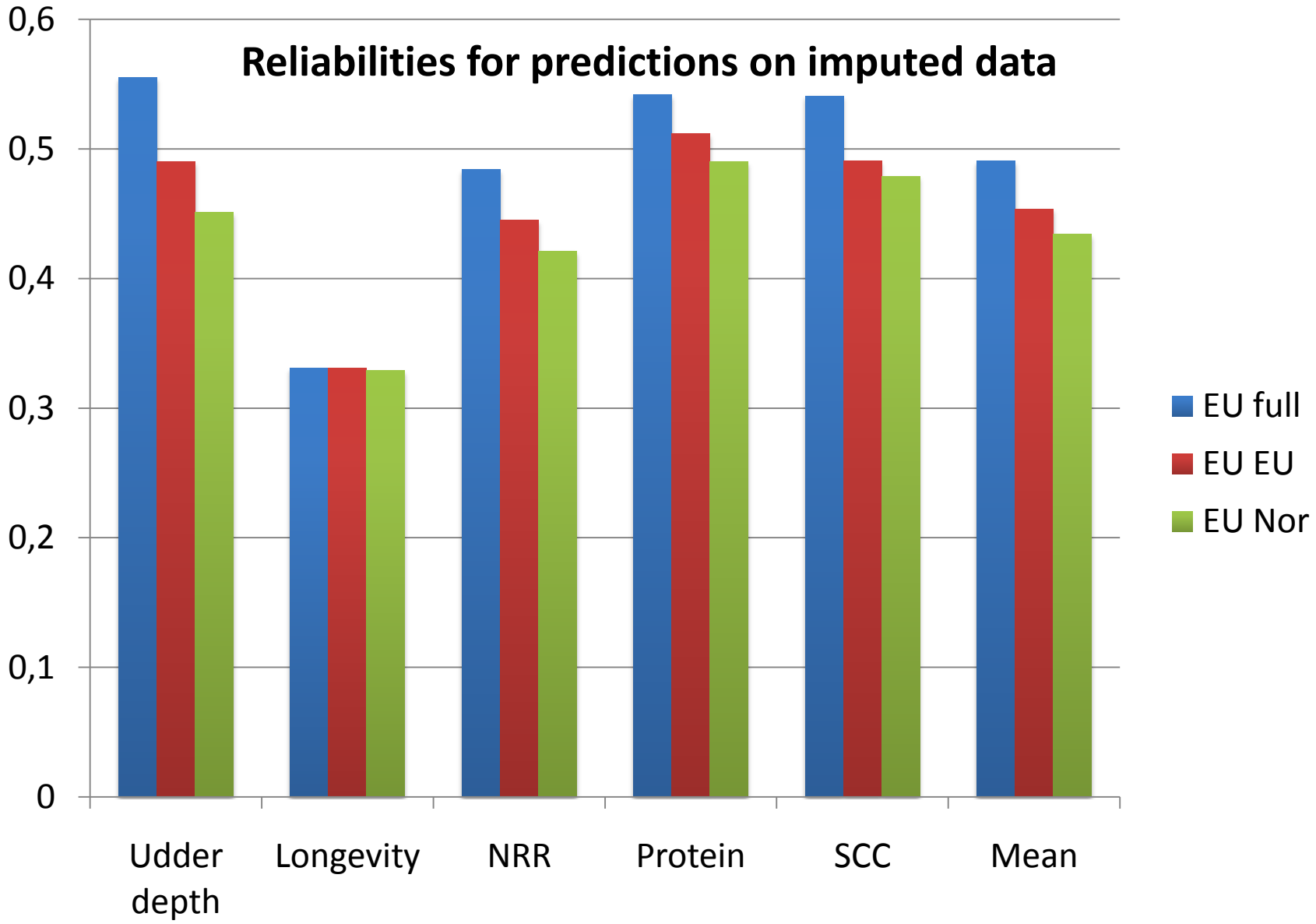
	Reference		Test
	Nordic	Eurogenomics	Both
No. of animals	3,058	10,880	1,086
No. of markers	38,545	38,545	2,285

Holstein: Results (Imputations with DAGPHASE)

	Nordic		Eurogenomics	
	N	Error rate (%)	N	Error rate (%)
All	1086	5.48	1086	3.97
Sire in ref	795	4.45	1039	3.83
Sire not in ref	291	8.31	47	7.04
Sire+MGS in ref	650	4.34	953	3.76

Holstein Results (genomic predictions)

Trait	N	EU-ref Full	Nordic-ref Full	EU-ref EU-imp	EU-ref Nordic-imp	Nordic-ref Nordic-imp
Depth	948	0.555	0.407	0.490	0.451	0.362
Longevity	528	0.331	0.234	0.331	0.329	0.234
NRR	942	0.484	0.434	0.445	0.421	0.402
Protein	942	0.542	0.399	0.512	0.490	0.364
SCC	947	0.541	0.401	0.491	0.479	0.392
Mean	N/A	0.491	0.375	0.454	0.434	0.351


Reliabilities for predictions on imputed data
0,6
0,5
0,4
0,3
0,2
0,1
0
Udder depth
Longevity
NRR
Protein
SCC
Mean
EU full
EU EU
EU Nor

Conclusions Holstein

- Imputation error rate of 4-5 %
- Lower error rates with Eurogenomics data
- 3-6% points loss in mean reliability for genomic predictions
- Around 20% loss of genomic predictive ability
- Most of the increase from using 3K to 50K data is captured by imputation
  - Using 3K decreases reliability 12%
  - Using 3K imputed -> 50K decreases reliability by 2 %

RDC – Can we impute HD-chip from 50K?

- Mimik by imputing 15K to 50K
- 15K chosen by MAF and distance
- Delete remaining 35K SNPs
- Impute 35K missing SNPs and check with observed
- Investigate different strategies to choose bulls
- Investigate two imputation methods

RDC: data

	RDCDNK	RDCSWE	RDCFIN
Top 500	74	154	242
Top 1000	153	263	554
Top 1500	517	333	620
Top 2000	1,004	344	622
Top 1000 new	300	350	350
Top 1500 new	500	500	500
1000 random	300	350	350

- Top Animals ranked by genetic impact
- 6494 animals with genotypes

Ref. Size	DK	SWE	FIN	All
500	13,2	6,5	6,3	7,8
1000	11	5,3	4,2	6
1500	6,3	4,6	4,1	4,6
2000	3,1	4	3,9	3,9

Error rates from Beagle imputations

Reference	DK	SWE	FIN	All
Top 1000	3.2	1.6	1.3	1.8
Top 1000 new	2.7	1.5	1.5	1.7
Top 1500 new	2.2	1.4	1.3	1.5

- Switch to pure LD model
- Demands more computation time

Conclusions RDC

- Improvement of error rates by using only LD model, at the cost of longer computation times
- Imputation using LD model is better in Beagle than in DAGPHASE
- **Recommend to type 1000 bulls with HD chip**

Summary

› Reliability of GEBVs increase with SNP density (sequence best)

› Large increase expected for across breeds predictions

› Large potential to use HF-reference in other breeds

› Imputation can capture most of the increase

› Imputation error rate higher with low density chips

› Regenotyping bulls of all breeds with HD-chip is essential

› Resequencing bulls is highly desirable