



# Integrating additional SNPs selected from **whole genome sequence** for genomic prediction in **DK and US Jersey**

A. Liu<sup>12</sup>, M. S. Lund<sup>1</sup>, D. Boichard<sup>3</sup>, S. Fritz<sup>3</sup>, E. Karaman<sup>1</sup>, Y. Wang<sup>2</sup>, G. Su<sup>1</sup>

<sup>1</sup>Department of Molecular Biology and Genetics, Aarhus University, 8830, Tjele, Denmark

<sup>2</sup>College of Animal Science and Technology, China Agricultural University, 100193, Beijing, China

<sup>3</sup>GABI, INRA, AGROParisTech, Université Paris Saclay, 78350, Jouy-en-Josas, France

# Background

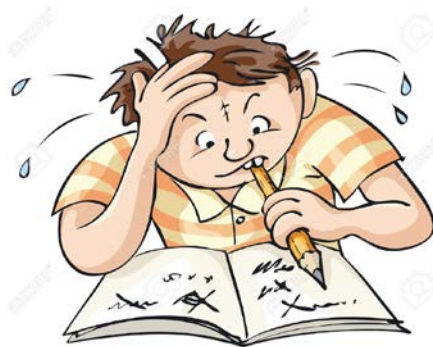


High throughput  
genotyping



Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density  $\rightarrow$  better LD  $\rightarrow$  higher reliability

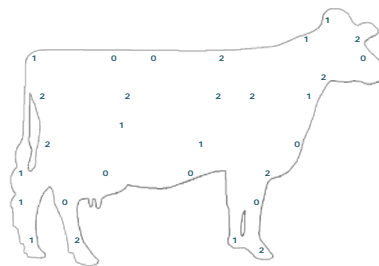


**Increase SNP  
density!**

# Background



High throughput  
genotyping



Low density  
(7 K)



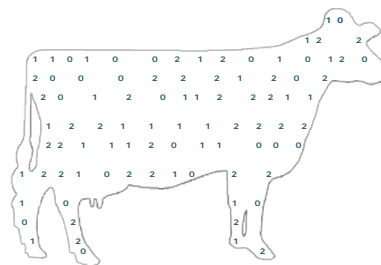
Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density -> better LD -> higher reliability

# Background



High throughput  
genotyping



Medium density  
(54 K)



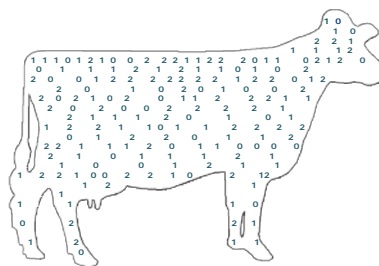
Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density -> better LD -> higher reliability

# Background



High throughput  
genotyping



High density  
(777 K)



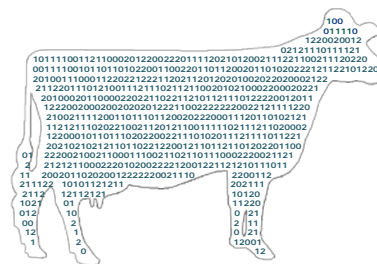
Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density -> better LD -> higher reliability

# Background



High throughput  
genotyping



WGS (~26,700 K)



Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density -> better LD -> higher reliability

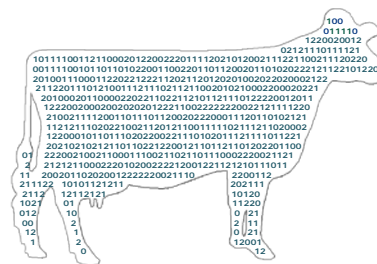
## Results from real data:

- HD  $\approx$  54K (Su et al., 2012)
- Imputed WGS  $\approx$  HD (Van Binsbergen et al., 2015)

# Background



High throughput  
genotyping



WGS (~26,700 K)



Reliable  
genomic prediction

**Hypothesis1** : Higher SNP density -> better LD -> higher reliability

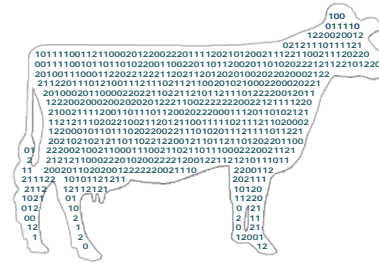
## Explanations:

- Only variants very close to causative mutations can improve reliability (van den Berg et al., 2016)
- non-causative mutations bring noise

# Background



High throughput genotyping



WGS (~26,700 K)



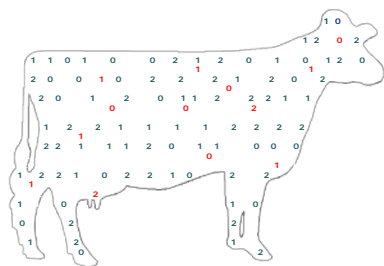
Reliable genomic prediction

Hypothesis1 : Higher SNP density -> better LD -> higher reliability

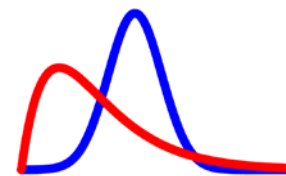
**Hypothesis2:** Causative mutations or SNPs in high LD with causative mutations -> high reliability



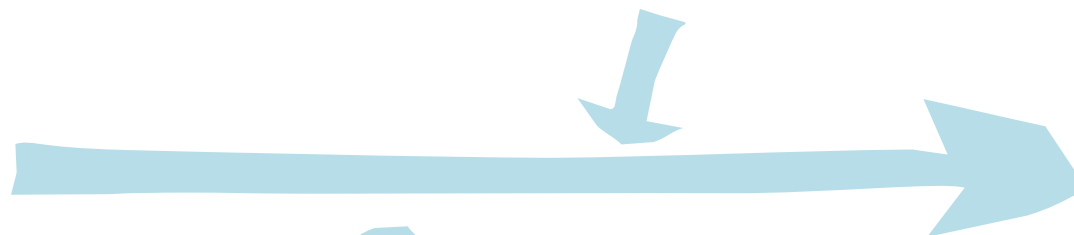
# Background



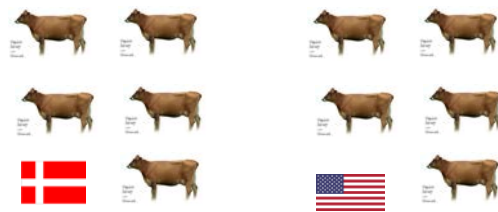
**54 K +**  
**SNPs selected from WGS**



**Models**

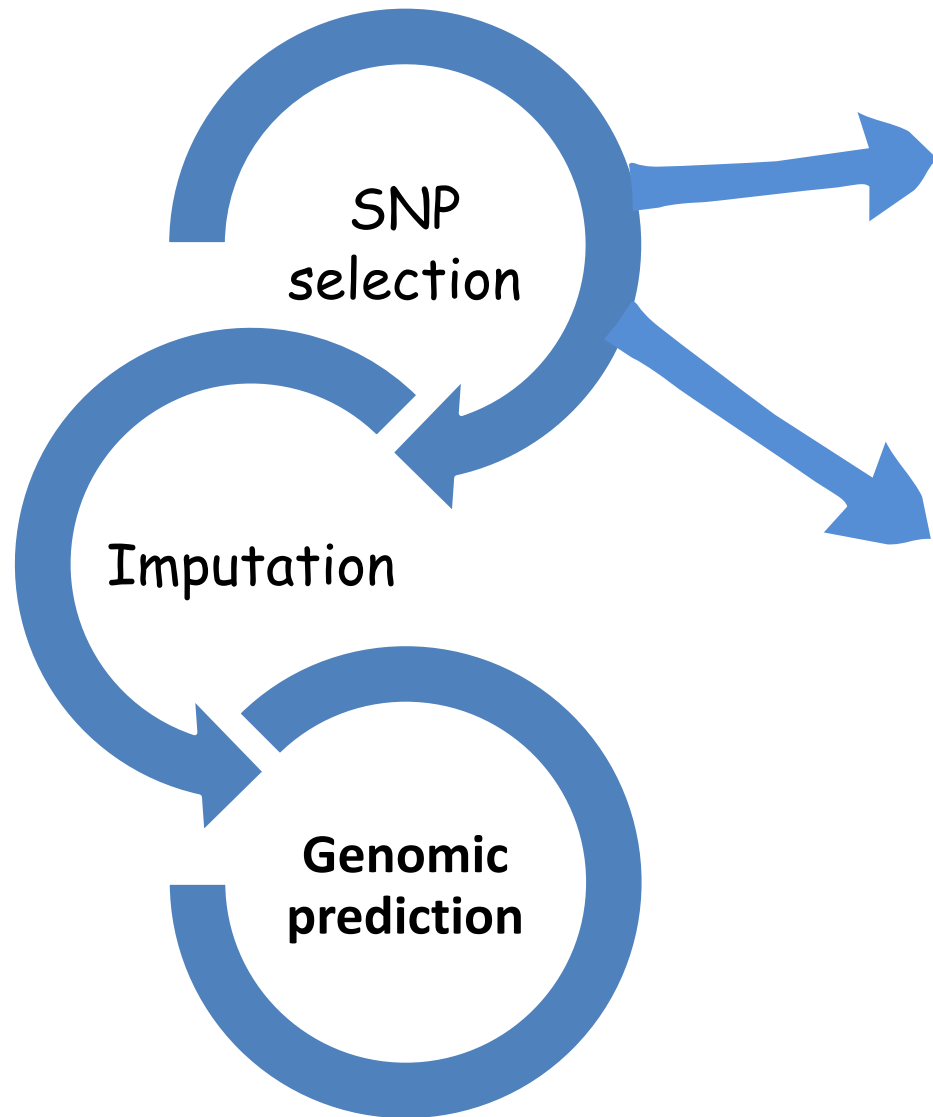


**Joint reference**



**Reliable**  
**genomic prediction**

# Workflow



## **NOR SNPs** (Brondum et al., 2015):

- peaks of QTL from Nordic Holsteins, Nordic Red and Danish Jersey

## **FR SNPs** (Boichard et al., unpublished):

- literature
- a strong variant effect predictor annotation (e.g. non-synonymous substitution)
- regulatory regions of genes
- peaks of QTL
- breakpoints of structural SNPs

# Imputation



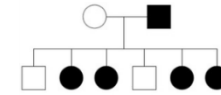
## Animal

- DK bulls: 1,310
- DK cows: 30,811
- US bulls: 1,153

## Genotype



- 54K chip
- standard LD chip
- customized LD chip
  - standard LD chip
  - NOR SNPs
  - FRA SNP



## Pedigree

- 6,102 males
- 66,466 females

**Two-step imputation**  
(*Fimpute*)

(1)

54K

(2)

54K

+ NOR SNPs + FRA SNPs

# Prediction: SNP

## Quality control

- Minor allele frequency > 0.01
- Imputation accuracy
  - correlation > 0.8
  - concordance rate > 0.8

SNPs	No. of SNPs	
	before	after
54K	40,452	33,166
NOR SNPs	1,754	1,270
FRA SNPs	4,325	2,427

# Prediction: Genomic feature model

## ➤ One-component GBLUP model

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



54K/ 54K+selected SNPs

Scenarios	One
54K	54K
54K_NOR	54K+NOR
54K_FRA	54K+FRA
54K_NOR_FRA	54K+NOR+FRA

## ➤ Two-component GBLUP model

$$y = \mathbf{1}\mu + X_{54K}g_{54K} + X_{WGS}g_{WGS} + e$$



54K






Selected SNPs

Scenarios	One	Two
54K_NOR	54K	NOR
54K_FRA	54K	FRA
54K_NOR_FRA	54K	NOR+FRA

# Prediction: Reference and validation

## ➤ Reference

- DK: 1,029 DK bulls born before 2005 
- Joint DK-US: 1,029 DK bulls born before 2005   
1,153 US bulls 

## ➤ Validation

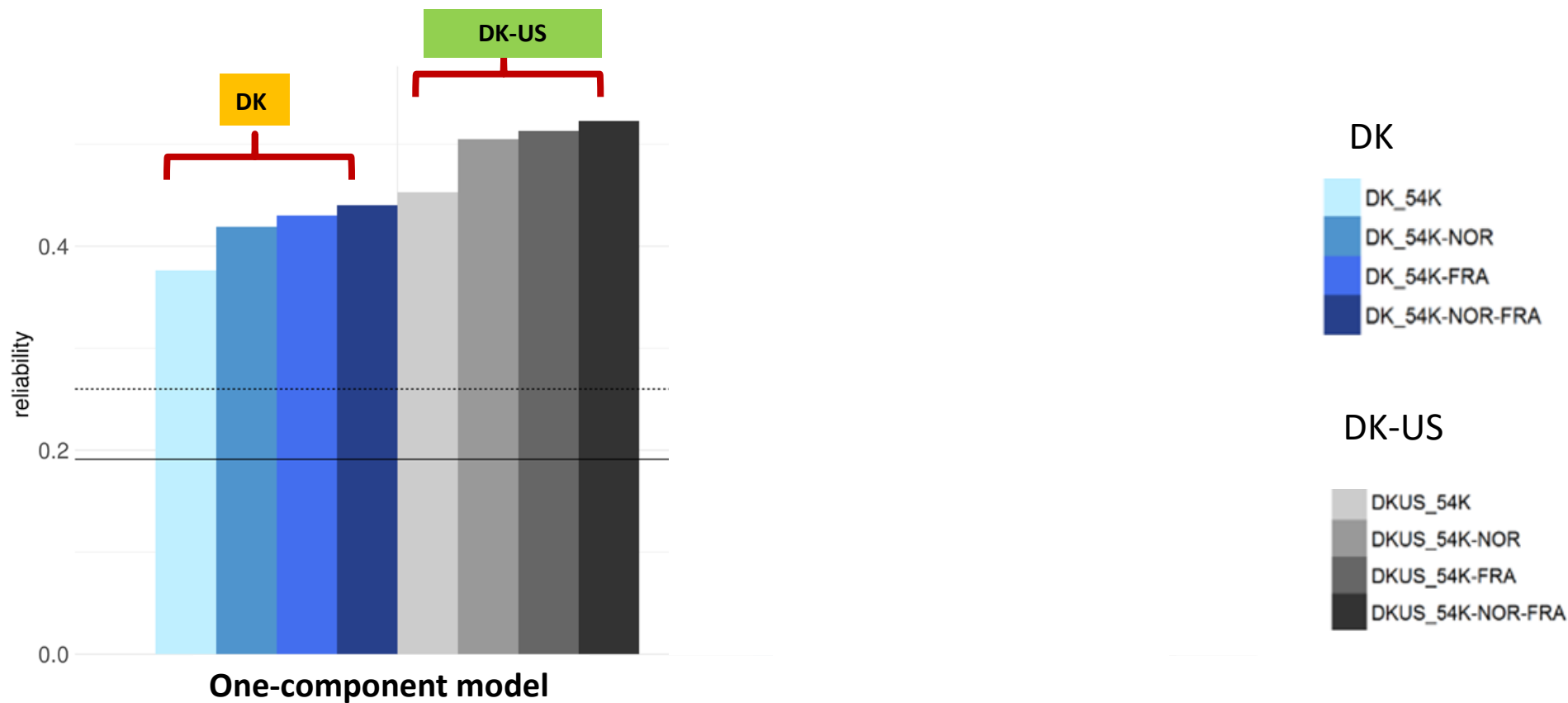
- 281 DK bulls born after 2005 

# Results-Reliability of milk



- Inclusion of additional WGS SNPs improved reliability (11.4-17.0%)
- Inclusion of all additional WGS SNPs achieved highest reliabilities

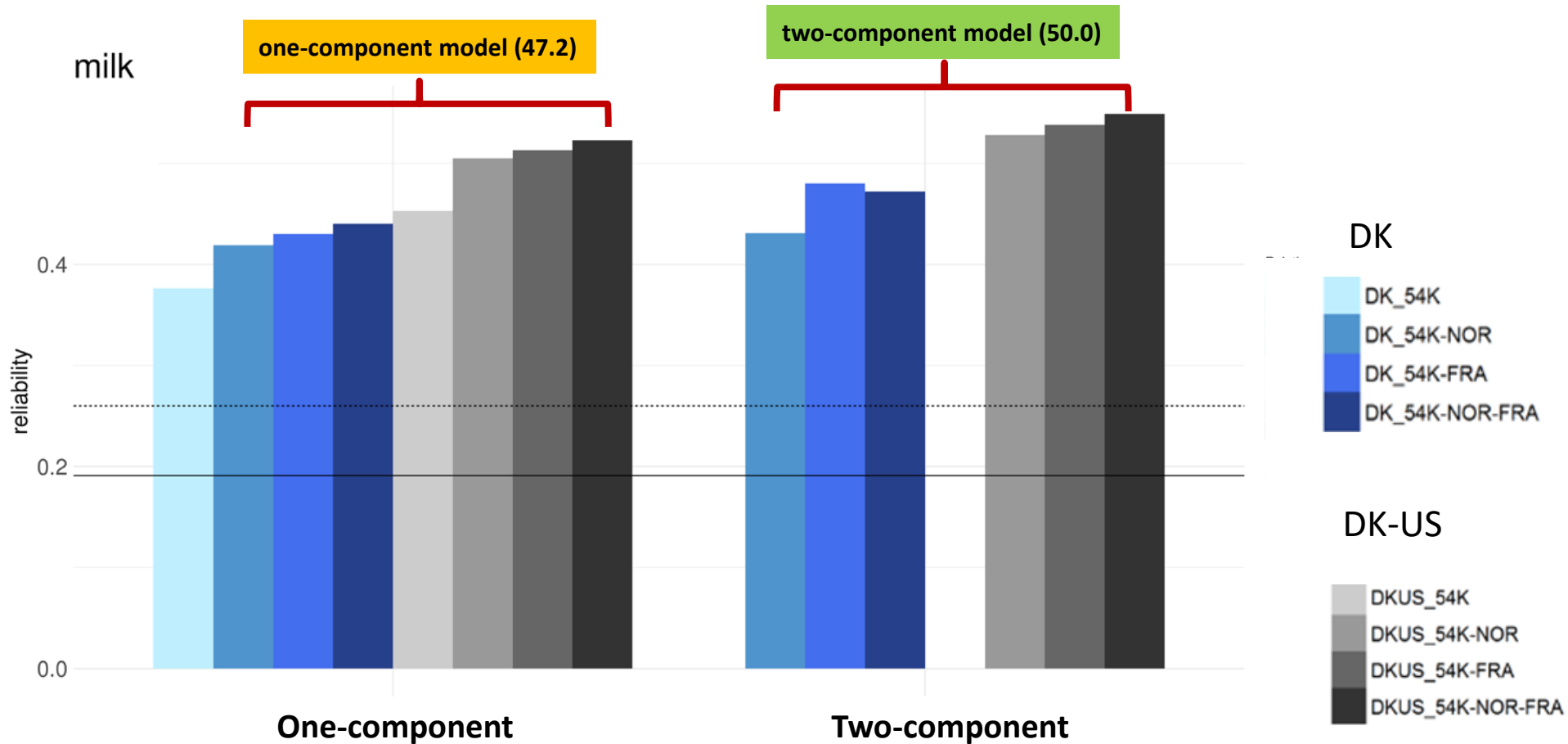
# Results-Reliability of milk



- A joint DK-US reference better than a DK reference for all traits (20%)
- Additional WGS SNPs improved reliabilities of a joint reference (11.5-13.6%)



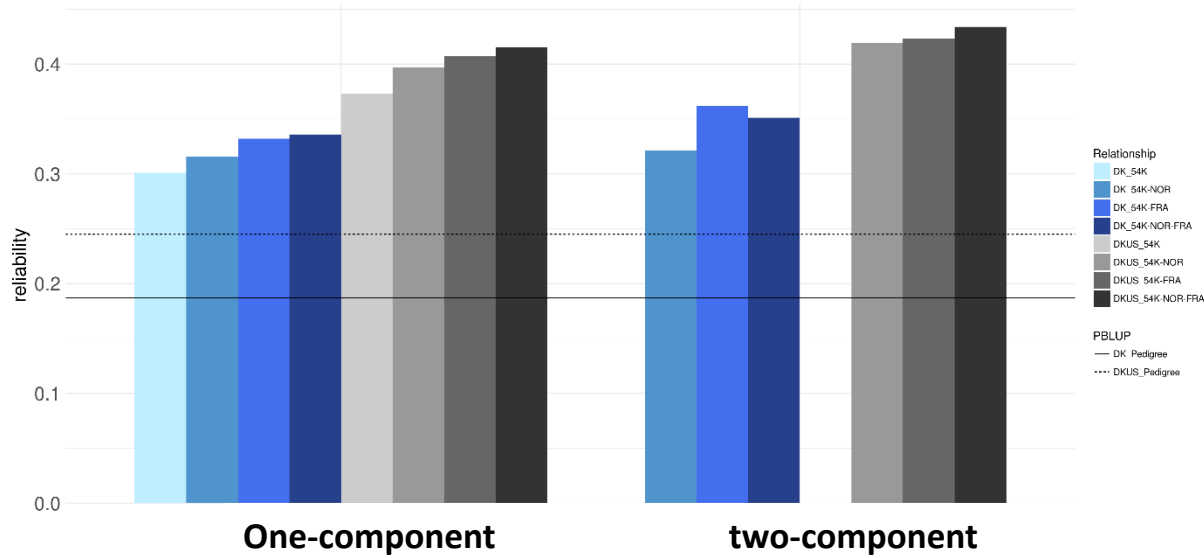
# Results-Reliability of milk



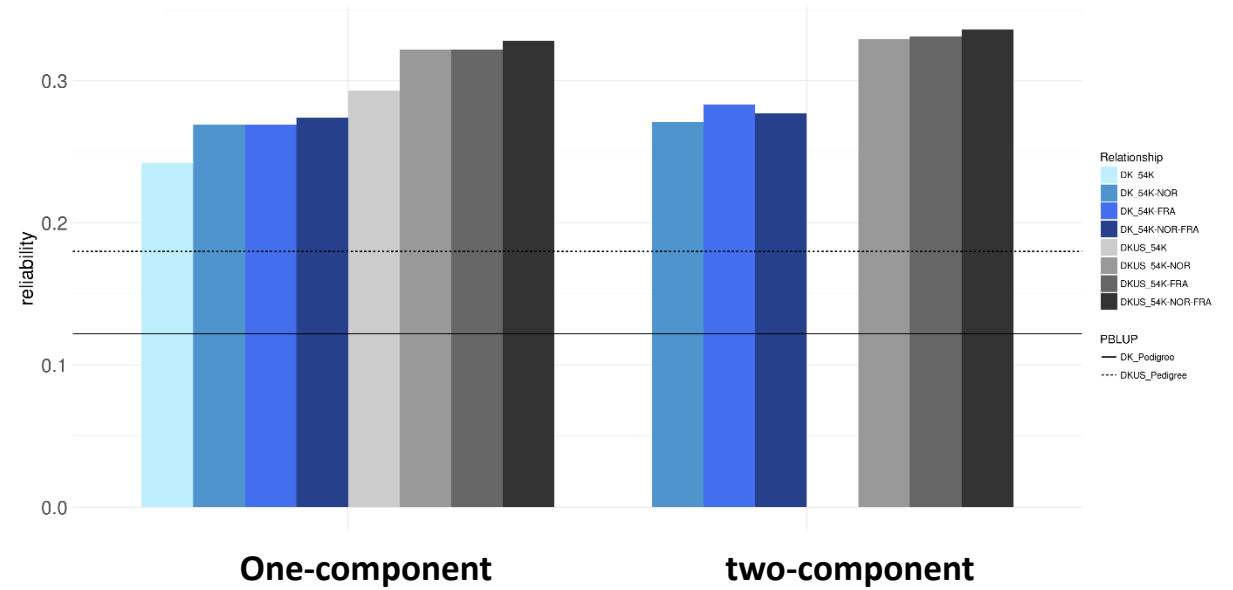
➤ A two-component model improved reliabilities (4.8%)

# Results-Reliability of protein & fat

protein



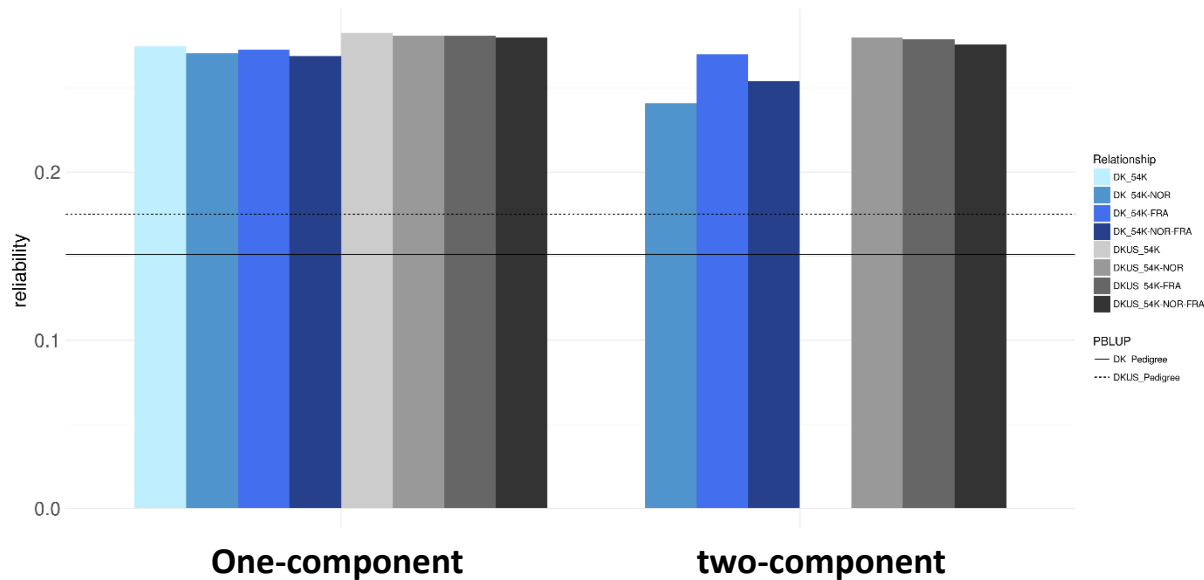
fat



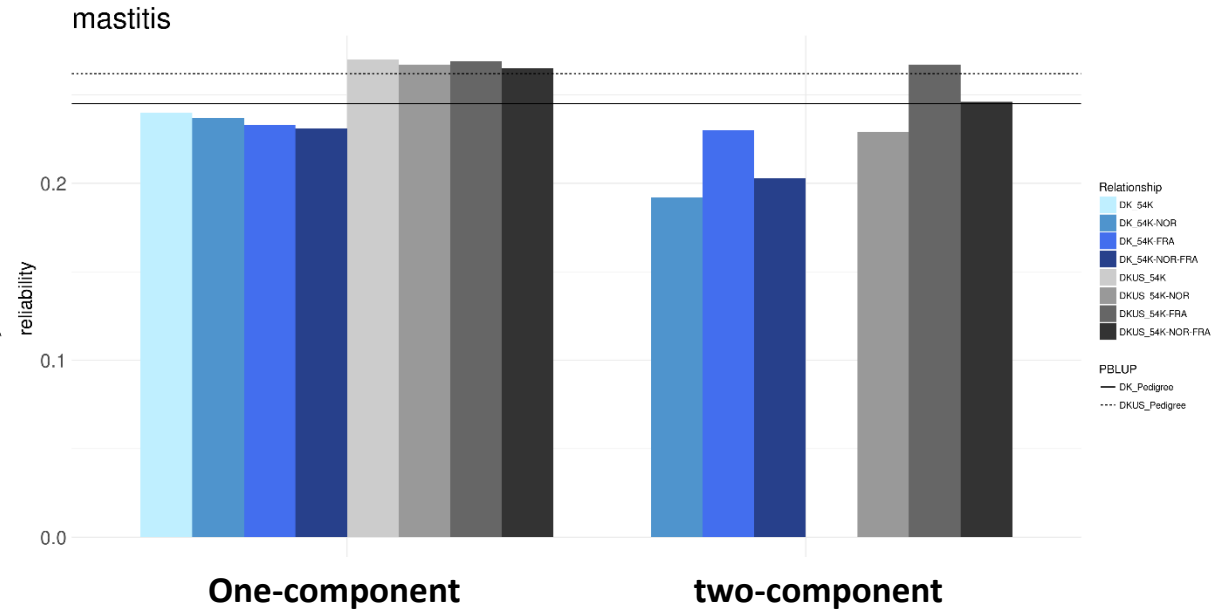
➤ Similar to milk

# Results-Reliability of fertility & mastitis

## fertility



## mastitis

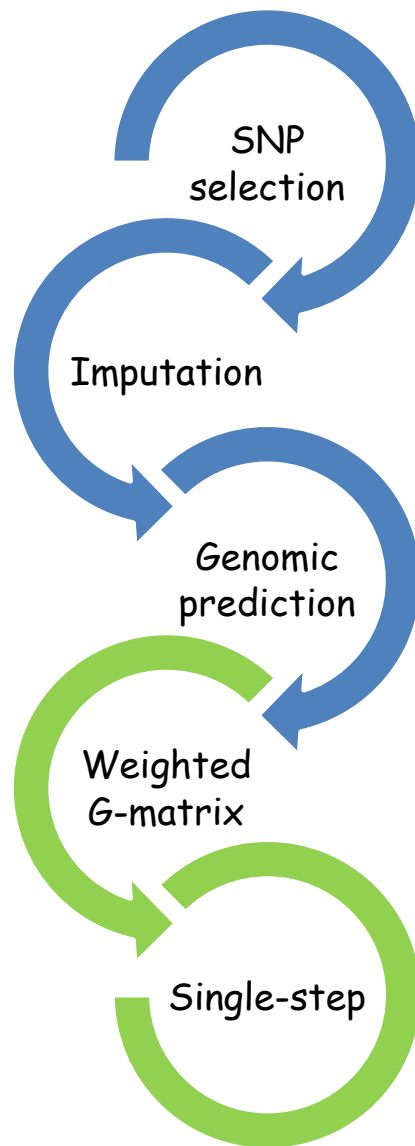


- No significant difference between 54K and 54K + selected SNPs
- A joint DK-US reference better than a DK reference for all traits

# Conclusions

- Additional WGS SNPs improved reliabilities for milk production, not for fertility and mastitis
- The inclusion of all additional WGS SNPs achieved highest reliabilities
- A joint DK-US reference better than a DK reference for all traits
- Additional WGS SNPs further improved reliabilities of a joint DK-US reference
- A two-component model improved reliabilities for milk production

# Next step



# Imputation accuracy

## Imputation accuracy

➤ Correlation = COR (TRUE, IMPUTED)

➤ Concordance rate =  $\frac{\text{No. of animals with corectly imputed genotypes}}{\text{No. of animals with imputed genotypes}}$

# Results-Reliability

*Reliability using a one-component GBLUP model (M1) and a two-component GBLUP model (M2)*

Trait	Reference	54K	54K + NOR		54K + FRA		54K+ NOR + FRA	
		M1	M1	M2	M1	M2	M1	M2
Milk	DK	37.6	41.9	43.1	43.0	48.0	44.0	47.2
	DKUS	45.3	50.5	52.8	51.3	53.8	52.3	54.9
Protein	DK	30.1	31.6	32.1	33.2	36.2	33.6	35.1
	DKUS	37.3	39.7	41.9	40.7	42.3	41.5	43.4
Fat	DK	24.2	26.9	27.1	26.9	28.3	27.4	27.7
	DKUS	29.3	32.2	32.9	32.2	33.1	32.8	33.6
Fertility	DK	27.5	27.1	24.1	27.3	27.0	26.9	25.4
	DKUS	28.3	28.1	28.0	28.1	27.9	28.0	27.6
Mastitis	DK	24.0	23.7	19.2	23.3	23.0	23.1	20.3
	DKUS	27.0	26.7	22.9	26.9	26.7	26.5	24.6

# Results-Bias

*Bias using a one-component GBLUP model (M1) and a two-component GBLUP model (M2)*

Trait	Reference	54K	54K + NOR		54K + FRA		54K+ NOR + FRA	
		M1	M1	M2	M1	M2	M1	M2
Milk	DK	0.832	0.853	0.832	0.871	0.909	0.872	0.891
	DK-US	0.826	0.842	0.832	0.849	0.859	0.852	0.858
Protein	DK	0.722	0.726	0.692	0.743	0.772	0.741	0.741
	DK-US	0.744	0.755	0.751	0.759	0.766	0.763	0.765
Fat	DK	0.706	0.729	0.680	0.738	0.751	0.738	0.719
	DK-US	0.694	0.711	0.705	0.712	0.713	0.715	0.714
Fertility	DK	1.036	1.027	0.954	1.030	1.024	1.021	0.985
	DK-US	0.991	0.990	0.987	0.981	0.974	0.980	0.970
Mastitis	DK	0.830	0.822	0.710	0.812	0.803	0.804	0.731
	DK-US	0.879	0.872	0.792	0.874	0.870	0.866	0.823





## ➤ Validation

- genotyped cows born after 2014-01-01
- these cows and their paternal female half-sibs born after 2008-07-01 as cow validation set
- excluding the half-sib families with size > 500
- **5,829 validation cows from 155 paternal half-sib families**

## ➤ Reference

- validation cows' maternal female and male half-sibs born after 2008-07-01 were excluded
- progenies of these animals (validation cows and the sibs) were removed



## ➤ No. of animals in Reference

Reference	No. of bulls	No. of cows
<b>COW</b>	--	<b>8,763</b>
<b>DK</b>	<b>1,282</b>	--
<b>DKUS</b>	<b>2,430</b>	--
<b>DKCOW</b>	<b>1,282</b>	<b>8,602</b>
<b>DKUSCOW</b>	<b>2,430</b>	<b>8,602</b>

# Prediction: validation on cows



## Reliabilities of milk

Reference	54K	54K+ NOR		54K+ FRA		54K+ NOR +FRA	
	M1	M1	M2	M1	M2	M1	M2
DK	0.248	0.318	0.363	0.323	0.363	0.340	0.369
DKUS	0.330	0.412	0.437	0.415	0.434	0.427	0.441
COW	0.450	0.513	0.517	0.520	0.527	0.523	0.524
DKCOW	0.470	0.534	0.551	0.538	0.549	0.544	0.554
DKUSCOW	0.503	0.558	0.570	0.562	0.570	0.567	0.573

# Prediction: validation on cows



## Bias of milk

Reference	54K	54K+ NOR		54K+ FRA		54K+ NOR +FRA	
	M1	M1	M2	M1	M2	M1	M2
DK	1.05	1.15	1.16	1.16	1.16	1.17	1.15
DKUS	1.11	1.20	1.20	1.20	1.19	1.21	1.20
COW	0.94	0.92	0.91	0.92	0.91	0.92	0.91
DKCOW	1.03	1.03	1.04	1.03	1.02	1.03	1.03
DKUSCOW	1.07	1.07	1.08	1.07	1.06	1.07	1.07