AARHUS
UNIVERSITY

CENTER FOR QUANTITATIVE
GENETICS AND GENOMICS
QGG

# Use of Whole Genome Sequence variants in genomic prediction

# GenSAP challenges

- Use of full sequence data
  - Inc integration of external information
- Predictions across breeds and populations
- Plants and new animal species
- Non-additive genetic effects
- Analysis of large datasets
- Estimation and control of inbreeding using Genomics

# Lessons learned

- Sequence include causal and high LD variants
- Real data: More markers only marginally improve predictions
  - Real data: HD ≈ 54K (Su et al., 2012)
  - Imputed WGS ≈ HD (Van Binsbergen et al., 2015 )
- Theoretic study by Van den Berg et al., 2016 (SFA2)
  - Small improvements within breeds
  - Across breeds: Only use markers very close to causal variants (others add noise)
- Low MAF variants are poorly imputated

# Lessons learned

- Not all SNPs are equal
- Feature models may improve predictions (SFA1)
  - Need SNP set highly enriched with causal and high LD variants
  - Best in for unrelated individuals
- Bayesian Variable Selection Models (SFA2)
  - Discriminate between high/low variance SNPs
  - Require high computation time
- Weighted G(SNP)BLUP can specify identical models/predictions (Su et al., 2014) (SFA2)

# Challenge

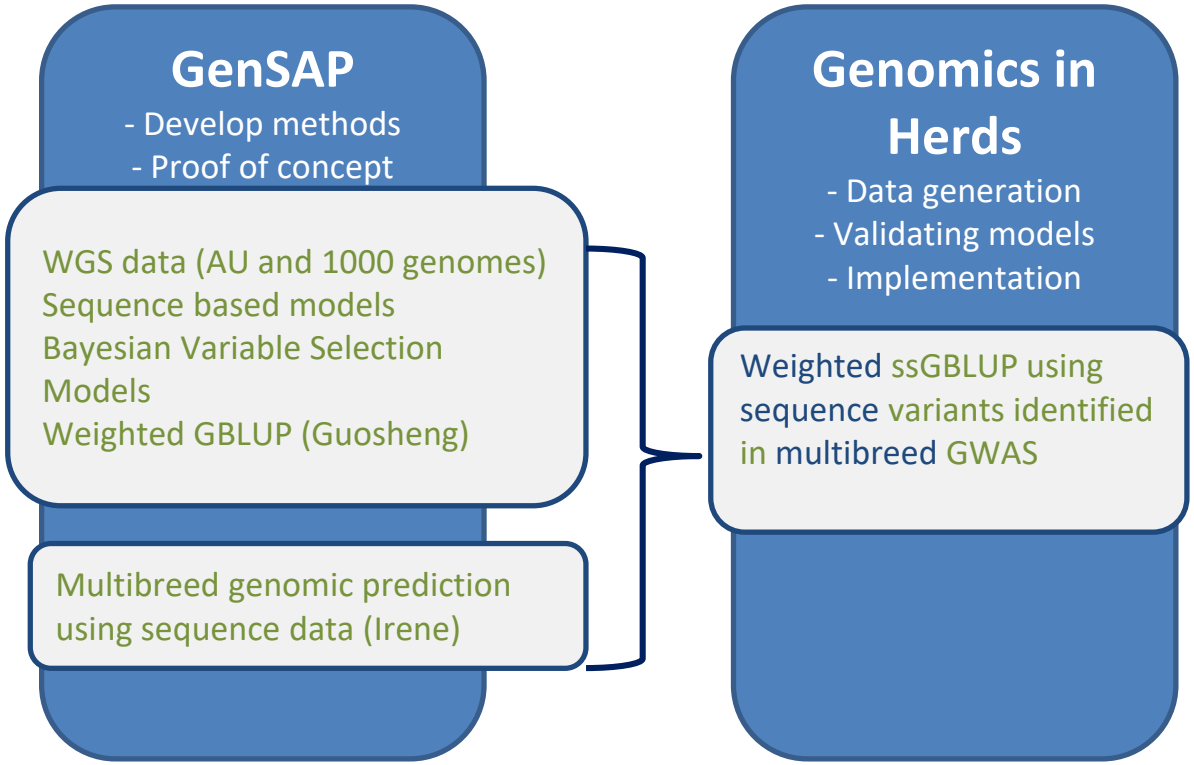
**Can we develop models that:**

- **improve genomic predictions by using whole genome sequence variants**
- **Can be implemented in routine evaluations**

# Strategy to meet challenge

- Identify set of SNPs enriched with causal variants
  1. GWAS using multi breed data: Identify 3-5 top SNPs/QTL
  2. Functional annotated information
- Genotype large number of cows (custum chip)
- Estimate parameters in BVS models or genomic feature model
- Develop equivalent model by weighted G(SNP)BLUP

# GenSAP and industry projects



CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

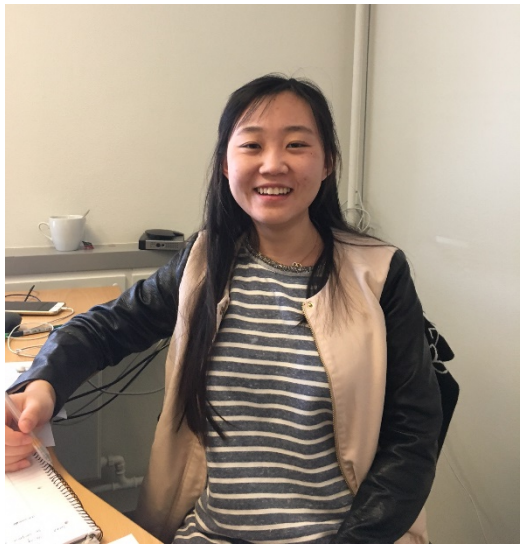
# Using additional SNPs selected from whole genome sequence (WGS) data for genomic prediction in Danish Jersey

Aoxing Liu,

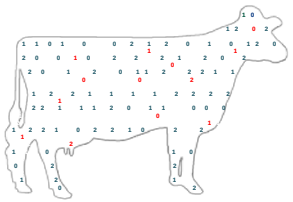
Mogens Sandø Lund, Didier Boichard,Sebastien Fritz, Emre Karaman, Yachun Wang, Guosheng Su

AARHUS UNIVERSITY

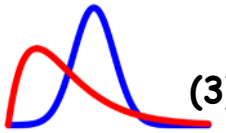
INRA SCIENCE & IMPACT



# Objectives



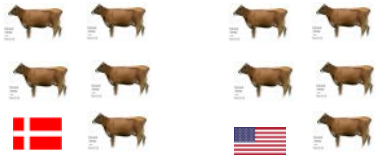
(1) 54 K +
SNPs selected from WGS



(3) Models



(2) Joint reference

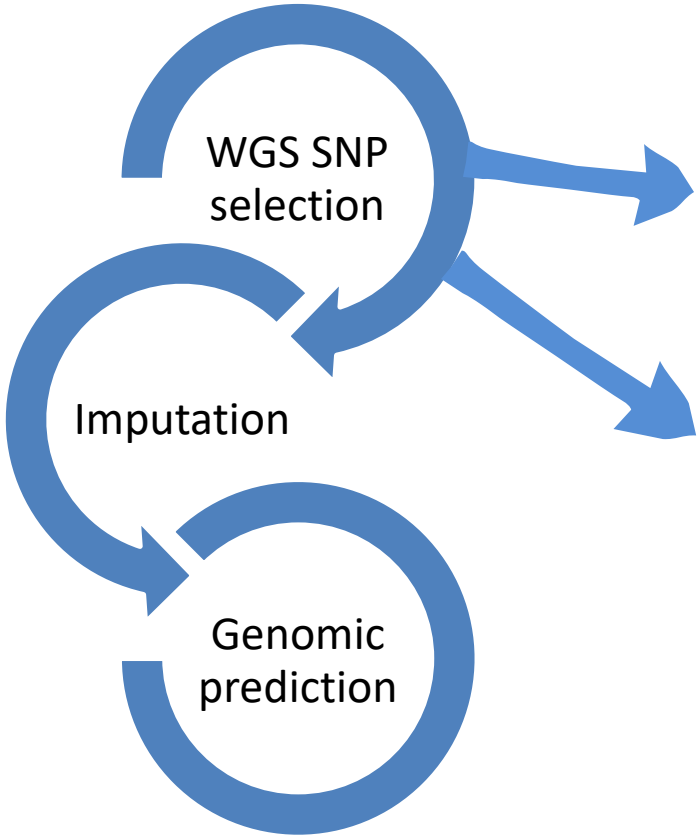


Reliable
genomic prediction

- Investigate effects of additional WGS SNPs on genomic prediction
- Effects of using additional WGS SNPS in a joint reference
- Assessed models on their efficiency to use information of additional WGS SNPs

# Workflow






**NOR SNPs (Goutam, Bernt, Xiaoping, Zexi)**

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

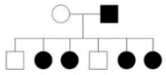
**FR SNPs (Didier et al.)**

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

# Imputation

**Animal**
- DK bulls: ~1,300
- US bulls: ~1,200
- DK cows: ~31,000

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



**Pedigree**
- 6,100 males
- 66,000 females

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

(1)

**54K**

(2)

**54K**
**+ NOR SNPs + FRA SNPs**

# Reference and validation populations

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QCG

- **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

| Reference | N_BULL | N_COW |
|---|---|---|
| COW | -- | 8,763 |
| DK | 1,282 | -- |
| DKUS | 2,430 | -- |
| DKCOW | 1,282 | 8,602 |
| DKUSCOW | 2,430 | 8,602 |

# Prediction: GBLUP and BVS models

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

- **One-component model**

$$y = 1\mu + X\,g + e$$

54K/ 54K+selected WGS SNPs

| Scenarios |
|---|
| 54K |
| 54K_NOR |
| 54K_FRA |
| 54K_NOR_FRA |

- **Two-component model**

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

54K — Selected WGS SNPs

| Component_One | Component_Two |
|---|---|
| 54K | NOR |
| 54K | FRA |
| 54K | NOR+FRA |

Bootstrap to asses significance

# Prediction: validation on cows

**Milk**

**GBLUP**

| Reference | PBLUP | 54K G1 |
|---|---|---|
| DK[1] | 13.2 | 31.2 |
| DKUS[2] | 17.4 | 41.5 |
| COW[3] | 10.5 | 56.3 |
| DKCOW[4] | 14.9 | 59.8 |
| DKUSCOW[5] | 18.9 | 63.7 |

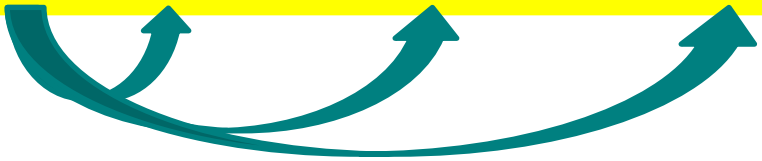
Large improvement with increased reference

# GBLUP: Large improvements from sequence variants

**Milk**

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 13.2 | 31.2 | 40.0 | | 40.7 | | 42.8 | |
| DKUS[2] | 17.4 | 41.5 | 51.7 | | 52.2 | | 53.6 | |
| COW[3] | 10.5 | 56.3 | 64.1 | | 65.0 | | 65.3 | |
| DKCOW[4] | 14.9 | 59.8 | 67.8 | | 68.4 | | 69.0 | |
| DKUSCOW[5] | 18.9 | 63.7 | 70.4 | | 71.0 | | 71.4 | |



Large improvement by adding sequence variants

# Models

**Milk**

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 13.2 | 31.2 | 40.0 | 45.3 | 40.7 | 45.5 | 42.8 | 46.0 |
| DKUS[2] | 17.4 | 41.5 | 51.7 | 54.4 | 52.2 | 54.5 | 53.6 | 55.1 |
| COW[3] | 10.5 | 56.3 | 64.1 | 64.5 | 65.0 | 65.9 | 65.3 | 65.4 |
| DKCOW[4] | 14.9 | 59.8 | 67.8 | 69.5 | 68.4 | 69.7 | 69.0 | 70.0 |
| DKUSCOW[5] | 18.9 | 63.7 | 70.4 | 71.6 | 71.0 | 71.8 | 71.4 | 72.0 |

**BVS**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | B1 | B1 | B2 | B1 | B2 | B1 | B2 |
| DK[1] | 13.2 | 41.3 | 47.5 | 48.2 | 48.2 | 49.9 | 48.6 | 48.3 |
| DKUS[2] | 17.4 | 50.6 | 57.3 | 57.4 | 57.2 | 57.8 | 57.8 | 58.1 |
| COW[3] | 10.5 | 64.4 | 67.3 | 67.0 | 67.5 | 62.4 | 67.5 | 67.7 |
| DKCOW[4] | 14.9 | 71.2 | 71.6 | 71.6 | 48.3 | 72.2 | 72.0 | 72.2 |
| DKUSCOW[5] | 18.9 | 71.2 | 73.8 | 73.7 | 74.0 | 73.9 | 74.0 | 74.2 |

Bayesian models better than GBLUP

# Improvement from WGS SNPs and models for milk yield

## Milk

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 13.2 | 31.2 | 40.0 | 45.3 | 40.7 | 45.5 | 42.8 | 46.0 |
| DKUS[2] | 17.4 | 41.5 | 51.7 | 54.4 | 52.2 | 54.5 | 53.6 | 55.1 |
| COW[3] | 10.5 | 56.3 | 64.1 | 64.5 | 65.0 | 65.9 | 65.3 | 65.4 |
| DKCOW[4] | 14.9 | 59.8 | 67.8 | 69.5 | 68.4 | 69.7 | 69.0 | 70.0 |
| DKUSCOW[5] | 18.9 | 63.7 | 70.4 | 71.6 | 71.0 | 71.8 | 71.4 | 72.0 |

**BVS**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | B1 | B1 | B2 | B1 | B2 | B1 | B2 |
| DK[1] | 13.2 | 41.3 | 47.5 | 48.2 | 48.2 | 49.9 | 48.6 | 48.3 |
| DKUS[2] | 17.4 | 50.6 | 57.3 | 57.4 | 57.2 | 57.8 | 57.8 | 58.1 |
| COW[3] | 10.5 | 64.4 | 67.3 | 67.0 | 67.5 | 62.4 | 67.5 | 67.7 |
| DKCOW[4] | 14.9 | 68.5 | 71.6 | 71.6 | 48.3 | 72.2 | 72.0 | 72.3 |
| DKUSCOW[5] | 18.9 | 71.2 | 73.8 | 73.7 | 74.0 | 73.9 | 74.0 | 74.2 |

1)54K+ WGS > 54K
2) Reference increase, 54K+ WGS > 54K

3) Bayesian > GBLUP
2) Two > one for GBLUP, not consist for Bayesian

Potential: 63.7 → 74.2

# Improvement from WGS SNPs and models for protein yield

## Protein

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 17.5 | 26.6 | 29.2 | 30.4 | 29.9 | 30.7 | 30.7 | 31.0 |
| DKUS[2] | 20.9 | 32.7 | 35.7 | 36.7 | 36.3 | 36.8 | 37.0 | 37.2 |
| COW[3] | 11.6 | 35.8 | 38.6 | 38.2 | 39.6 | 39.7 | 39.7 | 39.3 |
| DKCOW[4] | 15.0 | 39.0 | 42.1 | 42.7 | 42.8 | 43.2 | 43.1 | 43.3 |
| DKUSCOW[5] | 17.8 | 41.9 | 44.5 | 44.7 | 45.0 | 45.2 | 45.2 | 45.2 |

**BVS**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | B1 | B1 | B2 | B1 | B2 | B1 | B2 |
| DK[1] | 17.5 | 29.1 | 30.9 | 31.0 | 31.4 | 31.1 | 31.5 | 30.7 |
| DKUS[2] | 20.9 | 36.2 | 38.4 | 38.2 | 38.2 | 37.6 | 38.5 | 37.6 |
| COW[3] | 11.6 | 38.1 | 39.6 | 39.4 | 40.2 | 40.4 | 40.1 | 39.8 |
| DKCOW[4] | 15.0 | 41.7 | 43.6 | 43.6 | 31.1 | 44.2 | -- | -- |
| DKUSCOW[5] | 17.8 | 44.7 | 46.0 | 46.0 | 46.1 | 46.2 | -- | -- |

Same conclusions but at lower level

Potential: 41.9 → 46.2

# Improvement from WGS SNPs and models for fat yield

## Fat

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 19.9 | 26.7 | 28.1 | 28.3 | 27.7 | 27.4 | 28.2 | 27.8 |
| DKUS[2] | 21.4 | 29.8 | 31.3 | 31.3 | 30.9 | 30.7 | 31.3 | 31.1 |
| COW[3] | 14.8 | 33.1 | 33.9 | 34.0 | 34.1 | 34.3 | 34.3 | 34.5 |
| DKCOW[4] | 22.2 | 37.1 | 37.7 | 37.6 | 37.6 | 37.7 | 37.8 | 37.8 |
| DKUSCOW[5] | 23.1 | 37.9 | 38.6 | 38.6 | 38.4 | 38.4 | 38.6 | 38.6 |

**BVS**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | B1 | B1 | B2 | B1 | B2 | B1 | B2 |
| DK[1] | 19.9 | 27.2 | 28.2 | 28.5 | 27.8 | 27.6 | 28.3 | 27.6 |
| DKUS[2] | 21.4 | 30.4 | 31.6 | 31.6 | 31.0 | 30.5 | 31.5 | 30.9 |
| COW[3] | 14.8 | 33.7 | 34.2 | 26.8 | 34.5 | 34.5 | 34.6 | 34.7 |
| DKCOW[4] | 22.2 | 27.2 | 37.7 | 37.8 | 28.0 | 37.6 | 37.8 | 37.7 |
| DKUSCOW[5] | 23.1 | 39.0 | 39.1 | 39.2 | 38.9 | 38.7 | 38.9 | 38.9 |

Same tendencies but at lower level and not significant

Potential: 37.9 → 38.9

# Improvement from WGS SNPs and models for mastitis

## Mastitis

**GBLUP**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | G1 | G1 | G2 | G1 | G2 | G1 | G2 |
| DK[1] | 20.5 | 32.4 | 32.8 | 33.0 | 33.4 | 32.3 | 33.1 | 33.7 |
| DKUS[2] | 19.6 | 33.9 | 34.1 | 34.5 | 34.8 | 33.4 | 34.6 | 35.0 |
| COW[3] | 16.8 | 34.0 | 34.5 | 34.4 | 34.8 | -- | -- | 34.3 |
| DKCOW[4] | 15.4 | 39.8 | 40.2 | 40.6 | 40.9 | 40.4 | 40.4 | 40.9 |
| DKUSCOW[5] | 14.6 | 40.0 | 40.3 | 40.8 | 41.0 | 40.3 | 40.6 | 41.0 |

**BVS**

| Reference | PBLUP | 54K | 54K + NOR | | 54K + FRA | | 54K+ NOR + FRA | |
|---|---|---|---|---|---|---|---|---|
| | | B1 | B1 | B2 | B1 | B2 | B1 | B2 |
| DK[1] | 20.5 | 31.6 | 30.8 | 30.1 | 32.2 | 32.3 | 32.2 | 31.2 |
| DKUS[2] | 19.6 | 33.8 | 33.8 | 32.0 | 34.3 | 33.9 | 33.9 | 32.8 |
| COW[3] | 16.8 | 34.2 | 34.4 | 34.8 | 34.4 | 34.6 | 34.7 | 34.8 |
| DKCOW[4] | 15.4 | 39.3 | 39.4 | 39.4 | 40.4 | 40.4 | 40.3 | 40.3 |
| DKUSCOW[5] | 14.6 | 39.9 | 40.2 | 40.1 | 40.7 | 40.6 | 41.1 | 41.0 |

1) Minor improvements and not significant

## Conclusions

- **Using additional WGS SNPs improve reliabilities for production traits considerably (also wth larger reference)**
- **Bayesian VSM were better than GBLUP**
- **No clear differences between one-component and two-component models**
- **Improvements will be testet and implemented by Weighted-SS-SNPBLUP**
- **Further improvements**
  - Better understanding of genome → improvements for other traits
  - More sequence data → improved imputation → Better markers for LDchip

# CHALLANGE

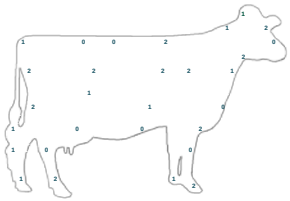
CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

- Large scale ssG(SNP)BLUP
- Sequence information
- Bayesian methods
- Multitrait heterogeneous (co)variance models

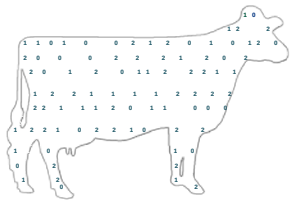
Integrate in model for rutine evaluation

# Background

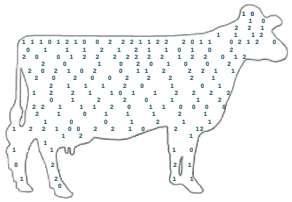
High throughput genotyping



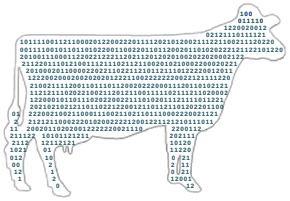
LD (7 K)



MD (54 K)



HD (777 K)



WGS (~26,700 K)

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

**Real data:** HD ≈ 54K (Su et al., 2012) & Imputed WGS ≈ HD (Van Binsbergen et al., 2015 )

- Only causative mutations or variants very close to causative mutations can improve reliability

  (van den Berg et al., 2016)

- non-causative mutations bring noise

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

# Current Danish (QGG) 'genomics in herds' team

Aoxing Liu, Emre Karaman, Zexi Cai, Yahui Gao

Goutam, Bernt, Guosheng, Mogens