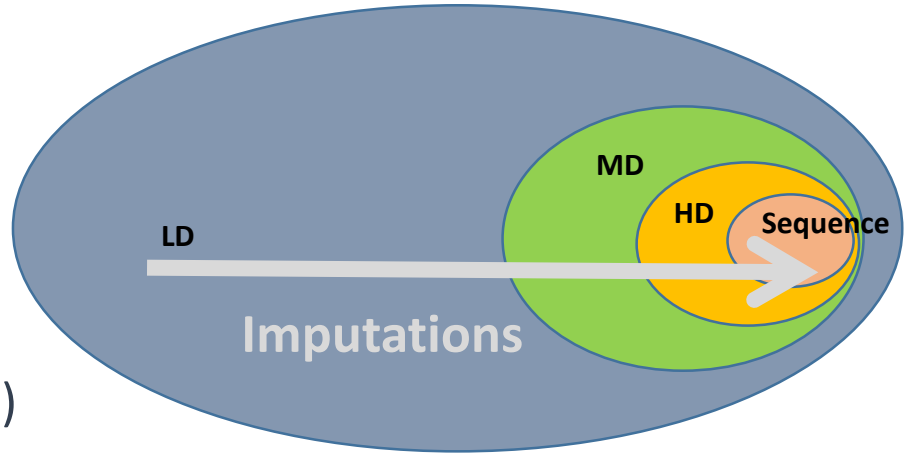
Inclusion of candidate mutations in genomic evaluation for French dairy cattle breeds.

Pascal Croiseau, Chris Hozé, Sébastien Fritz, Marie-Pierre Sanchez, Thierry Tribout

This study was conducted in the frame of the INCOMINGS project, funded by INRA (SELGEN Metaprogram, France)

Context

- The 1000 bull genomes project offers an excellent opportunity to move from QTL detection to causal mutation identification
- In this aim :
  - ✓ All French genotyped animals were imputed to the sequence level



  - ✓ GWAS were performed for all the available traits in Holstein and Montbéliarde breed (GCTA software)
  - ✓ Main candidate mutations were selected and designed on the custom part of the LD chip.

Aim of the study

- Assess whether the use of the genotypes for these candidate mutations could improve the accuracy of genomic predictions
- 2 scenarios were tested :
  - ✓ Virtual 50K chips (V50K) were designed by replacing all or part of the 50K by candidate mutation identified from GWAS at the sequence level
    - same number of SNP than the 50K
    - SNP « automatically » selected
  - ✓ The 50K was augmented with the candidate mutations of the custom part of the LD chip.
    - SNP « manually » selected

Context

- 2 breeds studied:
  - ✓ Montbéliarde : 2608 bulls (2086 in training pop. + 522 in validation pop.)
  - ✓ Holstein : 6382 bulls (5106 in training pop. + 1276 in validation pop.)

- 10 studied traits:
  - ✓ 5 production traits (Milk, protein anf fat yield, protein and fat content)
  - ✓ 2 type traits (rear udder width, height at sacrum)
  - ✓ 3 functional traits (somatic cell count, fertility for cow, milking speed)

- Genomic Evaluation
  - ✓ GBLUP
  - ✓ Correlation between DYD and DGV on the validation population

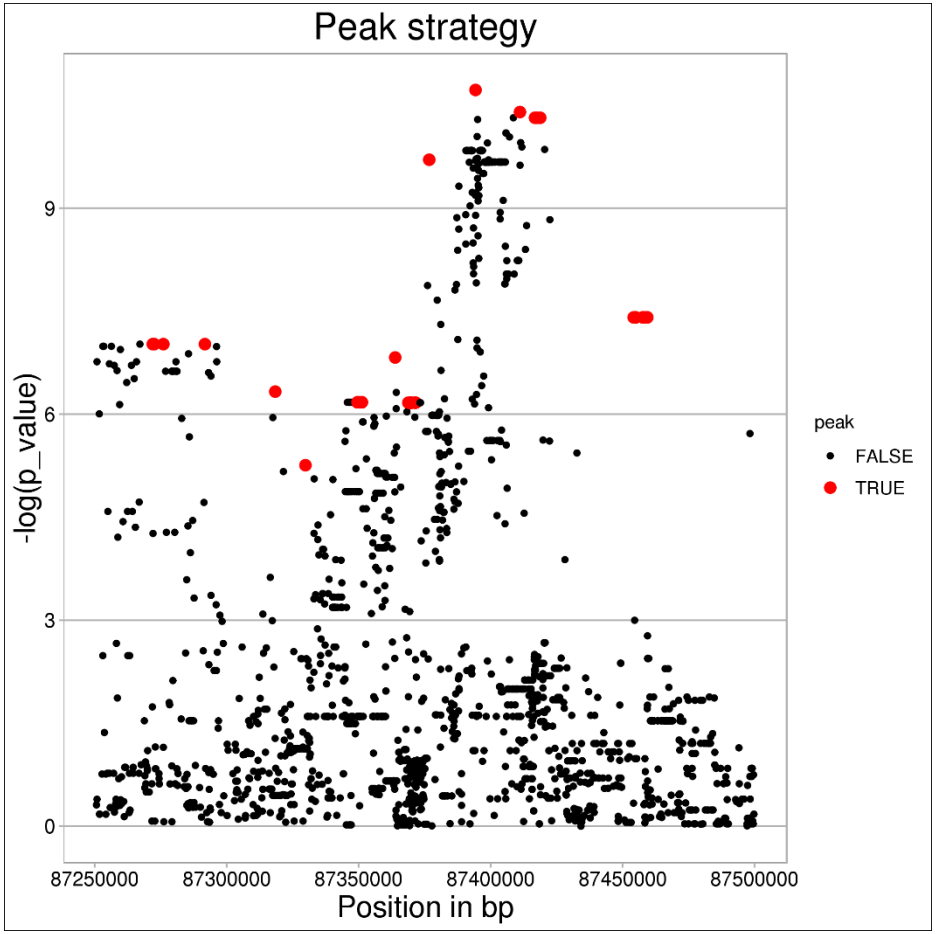
Virtual 50K chips strategies:

All SNP of the Illumina 50K were replaced (43,000 SNP routinely used in genomic evaluation)

✓ PEAK :

To be selected, a SNP should have the lowest p-value in a window of ±150 SNP.

When too many peaks were identified, only the most significant were retained



Virtual 50K chips strategies:

All SNP of the Illumina 50K were replaced (43,000 SNP routinely used in genomic evaluation)

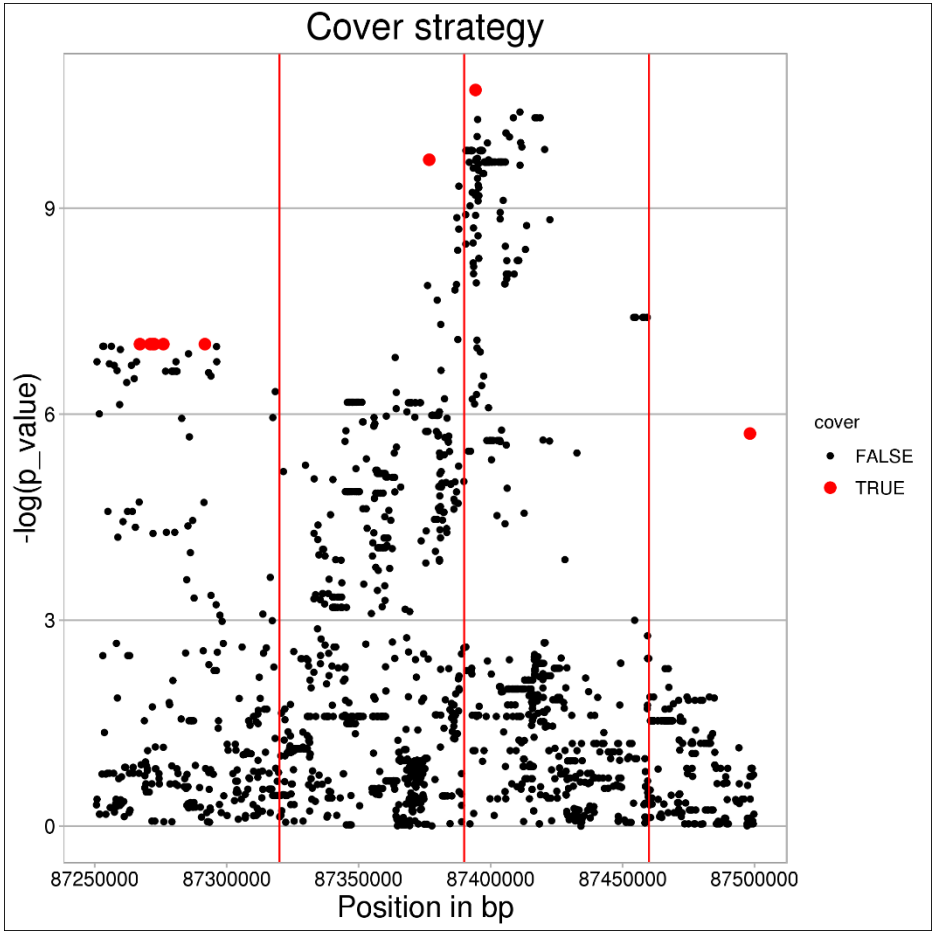
- ✓ PEAK
- ✓ COVER:

  The genome was divided in 43000 segments.

  In each segment, the SNP with a MAF >0.1 and the lowest p-value was retained.
- ✓ COVER2:

  Same as COVER but the retained SNP had a MAF>0.1and the lowest p-value with a

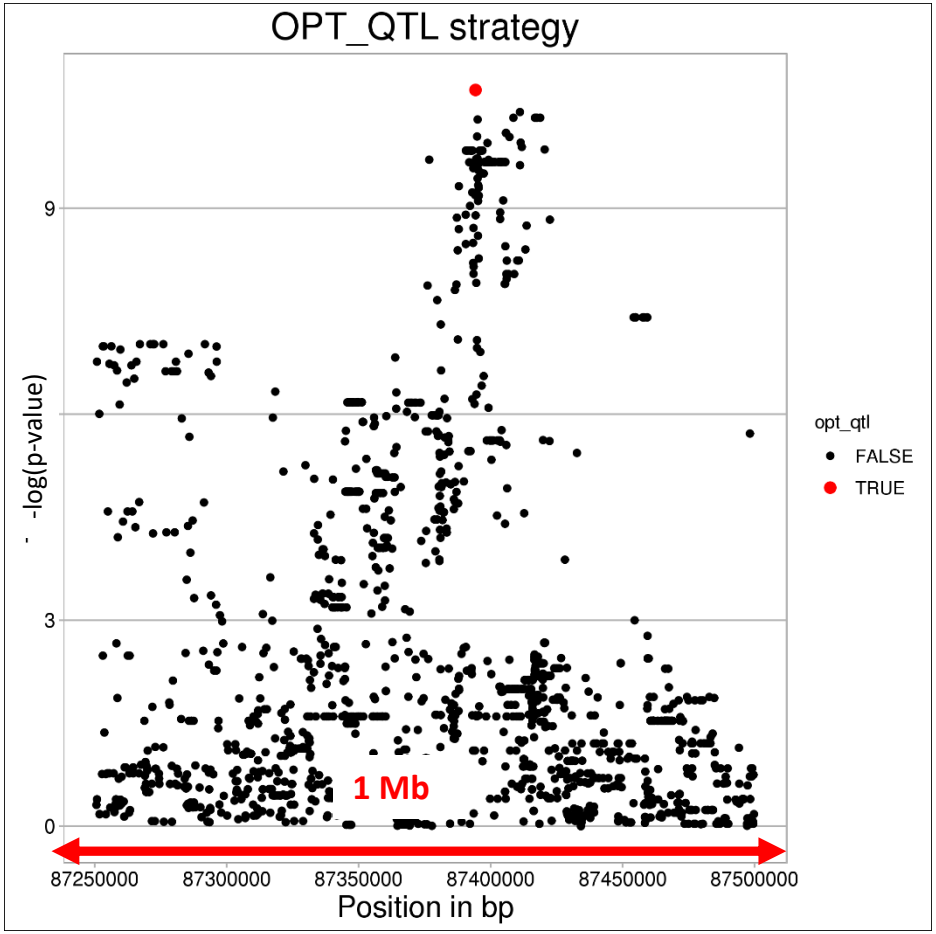
  priority for SNP located in a gene.



Virtual 50K chips strategies:
Only a part of the 50K SNP were replaced

- ✓ PEAK
- ✓ COVER
- ✓ COVER2
- ✓ OPT_QTL:

  Genome was split in region of 1Mb.
  All regions with at least one significant SNP were retained (using criteria on MAF and p-value threshold).
  In each of these selected regions, the corresponding SNP of the 50K chip with the lowest MAF was replaced by the most significant SNP.



Virtual 50K chips strategies:
Only a part of the 50K SNP were replaced

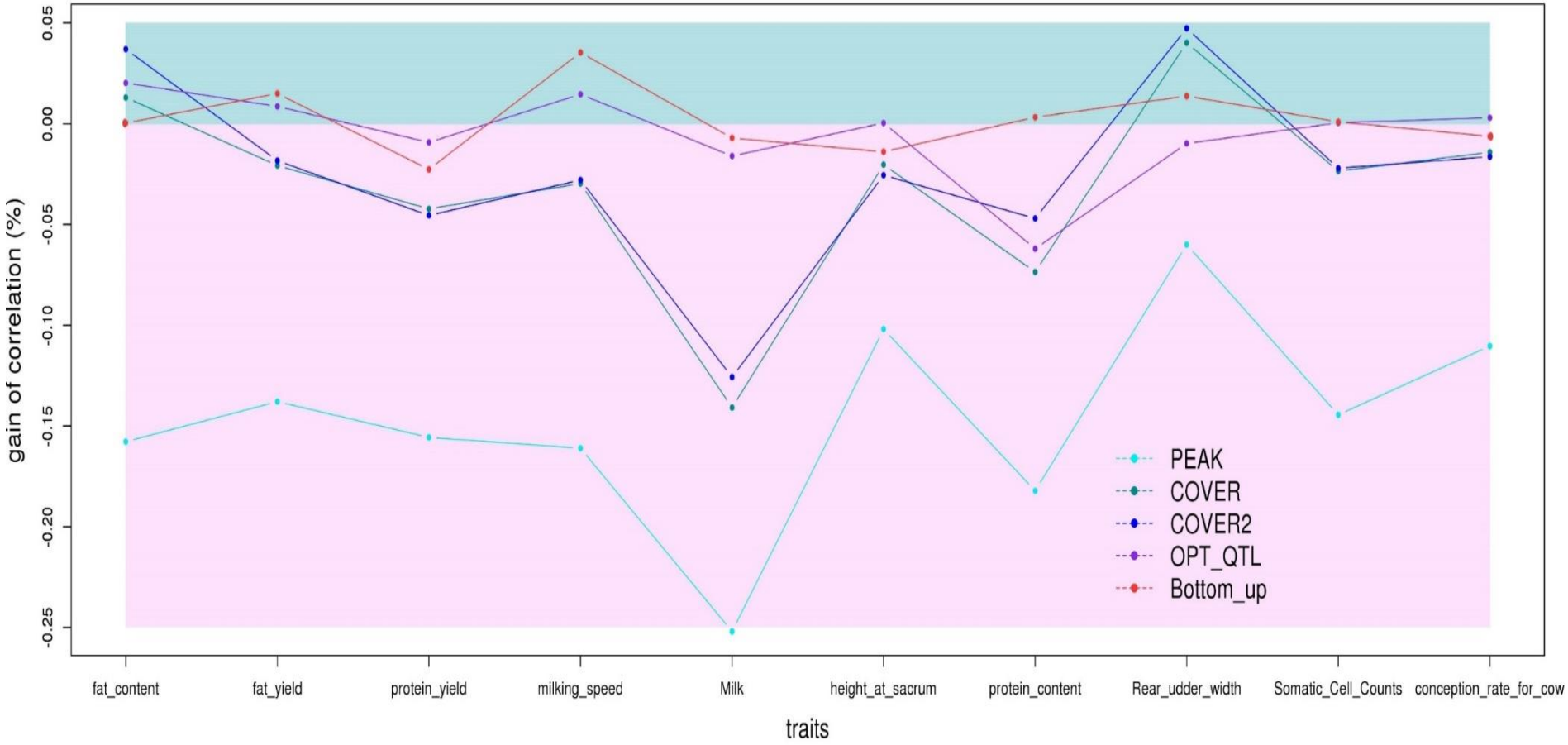
- ✓ PEAK
- ✓ COVER
- ✓ COVER2
- ✓ OPT_QTL
- ✓ Bottom-up:

  After running GBLUP on the 50K, part of the SNP (around 3000) had an estimated effect close to 0.

  These SNP were replaced by the SNP with the lowest p-value in a region of ± 0.5Mb around the 50K SNP.

Virtual 50K chips strategies

Gain of Correlation using V50K compared to the 50K chip



50K augmented with the custom part of the LD chip

Custom part

Causal Mutations described in the litterature

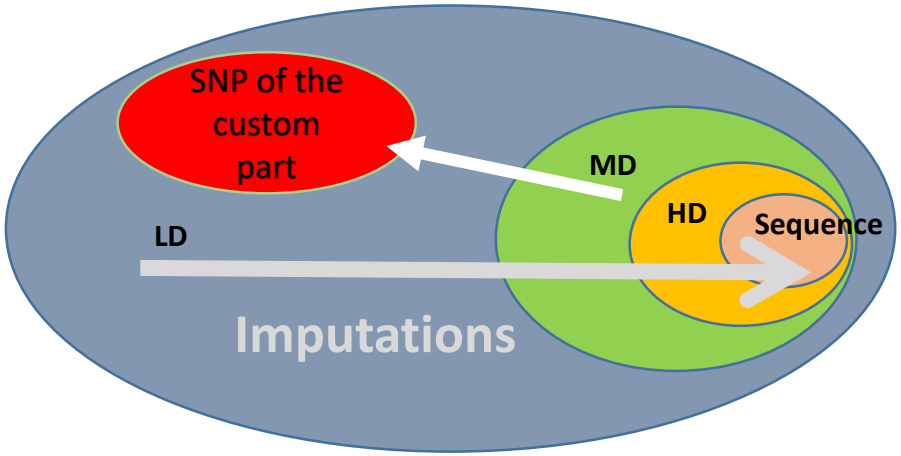
- ✓ 62 SNP in 36 genes
- ✓ Between 1 and 22 SNP included according to the traits

Candidate Mutations selected using

- ✓ P-value
- ✓ Annotation
- ✓ MAF
- ✓ Around 2000 selected SNP
  - 10-15 SNP per QTL
  - Between 151 and 416 SNP according to the traits

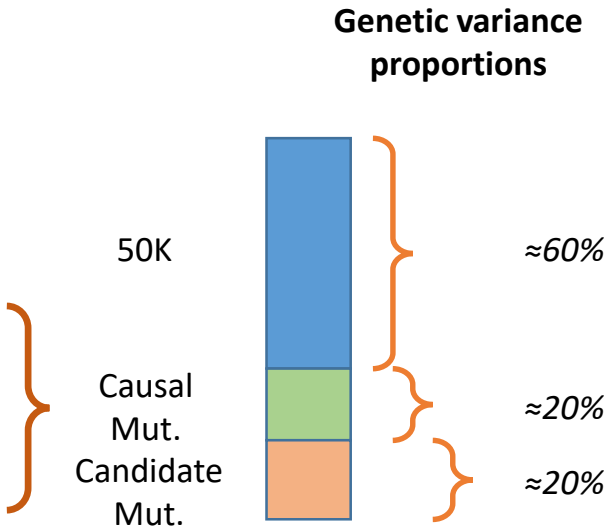
Reverse imputation

- ✓ Old animals were imputed from younger one
- ✓ Higher precision due to a larger reference population

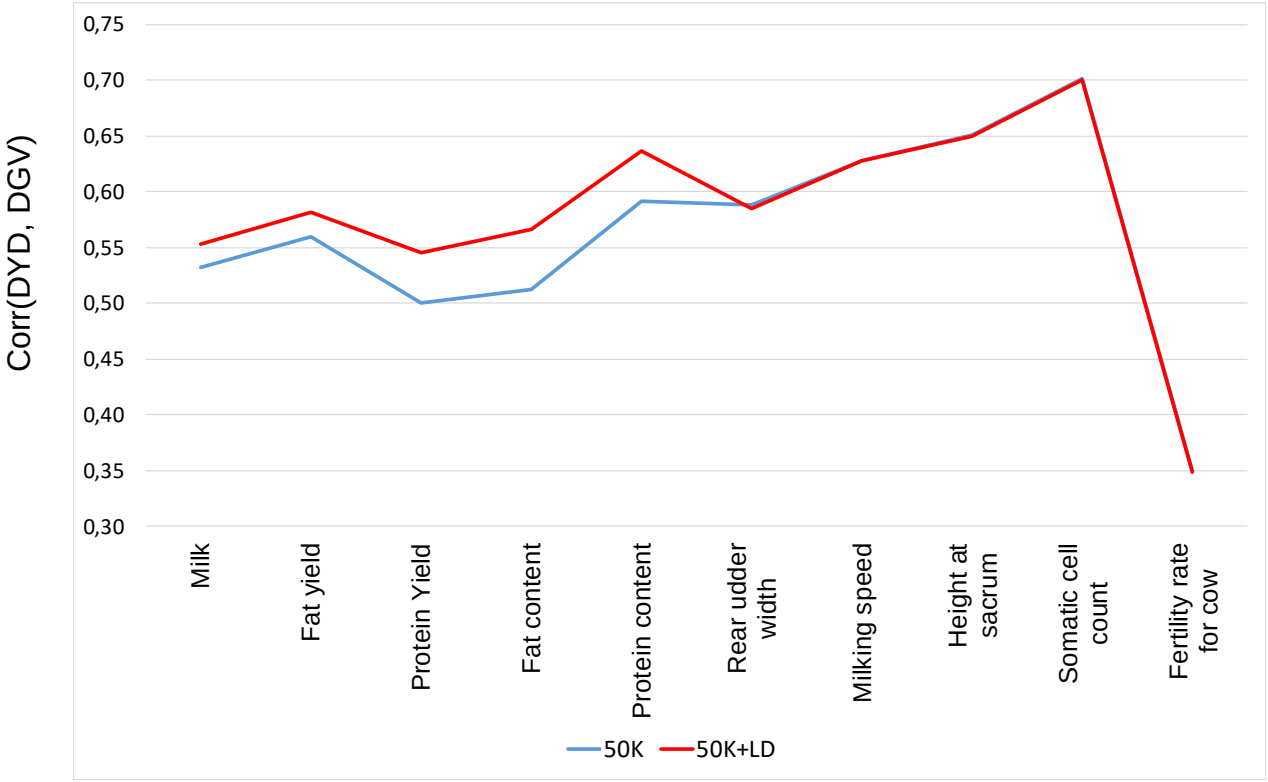


50K augmented with the custom part of the LD chip

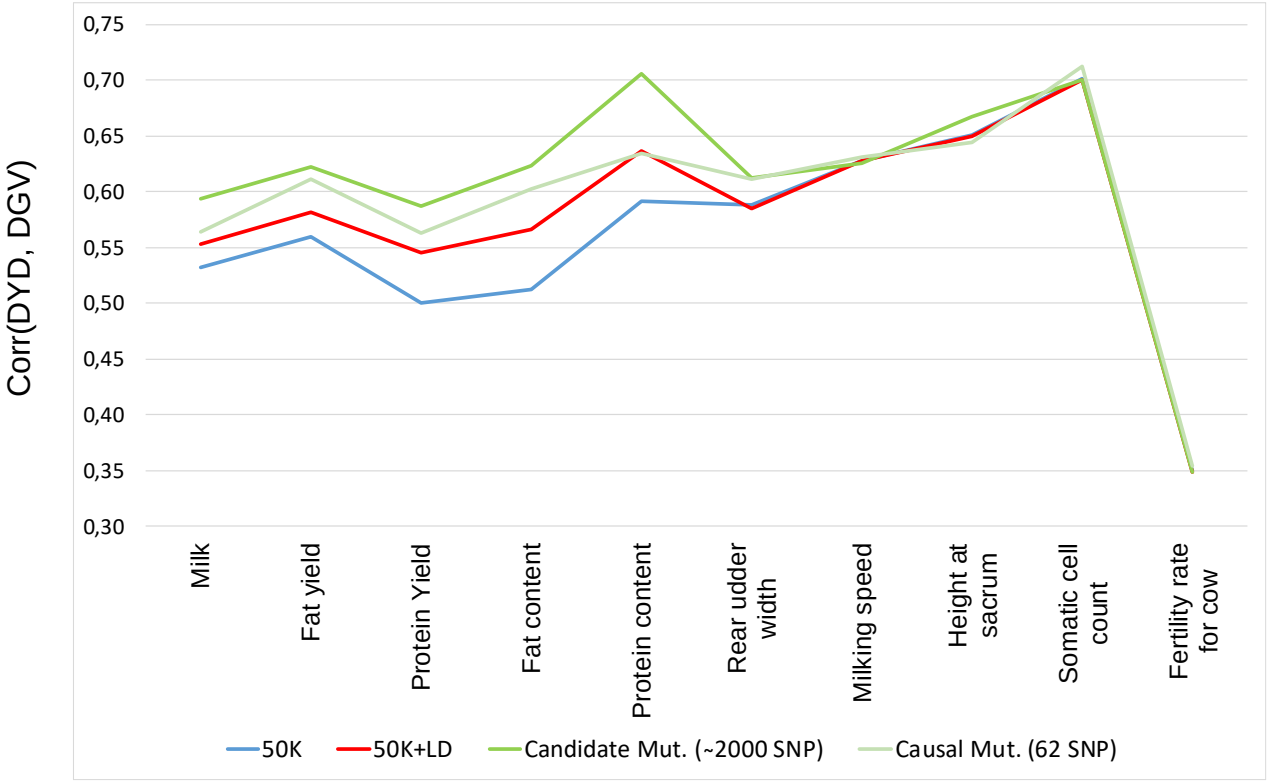
- GBLUP run with different dataset
  - ✓ 50K
  - ✓ 50K +LD (same genetic variance for each SNP)
  - ✓ Causal Mut. : 50K + causal mutations
  - ✓ Candidate Mut. : 50K + candidate mutations
  - ✓ 3 cat : 50K + causal mut. + candidate mut.



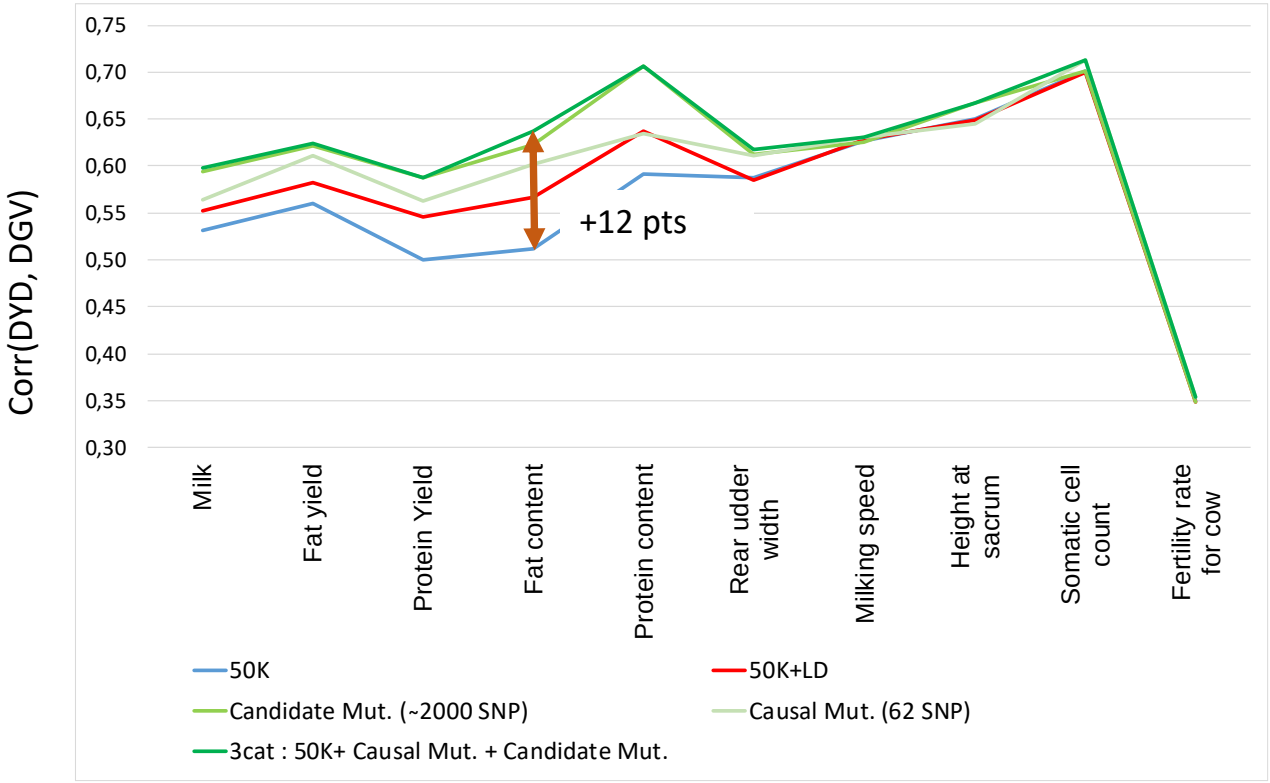
Results : Montbéliarde



Results : Montbéliarde

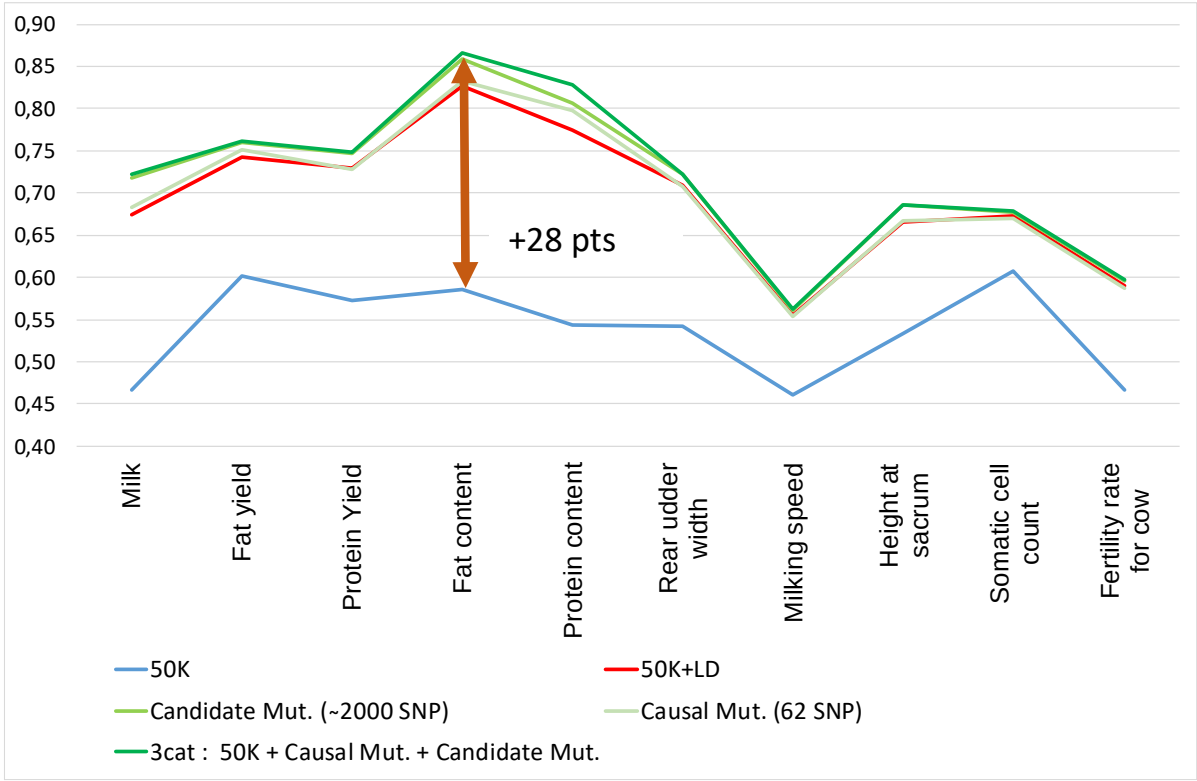


Results : Montbéliarde



Clear interest to well define genetic variance of each group

Results : Holstein



Conclusion and Perspectives

- According to the way candidate mutation are selected, the gain in accuracy vary drastically
  - ✓ Virtual chip failed to improve accuracy of genomic evaluations
  - ✓ Inclusion of 10-15 SNP per major candidate mutation on the custom part of the LD chip gives interesting results
    - Gain of accuracy included
      - between 0 and 12 points in Montbéliarde
      - Between 7 and 28 points in Holstein
    - Production traits were particularly positively impacted
      - Average gain of 9 points in Montbéliarde
      - Average gain of 28 points in Holstein

THANKS YOU FOR YOUR ATTENTION