# Using whole genome sequences to identify candidate mutations affecting Milk Fatty Acids in dairy cattle

*Govignon-Gion A.[1], Sanchez M.P. [1] Croiseau P. [1], Barbat A. [1], Fritz S. [2], Boussaha M. [1], Brochard M. [3], Boichard D. [1]*

*[1]INRA, 78350 Jouy en Josas*

*[2]ALLICE, 75012 Paris*

*[3]IDELE, 75012 Paris*

*France*

# *Introduction*

*Complex traits are influenced by **many QTLs**, each explaining a small part of their variability*

*QTL full characterization is especially challenging and **only a few QTLs have been identified so far**, in spite of large efforts*

*New tools have become available:*
*- High throughput **genotyping** of large populations*
*- **Whole genome sequencing***

*Is it now possible to re-address the question of QTL identification in **a more efficient way** ?*

# *Introduction*

*Objectives*

Identification of candidate **causal mutations** for milk fat composition

***PhénoFinLait** project*

8746 cows with milk fat composition & 50k genotypes

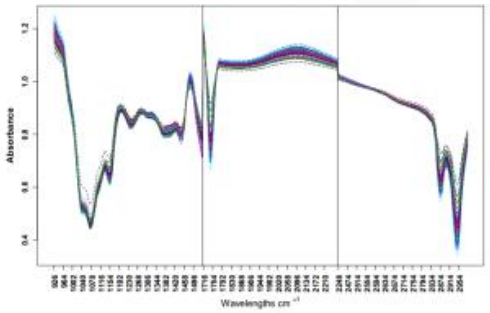
***« 1000 bull genomes »** project*

1147 bulls with whole genome sequences (RUN4)

Genome Wide Association Study (**GWAS**) at the **full sequence** level

# *Material & methods: 23 Fatty acids estimated by MIR*

Mid-Infrared **(MIR)** spectra




**Pre-correction** of data
for **non genetic** effects
Herd * test-day
Month * year of calving
Parity * days in milk

| **Total SAT** |
|---|
| **C4:0** |
| C6:0 |
| C8:0 |
| C10:0 |
| **C12:0** |
| C14:0 |
| **C16:0** |
| C18:0 |

| **Total UNSAT** | |
|---|---|
| Total MONO | **Total POLY** |
| **C18:1cis9** | C18:2cis9trans11 |
| C18:1cis12 | C18:2cis9cis12 |
| C18:1t11t10 | **C18:3n3** |
| TotC18:1 | TotC18:3 |
| TotC18:1cis | Omega 3 |
| TotC18:1trans | Omega 6 |

# *Material & methods: animals*

~ **120,000 cows with phenotypes**
(~ 600,000 test-day milk samples)

**8746** cows genotyped with the 50k Beadchip

| **2882** | **2816** | **3048** |
|---|---|---|
| Montbéliardes | Normandes | Holstein |
| **MON** | **NOR** | **HOL** |



# *Material & methods: genotypes & imputation*

***Imputation in two steps with FImpute** (Sargolzaei et al., 2014)*

**Bovine SNP50**

**Step 1** Within breed

**Bovine HD**

**Step 2** Within breed, with across breed reference

**Whole genome sequence**

***Reference populations***

**Within breed, HD genotyped bulls**
522 MON
546 NOR
776 HOL

**1 Reference population**
= 1147 bulls from
« 1000 Bull Genomes »
including
28 MON + 24 NOR + 288 HOL

**27 millions** of sequence variants imputed for **8746** cows

# *GWAS & Bayesian analyses*

Within breed single marker ***GWAS*** *with* ***GCTA*** *(Yang et al., 2011)*
**27 millions variants** , analyzed one at a time
Polygenic effects of animals, GRM calculated from HD 631,000 SNP

Selection of the most interesting QTL regions

***Bayesian analyses (BayesC)*** *with* ***GS3*** *(Legarra et al., 2013)*
Within breed, Multimarker (up to 30,000 markers)
Includes also a pedigree-based polygenic effect

# *Bayesian analyses*

Candidate variants were selected according to their **probability of inclusion** (based on 100,000 iterations, burn-in=20,000, thin = 50)

A **difficulty**: due to **very high LD**, inclusion probability of a region is distributed over many linked variants,
and can be low for individual variants

Inclusion probabilities were **summed over 5kb windows** to detect the largest signals

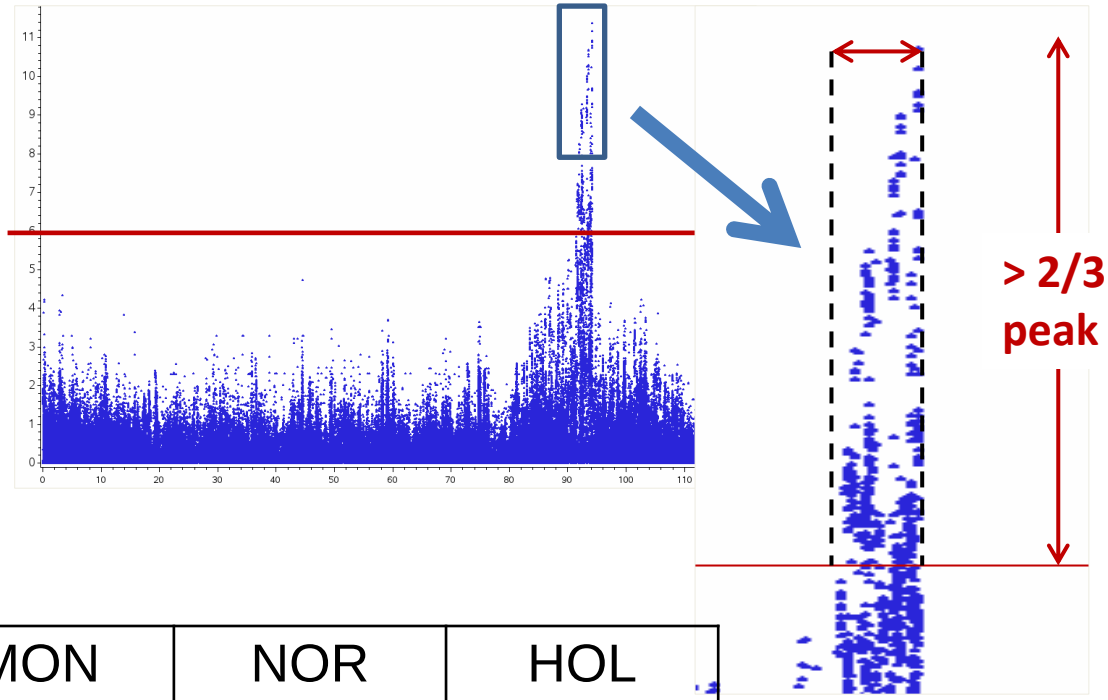
Candidate variants were searched within the best windows

Complementary information : (1) **Across breed comparison**
(2) **Variant Annotation** (1000 bull genomes)

# GWAS results

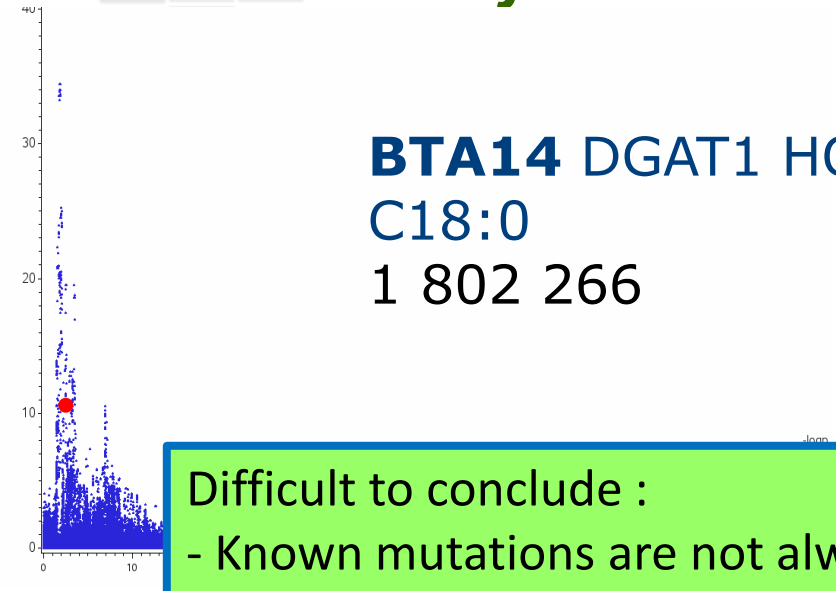
## Number of QTLs

- ✓ $-\log_{10}(p\_value) > 6$
- ✓ 1 QTL maximum in 2 Mb
- ✓ Drop-off value = max(2, 2/3 peak)




| Trait | MON | NOR | HOL |
|---|---|---|---|
| C4:0 | 24 | 30 | 28 |
| C12:0 | 24 | 18 | 31 |
| SAT | 23 | 11 | 26 |
| MONO | 22 | 15 | 20 |
| ω3 | 25 | 21 | 44 |

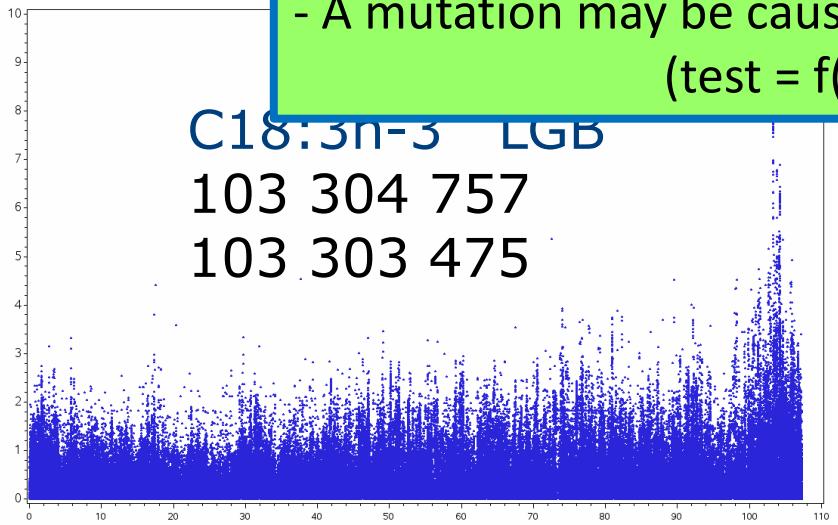
# GWAS results: status of known mutations



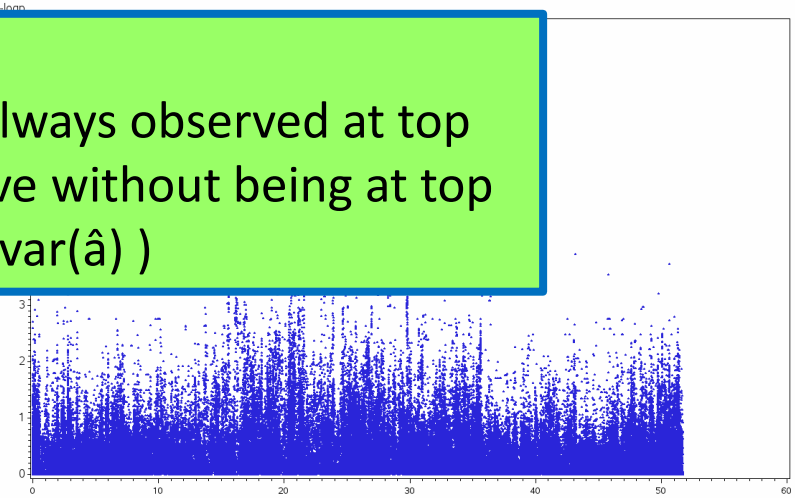
**BTA14** DGAT1 HOL
C18:0
1 802 266

**BTA26** – SCD - MON
C12:0
21 144 708

Difficult to conclude :
- Known mutations are not always observed at top
- A mutation may be causative without being at top
(test = f(â, var(â) )



C18:3n-3 LGB
103 304 757
103 303 475



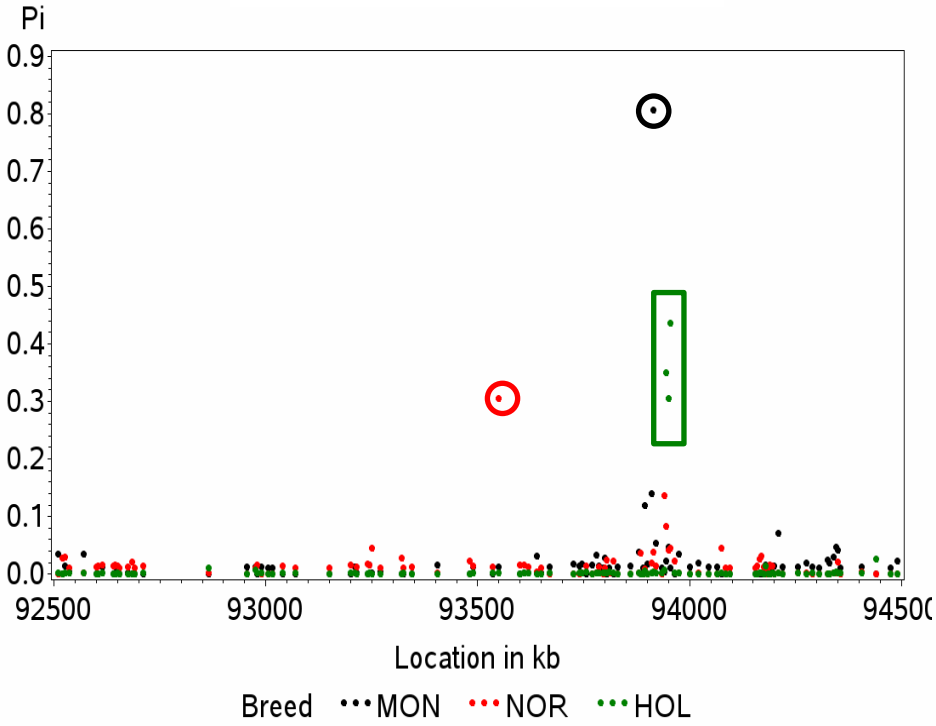
# Results: BayesC

## The regions studied

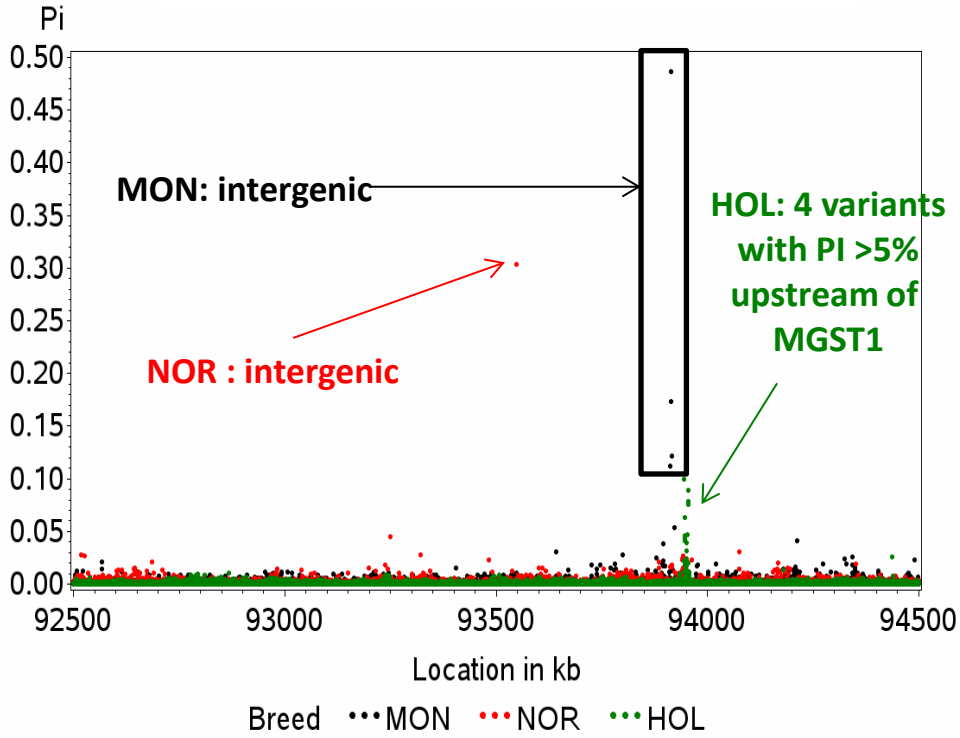
| Chromosome | Région (Mb) | Trait | GWAS test $Log_{10}(1/p)$ max MON – NOR – HOL |
|---|---|---|---|
| **5** | 92.5-94.5 | SAT | 14.2 – 13.2 – 24.4 |
| **14** | 1.3-3.8 | SAT | 34.4 – 79.9 – 169.8 |
| **17** | 52.5-55.0 | C4:0 | 30.2 – 47.8 – 12.3 |
| **19** | 50.0-53.0 | C12:0 | 28.2 – 15.2 – 38.8 |
| **27** | 36.0-36.5 | C16:0 | 16.8 – 9.3 – 9.7 |

# Results: Chromosome 5, SAT

Sum over 5kb windows

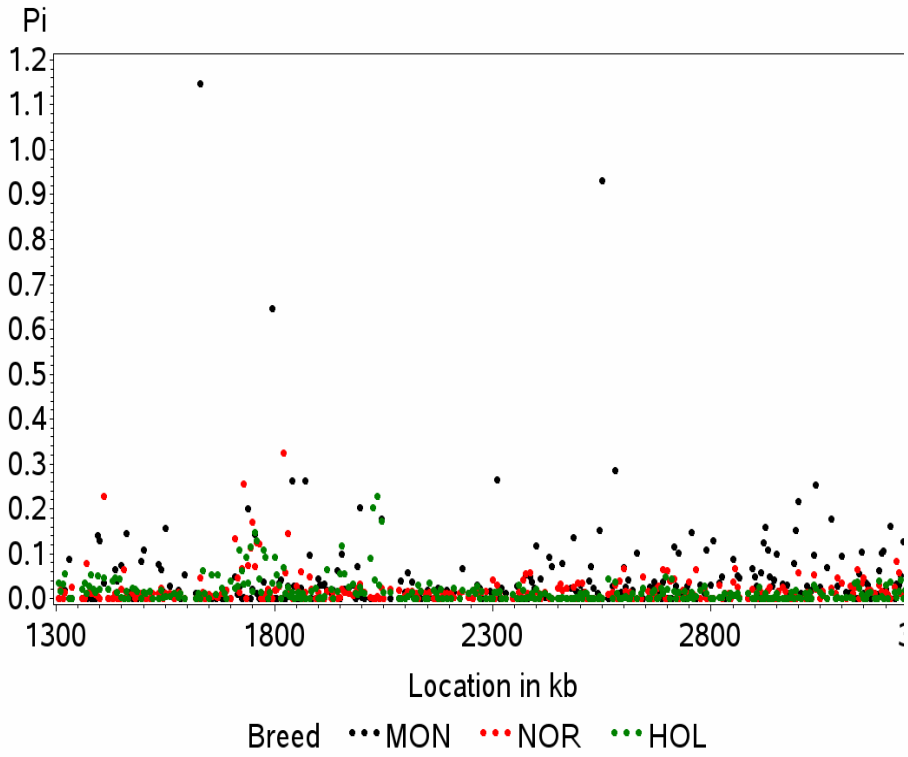



Inclusion probability for each SNP

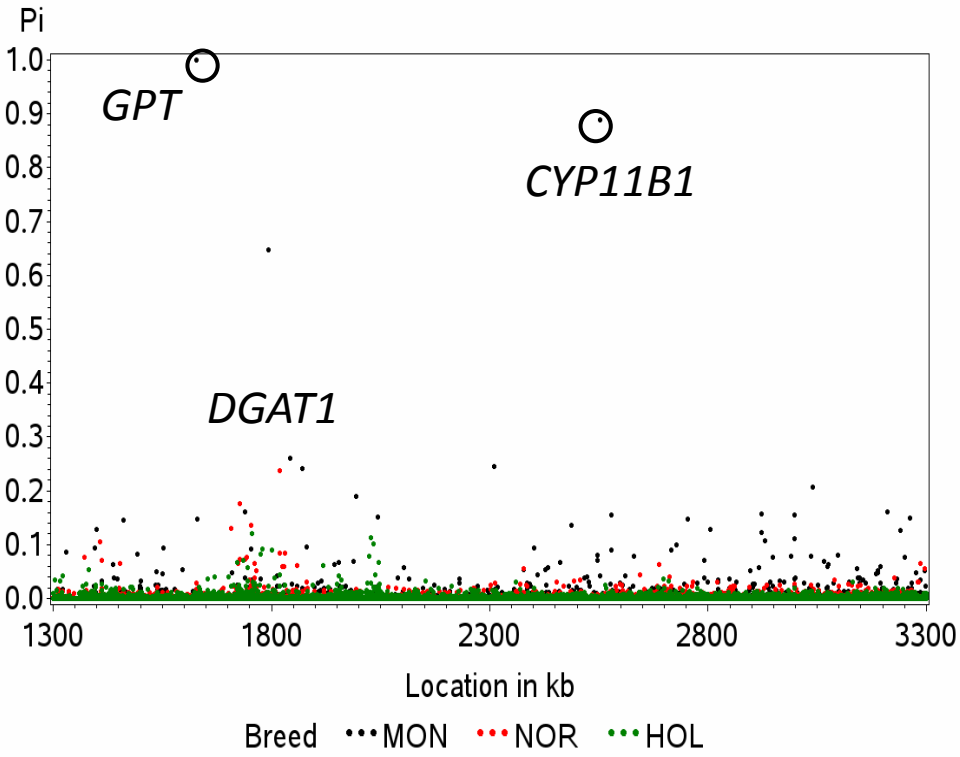


# Results: Chromosome 14, SAT

Sum over 5kb windows




Inclusion probability for each SNP

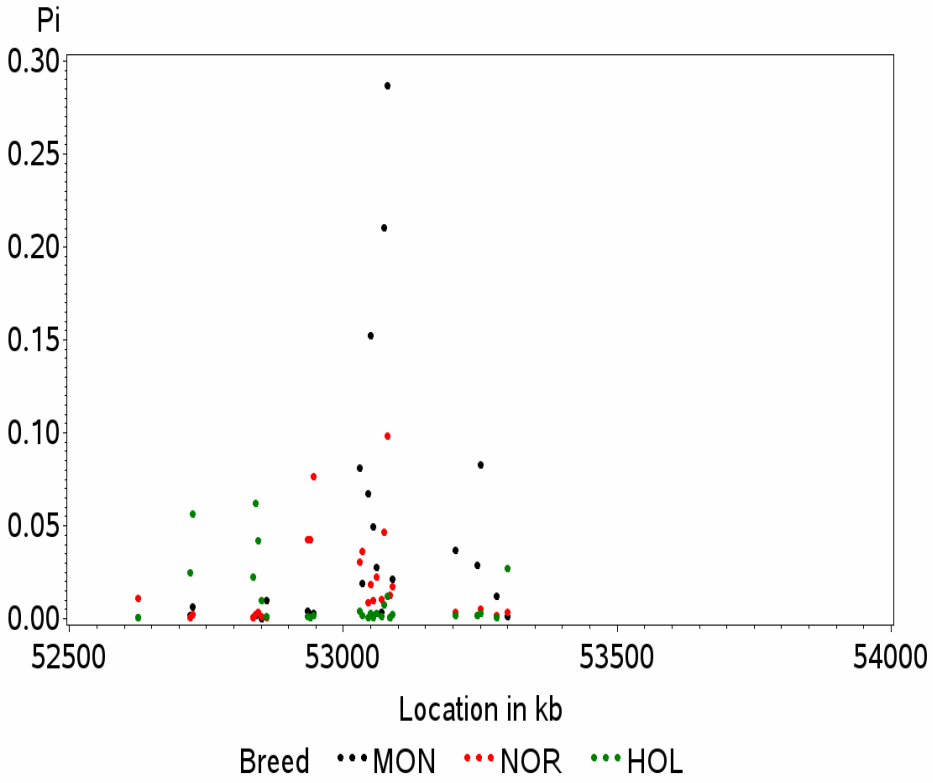



Other mutations than K232A suspected: QTL in populations fixed for K232A and in bulls homozygous for K232A

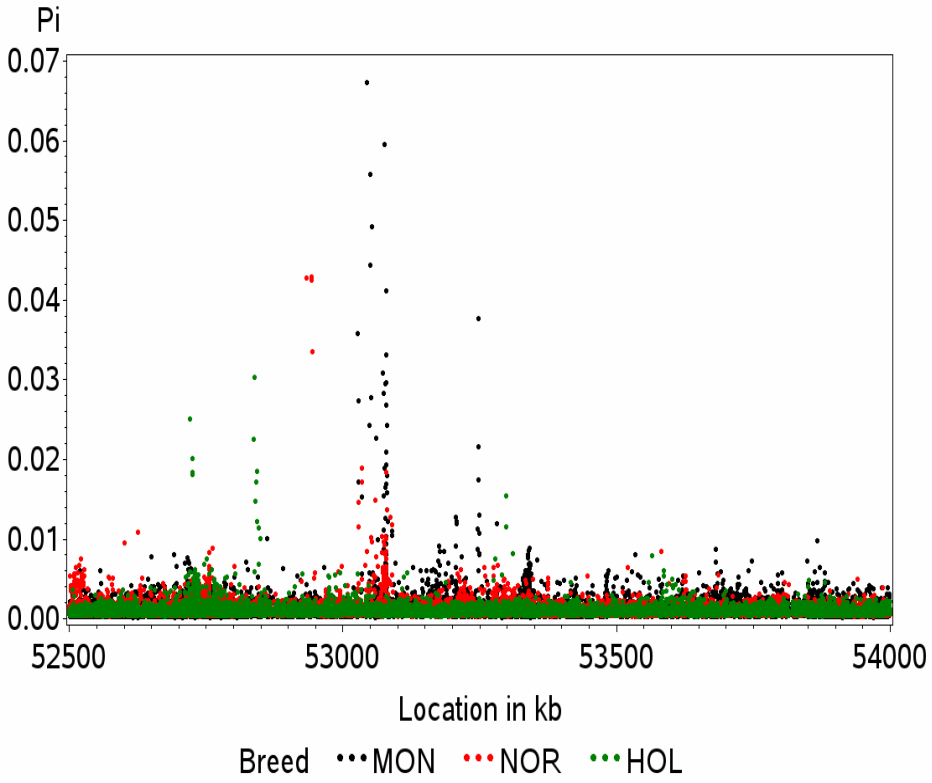
This region seems to be rich, with other mutations in DGAT1(4 ?) and in 2 other genes

# Results: Chromosome 17, C4:0

Sum over 5kb windows



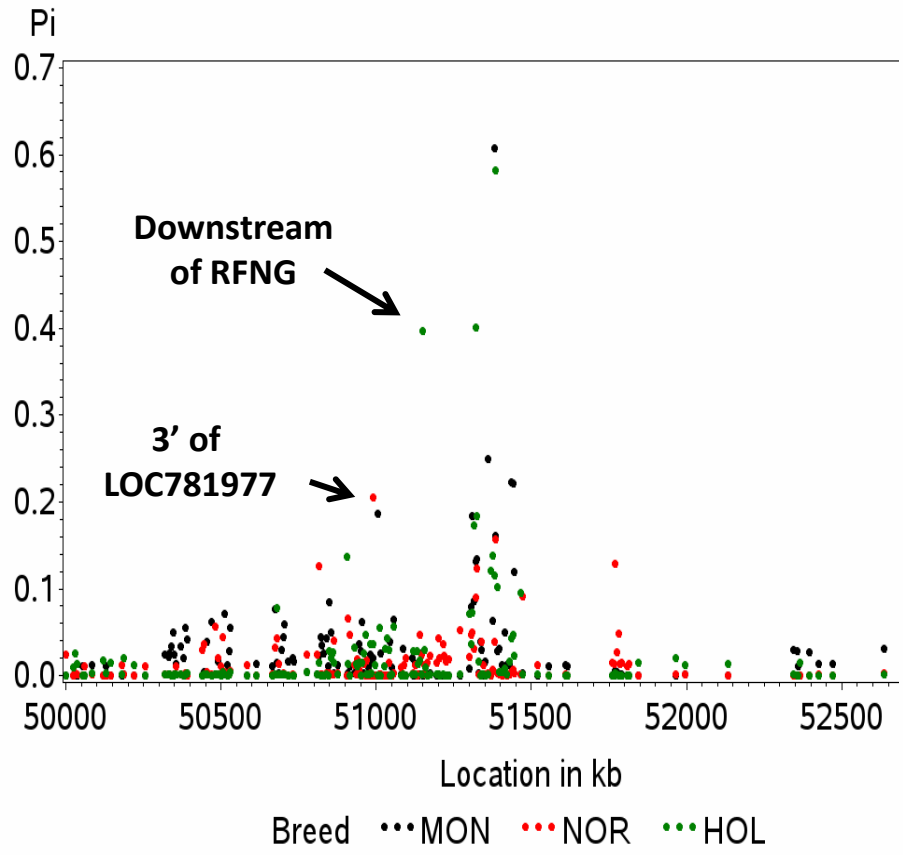

Inclusion probability for each SNP



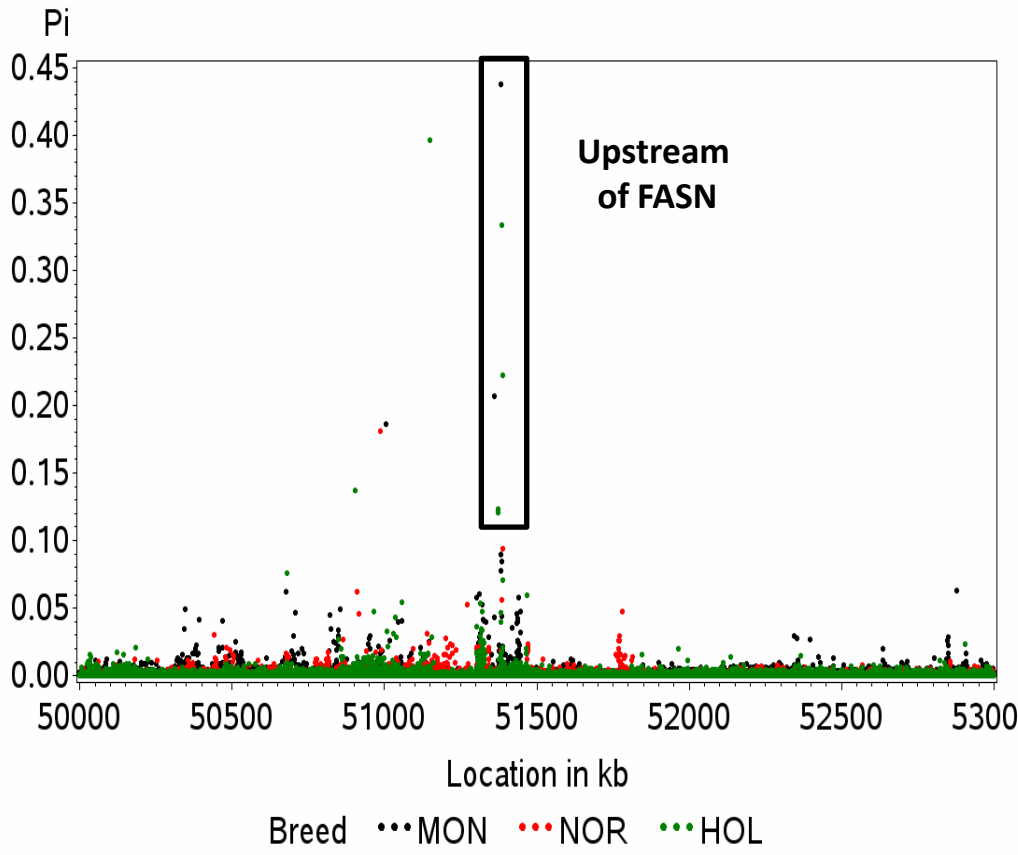

25 markers in very high LD, with similar probabilities, in the BRI3BP gene (all intronic)

# Results: Chromosome 19, C12:0

Sum over 5kb windows




Inclusion probability for each SNP

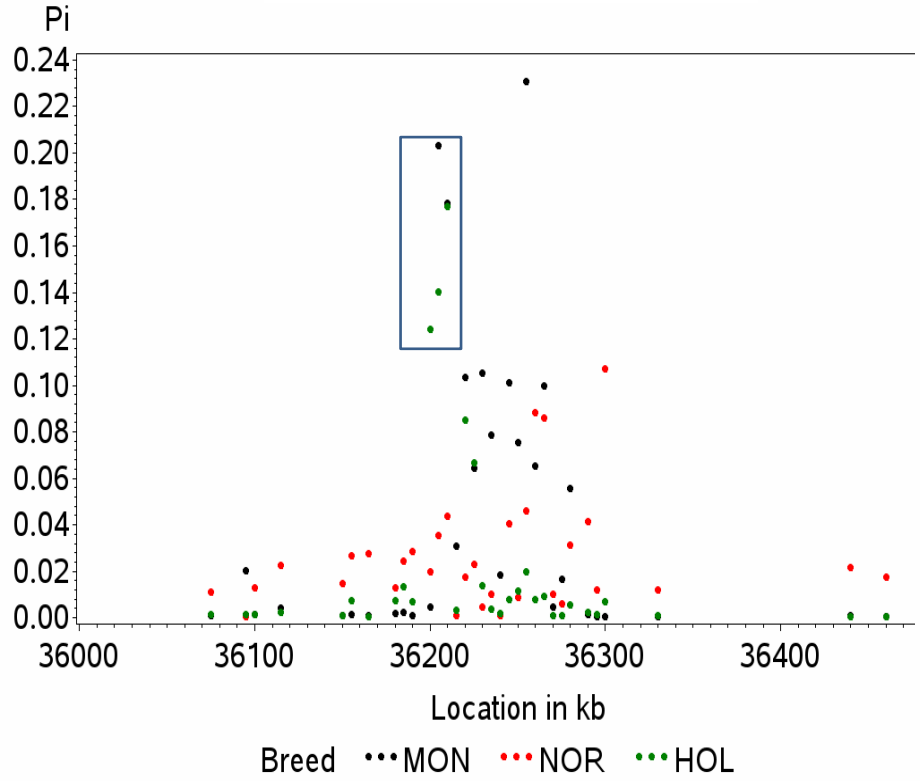



Several genes involved

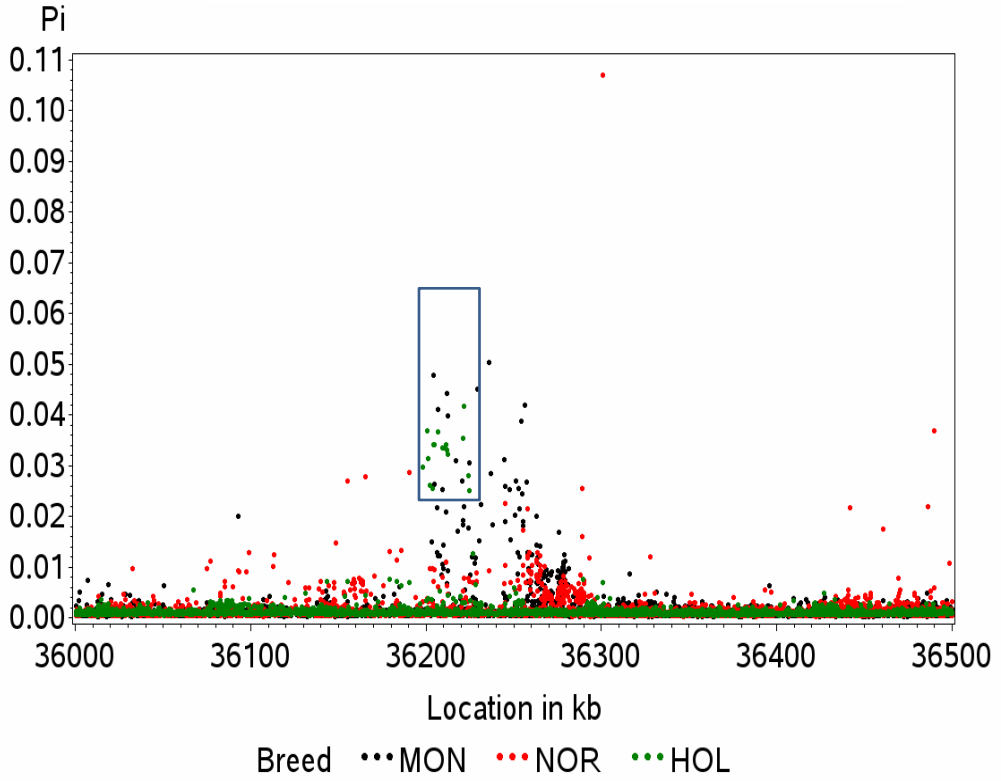
6 candidates in the upstream regulatory region of FASN

# Results: Chromosome 27, C16:0

Sum over 5kb windows




Inclusion probability for each SNP




12 variants very close to each other present PI between 2 and 5% in two breeds
4 with the highest probabilities are upstream of AGPAT6 and are the best candidates

# Results: Summary of results

| BTA | Bounds of peak (kb) | Trait | Candi-date variants | Genes | Annotation of variants in genes |
|---|---|---|---|---|---|
| 5 | 93 940-93 955 | SAT | 4 | MGST1 | Upstream |
| | 1620-1625 | SAT, POLY | 1 | GPT | 3'UTR |
| 14 | 1790-1870 | SAT | 4 | DGAT1 | Various |
| | 2700-2720 | POLY | 4 | CYP11B1 | Upstream / Downstream |
| 17 | 53 075-53 085 | C4:0 | 22 | BRI3BP | Intronic |
| 19 | 51 360-51 385 | C12:0 | 6 | FASN | Upstream |
| 27 | 36 205-36 220 | C16:0 | 4 | AGPAT6 | Upstream |

# Conclusion

BayesC, used to analyze targeted regions, is a good tool to select candidate mutations, in combination with functional annotation

Across breeds, when QTL co-localize, we observed that the same genes are involved

But across breed information is weaker than expected to target candidates. A majority of candidate mutations seem to be breed specific

Invitation: Talk 214 Session 20, Tuesday, 3:30 : Identification of causal variants for milk protein composition using sequence data in dairy cattle, by MP **Sanchez** et al

# Acknowledgements

ANR AGENCE NATIONALE DE LA RECHERCHE

APIS-GENE

« The 1000 bull genomes » consortium