Genome wide association study using single and multiple SNP analysis

Ghyslaine Schopen

Mario Calus, Marleen Visker, Johan van Arendonk and Henk Bovenhuis

QTL-MAS workshop, 17-18 May, Poznan, Poland

Background

- Single marker analysis has been widely used.
- Disadvantages:
  - calculation of variance explained by QTL not straightforward.
  - not all genetic variance captured.
  - multiple testing.

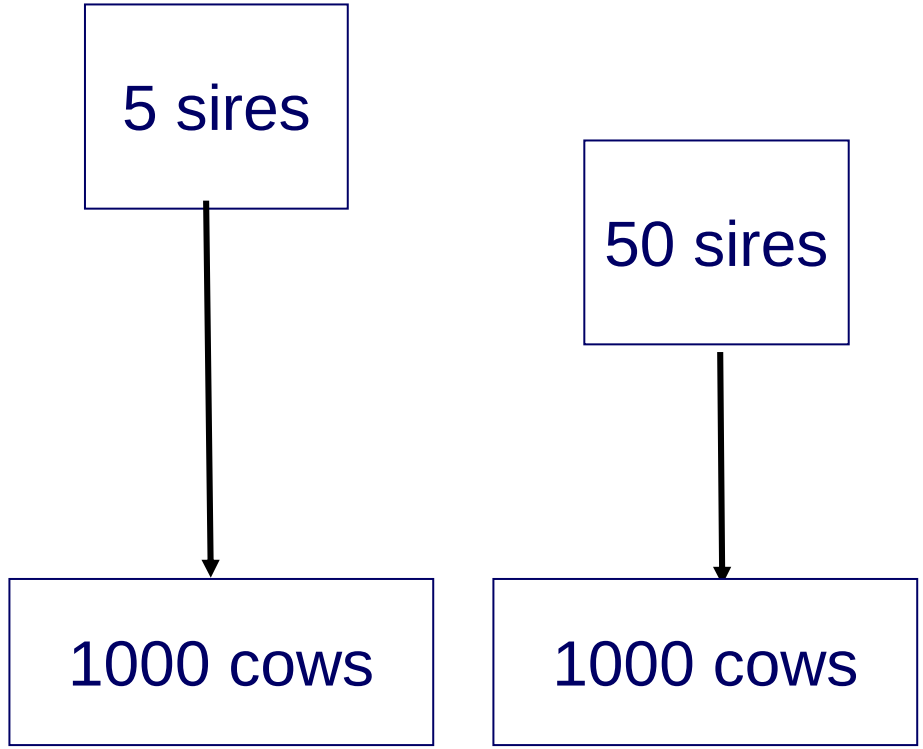
Background

- Fitting more markers simultaneously into the model.
- More power to detect QTL.
- Simulation studies comparing single and multiple QTL mapping.
- Using real data is good addition to simulation studies.

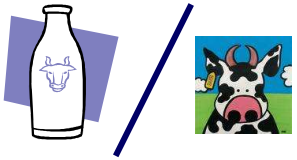
Aim

To compare single SNP analysis with multiple SNP analysis to detect QTL in Dutch Holstein-Friesians.

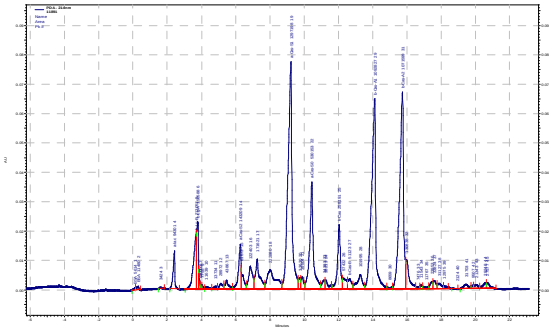
Resource population



Material & Methods

- 1912 
- 398 
- Pedigree
- 1 
- Winter 

- Milk samples
  - Capillary zone electrophoresis



- Blood and semen samples
  - DNA analyses

Phenotypes

- Casein
  - α_{s1}-casein
  - α_{s2}-casein
  - β-casein
  - κ-casein
- Whey
  - β-lactoglobulin
  - α-lactalbumin

Genotypes

- Custom-made 50K SNP designed by CRV.
- 4.857 SNPs excluded
- Finally 45.999 SNPs across the 29 bovine chromosomes.
- Final dataset consisted of 1,713 animals with phenotypes and genotypes.

Single SNP analysis

- General linear model:

$$Y_{ijk} = Sire_i + SNP_j + e_{ijk}$$

where Y was adjusted phenotype, Sire was fixed effect of sire, SNP was fixed effect of SNP, and e was residual

- Animal model:

$$Y_{ij} = \mu + SNP_j + animal_i + e_{ij}$$

where Y was adjusted phenotype, μ was overall mean, SNP was fixed effect of SNP, animal was random additive genetic effect and e was residual

Multiple SNP analysis

- Model (Meuwissen and Goddard, 2004):

$$Y_i = \mu + \sum_{j=1}^{45999} (q_{ij1} + q_{ij2}) v_j + \text{animal}_i + e_i$$

where Y was the adjusted phenotype, μ is the overall mean, v was the scale parameter of the QTL effect of the SNP at putative position j, $q_{1\,(2)}$ was the size of the QTL effect for the paternal (maternal) allele, animal was the random polygenic effect of the animal and e was the residual.

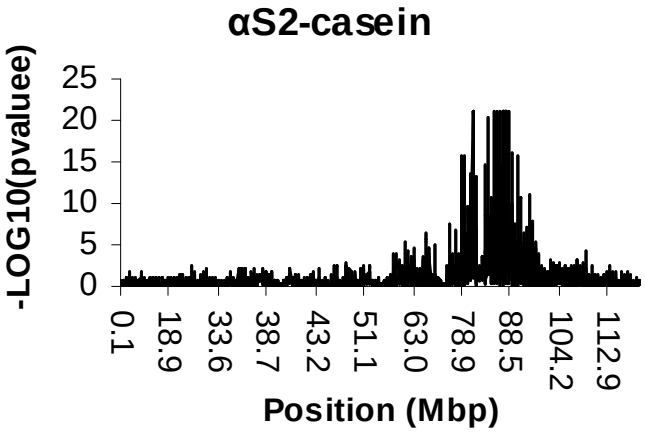
- Markov chain Monte Carlo method using Gibbs sampling:
  - 30,000 iterations
  - burn in 2,000 iterations.

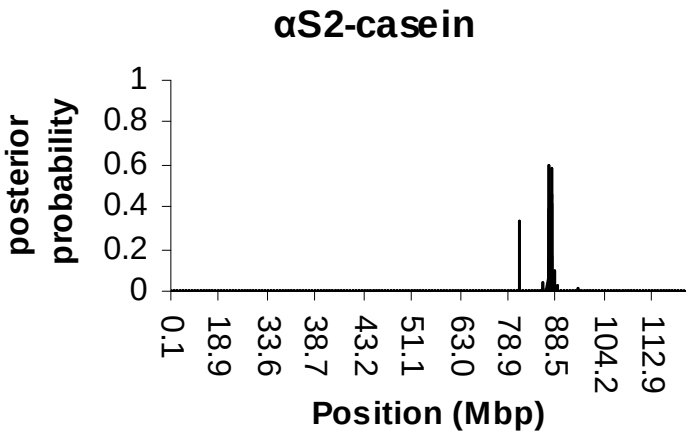
Results

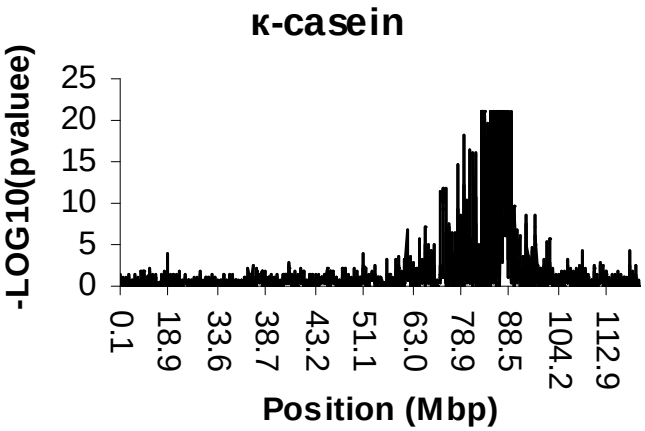
Detected SNPs

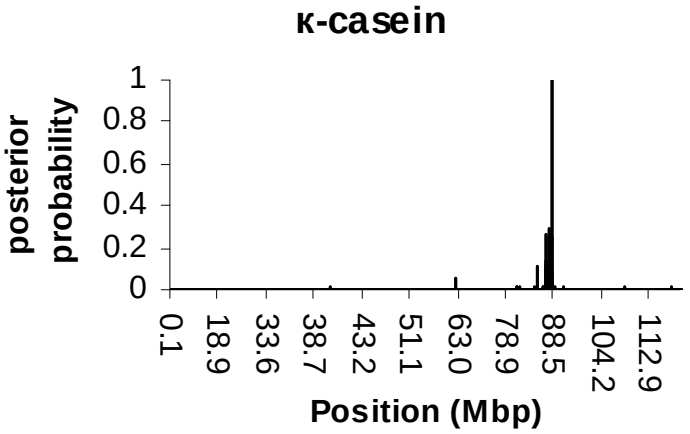
- Same four main regions (chromosome 5, 6, 11 and 14) detected in single and multiple SNP analysis.
- New region detected on chromosome 7 in multiple SNP analysis.
- Additional association detected on chromosome 27 in multiple SNP analysis.

BTA6


αS2-casein
-LOG10(pvaluee)
25
20
15
10
5
0
0.1
18.9
33.6
38.7
43.2
51.1
63.0
78.9
88.5
104.2
112.9
Position (Mbp)


αS2-casein
posterior probability
1
0.8
0.6
0.4
0.2
0
0.1
18.9
33.6
38.7
43.2
51.1
63.0
78.9
88.5
104.2
112.9
Position (Mbp)

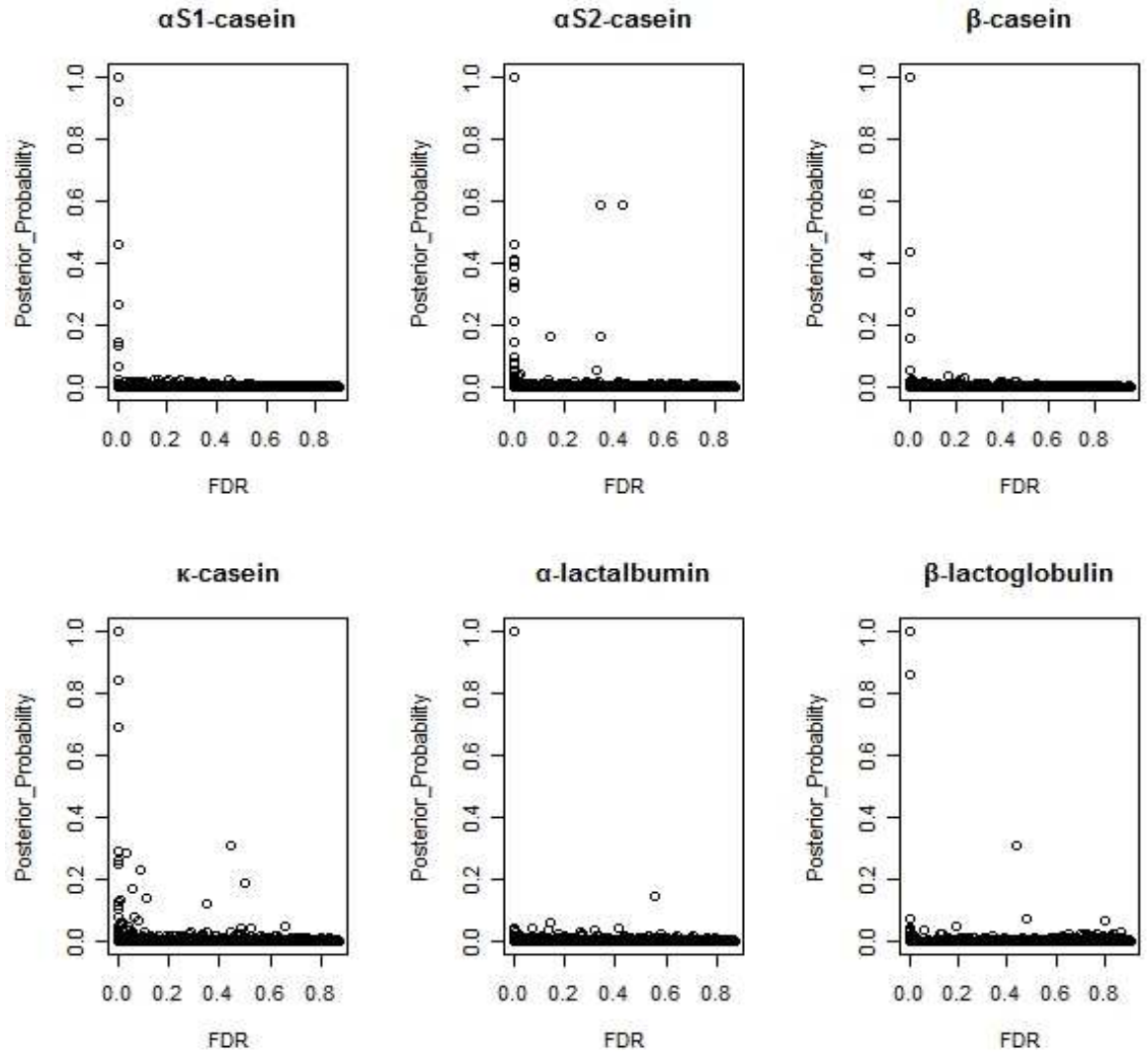

κ-casein
-LOG10(pvaluee)
25
20
15
10
5
0
0.1
18.9
33.6
38.7
43.2
51.1
63.0
78.9
88.5
104.2
112.9
Position (Mbp)


κ-casein
posterior probability
1
0.8
0.6
0.4
0.2
0
0.1
18.9
33.6
38.7
43.2
51.1
63.0
78.9
88.5
104.2
112.9
Position (Mbp)

Variance explained by SNPs

		Single	Multiple						
		VarSNP	VarSNP						
Chromosome	Trait	1	1	10	(Mbp)	20	(Mbp)	50	(Mbp)
5	α-lactalbumin	0.44	0.34	0.34	1.48	0.34	1.85	0.34	2.84
6	αS1-casein	17.02	5.72	5.72	0.88	5.72	1.59	6.34	2.32
6	αS2-casein	16.37	3.40	3.40	0.04	8.90	0.07	11.64	0.53
6	β-casein	42.70	1.88	1.88	0.04	1.88	0.07	1.88	0.33
6	κ-casein	4.11	4.70	4.72	0.26	4.73	0.64	4.86	1.95
6	α-lactalbumin	0.33	0.26	0.26	0.55	0.26	1.14	0.26	2.13
6	β-lactoglobulin	2.15	-	-	-	-	-	-	-
7	κ-casein	-	2.17E-02	0.03	1.08	0.03	1.76	0.03	3.17
11	αS1-casein	4.80	4.14	4.14	0.58	4.14	1.09	4.14	2.46
11	αS2-casein	3.87	2.96	2.96	0.56	2.96	1.04	2.96	2.58
11	β-casein	5.11	0.42	0.42	0.49	0.65	1.21	0.66	2.72
11	κ-casein	0.58	0.26	0.26	0.56	0.26	1.04	0.26	2.58
11	β-lactoglobulin	83.36	79.6	79.56	0.56	80.25	1.04	80.25	2.58
13	κ-casein	0.39	1.31E-01	0.13	0.66	0.13	1.17	0.13	3.11
14	αS1-casein	6.02	3.77	3.77	0.64	3.77	0.73	3.77	1.18
14	αS2-casein	6.14	4.16	4.16	0.64	4.16	0.73	4.16	1.18
14	κ-casein	-	0.02	0.02	0.64	0.02	0.73	0.02	1.18

Comparison FDR and posterior probability



Summary

- The same main four regions on chromosome 5, 6, 11 and 14 were detected in single and multiple SNP analysis.
- Limited number of SNPs identified in multiple SNP analysis compared to single SNP analysis.
- Additional region detected on chromosome 7.
- Multiple SNP analysis results in higher power and mapping precision than single SNP analysis.

Acknowledgements:

- Farmers
- Partners:

- Milk Genomics Team
  - www.milkgenomics.nl

Thank You for your attention!

Multiple SNP analysis results in higher power and in higher mapping precision than single SNP analysis!!

Linkage disequilibrium