Validation of candidate causative variants on milk composition and cheese-making properties in Montbéliarde cows

M.P. Sanchez[1], V. Wolf[2], M. El Jabri[3], E. Beuvier[4], O. Rolet-Répécaud[4], Y. Gauzère[5], S. Minéry[3], M. Brochard[6], S. Fritz[1,7], A. Michenet[1,7], S. Taussat[1,7], A. Barbat-Leterrier[1], A. Delacroix-Buchet[1], C. Laithier[3] & D. Boichard[1]

[1] GABI, INRA, AgroParisTech, Université Paris Saclay, F-78350 Jouy-en-Josas, France
[2] Conseil Elevage 25-90, 25640 Roulans, France
[3] Institut de l'Elevage, 75012 Paris, France
[4] URTAL, INRA, 39800 Poligny, France
[5] ENILBIO, 39801 Poligny, France
[6] Umotest, 01250 Ceyzériat, France
[7] Allice, 75012 Paris, France
marie-pierre.sanchez@inra.fr (Corresponding Author)

Summary

In a previous study, we identified candidate causative variants for milk protein and fatty acid composition. We designed these variants on the custom part of the EuroG10K BeadChip. In order to validate the effects of these candidate variants on milk composition and to estimate their effects on cheese-making properties (CMP), a genome wide association analysis (GWAS) was performed on 30 CMP and milk protein, fatty acid and mineral composition traits predicted from MIR spectra in 19,862 Montbéliarde cows. After genotypes (50K SNP + EuroG10K custom part SNP) have been imputed for all cows, each SNP effect was tested in a mixed linear model including random polygenic effects estimated with a genomic relationship matrix. We confirm here the effects of candidate causative variants located in 18 functional candidate genes on both CMP and milk composition traits. Five of these variants are missense in *ALPL*, *SLC26A4*, *CSN3*, *RECQL4* and *SCD* genes. Seven are located in 5'UTR (*AGPAT6*), 3'UTR (*GPT*) or upstream (*CSN1S1*, *CSN1S2*, *PAEP*, *DGAT1* and *PICALM*) regions and six are located in introns of the *SLC37A1*, *MGST1*, *CSN2*, *BRI3BP*, *FASN* and *ANKH* genes.

Keywords: Montbéliarde, cheese-making properties, milk composition, candidate variants

Introduction

Cheese-making properties (CMP) are strongly related to bovine milk composition (Wedholm *et al.*, 2006). In the PhénoFinlait project, protein and fatty acid milk composition was predicted using mid-infrared (MIR) spectrometry in the three main French dairy cattle breeds. GWAS on whole-genome sequences (imputed using data of the 1000 bull genome project) led to the identification of candidate causative mutations in 24 candidate genes (Sanchez *et al.*, 2017). In order to validate these mutations in an independent population, they were designed on the custom part of the EuroG10K Beadchip. The independent population was obtained from the FROM'MIR project, aiming at analyzing CMP and milk composition (proteins, fatty acids and minerals) predicted from MIR spectra in Montbéliarde cows. In the present study, we validate candidate causative variants evidenced in the PhénoFinlait project on both CMP

and milk composition traits predicted from MIR spectra on 19,862 FROM'MIR Montbéliarde cows.

Material and methods

Cheese-making properties and milk composition

A total of 420 milk samples were collected from 250 individual cows, 100 herds and 70 dairy vats in Protected Designation of Origin and Protected Geographical Indication cheese area of Franche-Comté (Eastern France). Milk samples were aliquoted and analysed within 24h by MIR spectroscopy using MilkoScan FT6000 (Foss, Hillerod, Denmark) and by CMP reference methods for soft and pressed cooked cheese parameters. Three cheese yields, 13 milk rennet coagulation (Formoptic) and 8 milk acidification by lactic bacteria (CINAC) parameters were measured.

Equations of predictions were developed for all the 24 cheese-making criteria using partial least square (PLS) regression. Accuracies of prediction (R^2) ranged from 0.08 to 0.89 according to cheese-making criteria (Laithier *et al.*, 2017). Only 15 criteria with medium to high prediction accuracies were retained for genetic analyses, *i.e.* 3 cheese yields ($0.54 \leq R^2 \leq 0.89$), 9 coagulation ($0.43 \leq R^2 \leq 0.76$) and 3 acidification ($0.39 \leq R^2 \leq 0.62$) traits (Table 1). Moreover, 15 milk composition traits (protein, fatty-acid and mineral contents) were predicted from MIR spectra using prediction equations developed in PhénoFinlait (Ferrand *et al.*, 2012) and Optimir (Gengler *et al.*, 2016) projects.

Prediction equations were applied on about 6 million MIR spectra collected from 330,000 Montbéliarde cows in the Franche-Comté region. Data from cows with at least three test-day records during the first lactation (1,506,037 test-day records from 194,934 cows) were adjusted for non-genetic effects using a mixed model. Herd x test-day x spectrometer and stage of lactation were included in this model as fixed effects while animal genetic and permanent environment effects were assumed random. Data adjusted for fixed effects were then averaged per cow.

Imputation and association analyses

A subset of 19,862 FROM'MIR cows were genotyped for the BovineSNP50 BeadChip (6,505 cows) or for the customized low-density EuroG10K BeadChip (13,357 cows mainly for versions 1 to 5) for routine genomic selection analyses. All genotypes were imputed to the 50K SNP and the custom part SNP of the version 7 of the EuroG10K BeadChip with FImpute software (Sargolzaei *et al.*, 2014) using 177,736 cows genotyped for the BovineSNP50 or EuroG10K (versions 1 to 7) BeadChips. Mean squared correlations (R^2) between imputed and true genotypes reached 91.6% in a validation set for variants with MAF $\geq$ 1%.

Single-trait association analyses were performed between all the polymorphic variants with MAF $\geq$ 1% (45,120 SNP) and the 30 traits (15 CMP and 15 milk composition traits). A mixed linear model was applied with the GCTA software (Yang *et al.*, 2011). It included a mean, the additive fixed effect of the candidate variant and random polygenic effects of animals, estimated with the genomic relationship matrix calculated from the 50K genotypes. The SNP effect was considered significant if its $-\log_{10}(P)$ value estimated assuming a Student distribution was higher than 6 (5% threshold after Bonferroni correction, 0.05/45,120).

Results and discussion

A total of 1069 variants had significant effects on at least one of the 30 traits analysed. Most of them were located in regions previously identified in the PhénoFinlait project for milk protein or fatty acid composition (Sanchez *et al.*, 2017). We confirm here the effects of these regions on both CMP and milk composition traits (proteins, fatty acids and minerals). We found the well-known regions of caseins (BTA6), *PAEP* (BTA11) or *DGAT1* (BTA14) genes as well as other regions on BTA1, 2, 4, 5, 17, 19, 20, 26, 27 and 29. In these regions, candidate variants of the EuroG10K custom part were systematically more significant than the 50K SNP (Figure 1).

We targeted 24 genes (1 to 5 per region) found to be the best candidates in the PhénoFinlait project (Table 2). For each gene, one to 33 candidate variants were present in the custom part of the EuroG10K BeadChip, *i.e.* 245 in total. Among them, 167, with MAF higher than 1%, had significant effects on at least one CMP or milk composition traits and 115 were ranked among the 10 most significant variants (TOP10) for at least one of the traits analysed. Effects of variants located in *BOP1* (7), *MROH1* (13) and *CYPB11* (1) genes on BTA14 were not tested because they had too low MAF (<0.01). Three other candidate genes could be excluded because all their polymorphisms had no significant effects (*ABCG2* and *DHX37*) or because significant variants were not located in the TOP10 of the peak (*GPSM1*).

Eighteen candidate genes, each containing 1 to 29 candidate variants were thus kept for further investigations. Ranks of each variant were then examined in peaks for all traits in order to find the best candidate causative variant in each gene (Table 3). In most of the genes, one variant, reported in Table 2, was ranked at the top of the peak for several traits analysed in this study. Five of these variants were missense in *ALPL*, *SLC26A4*, *CSN3* (κ casein A/B variants), *RECQL4* and *SCD*. Seven were located in 5'UTR (*AGPAT6*), 3'UTR (*GPT*) or upstream (*CSN1S1*, *CSN1S2*, *PAEP*, *DGAT1* and *PICALM*) regions. Finally, six variants were located in introns of the *SLC37A1*, *MGST1*, *CSN2*, *BRI3BP*, *FASN* and *ANKH* genes. Surprisingly, polymorphisms previously found as causal variants in *PAEP* and *DGAT1* genes were not the most significant for any traits in our study. In each of these genes, we identified an upstream variant that was the best candidate in both PhénoFinlait and FROM'MIR cows.

Analyses of both CMP and milk composition traits show that variants with significant effects on CMP were also significant on milk protein, fatty acid or mineral composition. This result therefore confirms the genetic links, previously described via genetic correlations, between CMP and milk composition and in particular with protein composition (Wedholm *et al.*, 2006). Moreover, considering all results together can help to establish the functional link existing between these traits and candidate genes. For example, the best candidate variant in the *SLC37A1* gene, that encodes a glucose-6-phosphate transporter, had significant effects on four CMP traits ($8 \leq -\log(P) \leq 30$) and five milk composition traits ($13 \leq -\log(P) \leq 167$) with the most significant effects obtained on phosphorous. Similarly, the best candidate variant in the *ANKH* gene, encoding an inorganic pyrophosphate transport regulator that helps to prevent the deposition of Ca and P in bones, had significant effects on ten CMP ($7 \leq -\log(P) \leq 46$) and 9 milk composition traits ($30 \leq -\log(P) \leq 175$) with the most significant effects found for α-LA that exhibits a high affinity to Ca. These two examples illustrate the interest to consider fine-scale phenotypes in complement to complex phenotypes, as CMP.

Conclusion

Effects of 13 genomic regions, previously identified on milk composition, are validated on CMP and milk protein, fatty acid and mineral composition predicted from MIR spectra in an independent population of 19,862 Montbéliarde cows. We show that analysing simultaneously fine-scale phenotypes and traits of interest can facilitate the identification of functional candidate genes. We report here candidate causative variants in 18 genes that have functional links with traits studied. In order to explore other genomic regions and to find other candidate variants, a GWAS will be performed on whole-genome sequence variants after imputations with the run 6 of the 1000 bull genome project. GWAS results will then be exploited to search for a set of interacting genes co-associated with CMP and milk composition.

Acknowledgements

This study was funded by the French Ministry of Agriculture, Agro-food and Forest, the French Dairy Interbranch Organization (CNIEL), the Regional Union of Protected Designation cheeses of Franche-Comté (URFAC) and the Regional Council of Bourgogne-Franche-Comté, under the project FROM'MIR.

List of References

Ferrand, M., G. Miranda, S. Guisnel, H. Larroque, O. Leray, F. Lahalle, M. Brochard & P. Martin. 2012. Determination of protein composition in milk by mid-infrared spectrometry. Pages 41-45 in Proc. International Strategies and New Developments in Milk Analysis. VI ICAR Reference Laboratory Network Meeting, Cork, Ireland.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M. Vanrobays, A. Laine, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont & P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. Journal of Dairy Science 99(5):4071-4079.

Laithier, C., V. Wolf, M. El Jabri, P. Trossat, S. Gavoye, D. Pourchet, P. Grosperrin, E. Beuvier, O. Rolet-Répécaud, Y. Gauzère, O. Belysheva, E. Notz & A. Delacroix-Buchet. 2017. Prediction of cheesemaking properties of Montbeliarde milks used for PDO/PGI cheeses production in Franche-Comté by mid-infrared spectrometry. In 12th international meeting on mountain cheese, 20-22 June 2017. Padova, Italy.

Sanchez, M. P., A. Govignon-Gion, P. Croiseau, S. Fritz, C. Hozé, G. Miranda, P. Martin, A. Barbat-Leterrier, R. Letaïef, D. Rocha, M. Brochard, M. Boussaha & D. Boichard. 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genet Sel Evol 49:68.

Sargolzaei, M., J. Chesnais & F. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.

Wedholm, A., L. B. Larsen, H. Lindmark-Månsson, A. H. Karlsson & A. Andrén. 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. J Dairy Sci. 89:3296-3305.

Yang, J., S. Lee, M. Goddard & P. Visscher. 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 88:76-82.

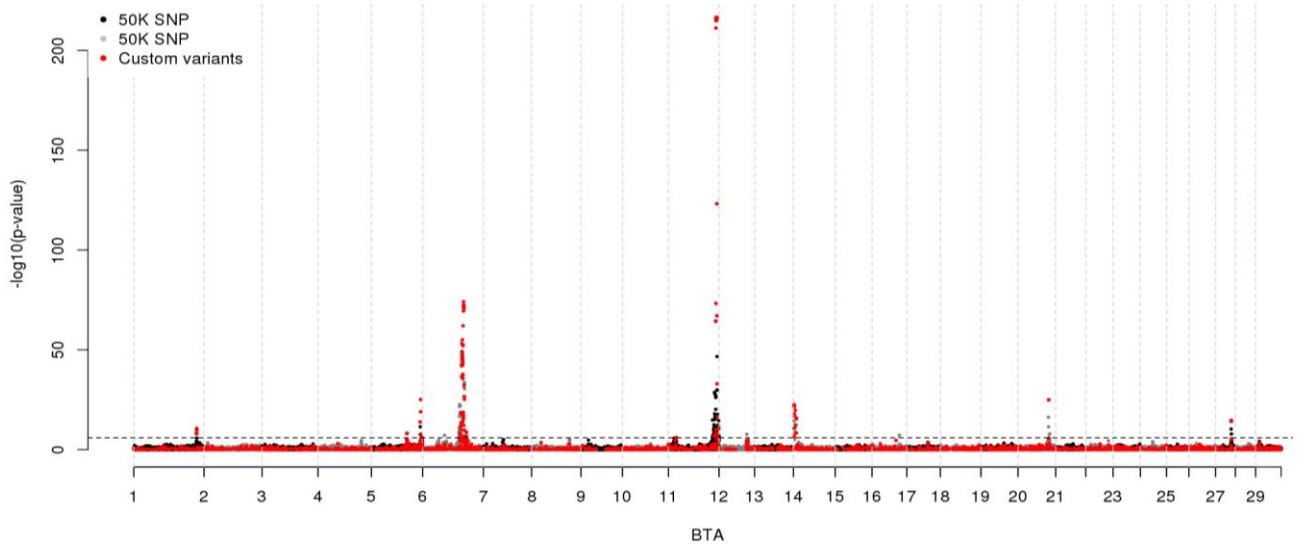

Figure 1. Manhattan plot for cheese yield.

Table 1. Means and standard deviations (SD) of cheese-making properties with medium to high accuracy MIR predictions (R^2_{val}).

Trait[1]		Unit[2]	R^2val[3]	Mean ± SD[4]
Cheese yields (CY)				
100 x (g curd / g milk)	CY_{CURD}	%	0.82	38.7 ± 8.6
100 x (g DM curd / g DM milk)	CY_{DM}	%	0.89	67.6 ± 5.7
(g milk fat + g milk protein) / kg curd	$CY_{FAT\text{-}PROT}$	g/kg	0.54	184.8 ± 23.7
Coagulation (FORMOPTIC)				
Curd firmness at Rennet Coagulation Time (RCT)	a_{SC}	FI	0.76	19.7 ± 3.1
Curd firmness at Rennet Coagulation Time (RCT)	a_{PCC}	FI	0.76	19.3 ± 2.9
Curd firmness at 2 times RCT	$a2_{SC}$	FI	0.69	23.4 ± 2.5
Curd organisation index (time to obtain 10 FI from RCT)	$K10_{SC}$	min	0.43	6.1 ± 1.7
Curd organisation index (time to obtain 10 FI from RCT)	$K10_{PCC}$	min	0.44	11.4 ± 3.5
Curd organisation index standardized for RCT	$K10/RCT_{SC}$		0.72	0.34 ± 0.75
Curd organisation index standardized for RCT	$K10/RCT_{PCC}$		0.68	0.35 ± 0.52
Curd organisation speed (curve slope at 10 FI[1])	$TG10_{SC}$	FI	0.49	14.9 ± 4.5
Curd organisation speed standardized for RCT	$TG10/RCT_{PCC}$	FI/min	0.48	0.25 ± 1.1
Acidification (CINAC)				
Started value of pH	pH_{0SC}	upH	0.45	6.6 ± 0.12
Started value of pH	pH_{0PCC}	upH	0.62	6.5 ± 0.14
upH/h between 170 and 230 min	ΔpH_{PCC}	upH/h	0.39	0.12 ± 0.15

[1] Soft cheese (SC) / Pressed cooked cheese (PCC); [2] Firmness measured by Formoptic in Volts (FV) is converted in Firmness Index FI = 10 x FV; [3]Accuracy of MIR prediction, R^2val = coefficient of determination calculated in the validation set (n=123); [4]Mean ± standard deviation (SD) calculated from 6,146,510 MIR predictions

Table 2. Best candidate variants in 24 candidate genes (R^2 = imputation accuracy).

BTA	Gene	# total variants	# sign. variants[1]	# Top10 variants[2]	Best candidate (bp)	Functional annotation	MAF	R^2	**Variant**
1	SLC37A1	14	12	10	144,398,764	Intronic	0.45	98.4	**1**
2	ALPL	8	7	7	131,812,821	Missense	0.38	95.8	**2**
4	SLC26A4	4	2	2	48,990,317	Missense	0.28	99.1	**3**
5	MGST1	26	19	7	93,945,738	Intronic	0.07	98.9	**4**
6	ABCG2	5	0	0	-	-	-	-	**-**
6	CSN1S1	11	9	4	87,141,456	Upstream	0.30	98.3	**5**
6	CSN1S2	10	7	5	87,261,372	Upstream	0.30	98.3	**6**
6	CSN2	14	10	8	87,187,426	Intronic	0.20	93.7	**7**
6	CSN3	21	17	10	87,390,612	Missense	0.40	95.6	**8**
11	GPSM1	3	3	0	-	-	-	-	**-**
11	PAEP	33	31	29	103,298,431	Upstream	0.45	1	**9**
14	RECQL4	3	3	2	1,617,841	Missense	0.01	87.3	**10**
14	GPT	2	2	2	1,623,927	3'UTR	0.48	96.3	**11**
14	DGAT1	9	1	1	1,795,176	Upstream	0.48	97.7	**12**
14	BOP1	7	0	0	-	-	-	-	**-**
14	MROH1	13	0	0	-	-	-	-	**-**
14	CYP11B1	1	0	0	-	-	-	-	**-**
17	BRI3BP	12	9	9	53,072,959	Intronic	0.06	98.5	**13**
17	DHX37	6	0	0	-	-	-	-	**-**
19	FASN	10	8	5	51,386,735	Intronic	0.37	94.7	**14**
20	ANKH	20	11	3	58,427,343	Intronic	0.07	98.8	**15**
26	SCD	6	5	5	21,144,708	Missense	0.46	96.2	**16**
27	AGPAT6	4	4	4	36,212,352	5'UTR	0.50	96.9	**17**
29	PICALM	3	2	2	9,611,304	Upstream	0.22	95.5	**18**

[1] Variants with MAF $\geq$ 0.01 and $-\log(\text{p-value}) \geq 6$; [2] Variants ranked in the top10 for at least one trait

Table 3. Ranks of the 18 best candidate variants in the peaks identified for each trait.

	Cheese-making properties[2]															Proteins					Fatty acids				Minerals				
Variant[1]	CY_{CURD}	CY_{DM}	$CY_{FAT\text{-}PROT}$	$K10_{PCC}$	$K10/RCT_{PCC}$	a_{PCC}	$TG10/RCT_{PCC}$	$K10_{SC}$	$K10/RCT_{SC}$	a_{SC}	$a2_{SC}$	$TG10_{SC}$	pH_{0PCC}	ΔpH_{PCC}	pH_{0SC}	α-LA	β-LG	αs1-CN	αs2-CN	β-CN	κ-CN	Saturated	Mono-Unsatured	Unsaturated	Poly-Unsaturated	Na	Ca	P	Mg	K
1			5			3							1		1			4								4		2	1	3
2						2	1												5		4					2				
3								2			1		1		1	1		1	1		1						1			
4	4	3	5	18	18			22			22	13						5				1	1	1	1					
5	32	26	47	22	37	51	64	46	42	26	39	43		88			19	28	5		26					9	8		40	46
6	2	3	28	4	3	13	41	23	4	5	3	4					58	50	66	50	38								7	
7	51	50	48	38	38	53	66	55	48	28	45	52		113			20	18	4		19						4		41	52
8	31	28	1	2	11		2	1	2	50	13	2		18		5	2	1	18	3	2									
9	1	1	7	35	5	21	35		5	3	2	2		4		1	12	23	2	18	2	10						9		7
10	58	70	65	45	38		43	26	28		44	45			1	42	35	78		11	40	62	56	57	68		48	56		36
11																						7	5	5	6					7
12	76	55	46	66	77		52	77	77		77	76				85	45	34			68	1	1	1	1		68	77		11
13				1	2		1	1	2		2	1		4					3	4	1									1
14																						1	1	1						
15	4		2			1	4	5		1		4	1	2	1	1		1	2		1						1		1	2
16	1	1		3								1						1	1			12	11	12	15					
17	1	1	4											2	1					1		1	1	1	2				4	
18														2				2	2							2	2			

[1]See Table 2 for descriptions of variants

[2]See Table 1 for descriptions of cheese-making properties traits