Random forests prediction of blood metabolic clusters of dairy cows: comparing three types of milk biomarkers

Leslie Foldager, Miel Hostens, Mazdak Salavati, Clément Grelet, Martin Tang Sørensen, Mark Crowe, Klaus Lønne Ingvartsen & Genotype Plus Environment Consortium

Department of Animal Science, Aarhus University, Blichers Allé 20, DK8830 Tjele, Denmark; Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark; Department of Reproduction, Obstetrics and Herd Health, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium; Royal Veterinary College, London, United Kingdom; Valorization of Agricultural Products Department, Walloon Agricultural Research Center, Gembloux, Belgium; UCD School of Veterinary Medicine, Veterinary Sciences Centre, University College Dublin, Belfield, Ireland

http://www.gpluse.eu/ leslie@anis.au.dk

Milk biomarkers as predictors of physiological imbalances and subclinical or clinical diseases in dairy cows may be used in management and selection. Cows were categorised into three clusters differentiating the metabolic status and defined by clustering of plasma hormone Insulin-like Growth Factor-1 (IGF-1) and three plasma metabolites: Glucose, beta-hydroxybutyrate (BHB), and non-esterified fatty acids (NEFA). We aimed to compare prediction of cluster category by use of milk metabolites and enzymes, selected milk MIR spectra wavenumbers, and milk IgG glycan profiles. Data between calving and 50 days in milk (DIM) were obtained from 235 Holstein Friesian cows in six research herds: AU (Denmark), UCD (Ireland) AFBI (UK), CRA-W (Belgium), FBN (Germany), CREA (Italy). Some diets were designed to challenge the cow and provoke production diseases but there were too few incidences to establish predictions. Instead, k-means clustering was performed over scaled residuals from linear mixed effects models for each of IGF-1, Glucose, log10(BHB) and log10(NEFA) measured around 14 and 35 DIM. Biomarker predictors were: Averages from 2 milkings/week of milk metabolites (free glucose, glucose-6-phosphate, BHB, isocitrate, uric acid, urea) and enzymes (LDH, NAGase) from same week as blood sampling; Closest MIR spectra sample within a limit of +/-3 days from blood sampling dates; IgG glycans sampled 14 and 35 days post-calving used for corresponding period. Random forests (RF) predictions by one biomarker-type and parity (1, 2 and 3+) were evaluated by leave-one-cow-out (internal) cross–validation. Results are shown in Table 1 as proportion of correctly classified observations (overall accuracy, i.e. diagonal of confusion matrix) for each of the three biomarkers and for each period (DIM14 and DIM35), and sensitivity, specificity and balanced accuracy (average of these) for each cluster. We obtained predictions of metabolic clusters from milk biomarkers of fair accuracy. Sample sizes may constrain both estimation of clusters, number of classes, and RF training. External validation in separate herds is warranted. Still, we see potential in using milk biomarkers to support management of dairy cows. Milk metabolites/enzymes resulted in slightly better accuracy of RF predictions but other methods should be investigated too. Practical matters may also influence choice of biomarkers.

Table 1 **Random forests prediction of metabolic clusters established with k-means clustering of centre scaled observed values for each of plasma glucose, plasma log10(BOHB), plasma log10(NEFA), and plasma log10(IGF-1). The class numbers cannot be compared between periods (numbering is arbitrary) but it is the same clusters across the three types of milk biomarker predictors.**

Milk biomarker	Period	Cluster	N	Sensitivity	Specificity	Balanced accuracy	Overall accuracy (95% CI)
Milk enzymes and metabolites	DIM14	Class 1	211	0.54	0.78	0.66	0.59 (0.52-0.65)
		Class 2		0.78	0.69	0.73	
		Class 3		0.33	0.89	0.61	
	DIM35	Class 1	212	0.66	0.72	0.69	0.60 (0.53-0.67)
		Class 2		0.53	0.89	0.71	
		Class 3		0.58	0.77	0.68	
MIR spectres (212 wavenos.)	DIM14	Class 1	197	0.48	0.73	0.60	0.49 (0.42-0.56)
		Class 2		0.52	0.58	0.55	
		Class 3		0.44	0.88	0.66	
	DIM35	Class 1	190	0.55	0.63	0.59	0.47 (0.40-0.55)
		Class 2		0.41	0.79	0.60	
		Class 3		0.44	0.78	0.61	
IgG glycan (19 peaks)	DIM14	Class 1	130	0.51	0.75	0.63	0.52 (0.43-0.61)
		Class 2		0.64	0.68	0.66	
		Class 3		0.38	0.85	0.61	
	35	Class 1	130	0.49	0.49	0.49	0.35 (0.26-0.43)
		Class 2		0.24	0.76	0.50	
		Class 3		0.27	0.74	0.50	

Acknowledgements
This article is based upon work from COST Action FA1308 DairyCare, supported by COST (European Cooperation in Science and Technology, www.cost.eu). COST is a funding agency for research and innovation networks. COST Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.