Tiplady et al. Journal of Animal Science and Biotechnology (2020) 11:39
https://doi.org/10.1186/s40104-020-00445-2

Journal of Animal Science and Biotechnology

REVIEW Open Access

The evolving role of Fourier-transform mid-infrared spectroscopy in genetic improvement of dairy cattle

K. M. Tiplady[1,2*], T. J. Lopdell[1], M. D. Littlejohn[1,2] and D. J. Garrick[2]

Abstract

Over the last 100 years, significant advances have been made in the characterisation of milk composition for dairy cattle improvement programs. Technological progress has enabled a shift from labour intensive, on-farm collection and processing of samples that assess yield and fat levels in milk, to large-scale processing of samples through centralised laboratories, with the scope extended to include quantification of other traits. Fourier-transform mid-infrared (FT-MIR) spectroscopy has had a significant role in the transformation of milk composition phenotyping, with spectral-based predictions of major milk components already being widely used in milk payment and animal evaluation systems globally. Increasingly, there is interest in analysing the individual FT-MIR wavenumbers, and in utilising the FT-MIR data to predict other novel traits of importance to breeding programs. This includes traits related to the nutritional value of milk, the processability of milk into products such as cheese, and traits relevant to animal health and the environment. The ability to successfully incorporate these traits into breeding programs is dependent on the heritability of the FT-MIR predicted traits, and the genetic correlations between the FT-MIR predicted and actual trait values. Linking FT-MIR predicted traits to the underlying mutations responsible for their variation can be difficult because the phenotypic expression of these traits are a function of a diverse range of molecular and biological mechanisms that can obscure their genetic basis. The individual FT-MIR wavenumbers give insights into the chemical composition of milk and provide an additional layer of granularity that may assist with establishing causal links between the genome and observed phenotypes. Additionally, there are other molecular phenotypes such as those related to the metabolome, chromatin accessibility, and RNA editing that could improve our understanding of the underlying biological systems controlling traits of interest. Here we review topics of importance to phenotyping and genetic applications of FT-MIR spectra datasets, and discuss opportunities for consolidating FT-MIR datasets with other genomic and molecular data sources to improve future dairy cattle breeding programs.

Keywords: Bovine milk, Cattle breeding genetics, Fourier-transform infrared spectroscopy, Trait prediction

* Correspondence: kathryn.sanders@lic.co.nz
[1]Research and Development, Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand
[2]School of Agriculture, Massey University, Ruakura, Hamilton 3240, New Zealand

Introduction

Characterisation of milk composition in dairy cattle has a long history of scientific and commercial interest, with many countries establishing formal milk testing programs by the early 1900's [1, 2]. Initial selection targets in these programs were yields of milk or fat, which were measured on a small scale from samples taken manually on farm. Over the course of the twentieth century, advances in refrigeration and transportation technologies, and the availability of automated on-farm milk meters, resulted in a shift to large-scale collection of samples, processed through centralised laboratories, with the scope extended to include quantification of traits such as protein yield and somatic cell counts. More recently, advances in analytical techniques have led to the widespread use of Fourier-transform mid-infrared (FT-MIR) spectroscopy for phenotyping major milk composition traits for dairy improvement programs.

Fourier-transform mid-infrared spectroscopy uses light from the mid-infrared region to scan milk samples and determine the presence of specific chemical bonds. Results are presented as an absorption profile, consisting of the absorbance values for individual infrared light wavenumbers across the mid-infrared region. Traits are predicted as a function of the individual FT-MIR wavenumber absorbances, enabling rapid, high-throughput phenotyping of milk traits such as fat and protein yields, at a fraction of the cost of estimating the components using other methods. Increasingly, there is interest in analysing the individual FT-MIR wavenumbers, and in utilising FT-MIR data to predict other novel traits of interest to the industry, because the spectra are already available as a by-product of routine milk testing. Many of these traits are relevant to consumer expectations and concerns about the nutritional quality of milk, and the impact of dairy production systems on animal health and the environment; and are also relevant to farmers as they seek to improve farming systems and select cows based on their productivity, reproductive performance and disease resistance.

Successful phenotyping using FT-MIR data is dependent on the magnitude of the phenotypic correlation between the predicted trait and the trait as measured by a benchmarked standard reference method. The successful incorporation of a FT-MIR predicted trait into a breeding program is further dependent on the heritability of the spectral-based predictions and on the genetic correlation between the spectral-based predictions and the trait as measured by the benchmarked standard [3, 4]. Improving our understanding of the genetics underlying the expression of FT-MIR predicted traits of interest is thus highly valuable. Conducting a genome wide association study (GWAS) is a widely used practice for identifying genomic regions that are influencing expression of complex traits, such as those predicted from FT-MIR data. However, linking complex traits, such as those predicted from FT-MIR spectra to specific genetic mechanisms is complex, as the phenotypic expression of traits are a function of a diverse range of molecular and biological mechanisms [5] that can obscure the underlying causal links between genotypes and phenotypes. These mechanisms may be characterised as a set of intermediate omics measures, including sugars, lipids and amino acids in the metabolome, proteins in the proteome, RNA molecules in the transcriptome and DNA in the genome, all of which interact with environmental factors to ultimately determine what is observed at the phenotypic level (Fig. 1).

Establishing causal links between the genome and observed phenotypes may be assisted by employing the individual FT-MIR wavenumbers, and other molecular phenotypes such as those related to the metabolome, chromatin accessibility, transcript levels, and RNA editing. Here we review the shifting role of FT-MIR datasets in dairy cattle improvement as we seek to predict new traits of importance to milk payment and animal evaluation systems. We discuss the broader topics of improving FT-MIR data quality and prediction model accuracy in phenotyping applications; and review existing studies of the genetics of FT-MIR predicted traits and individual FT-MIR wavenumbers. We also discuss opportunities for consolidating FT-MIR spectra datasets with other genomic and molecular data sources, to improve our knowledge of the genetic mechanisms of milk composition and enhance future dairy improvement programs.

Phenotyping applications of FT-MIR spectra

Fourier-transform mid-infrared spectroscopy uses infrared light to scan a milk sample and determine the presence of specific chemical bonds. As the light passes through the sample, it interacts with the molecules present, causing vibrations and rotational changes in the molecular bonds, resulting in absorption of some of the light. The light absorption is typically represented as an absorption spectrum, consisting of the absorbance values for individual infrared light wavenumbers across the mid-infrared range. Traits of interest are subsequently predicted as a function of the individual FT-MIR wavenumber absorbances. Utilising FT-MIR data for the prediction of milk composition and other novel traits has been widely studied and recently reviewed [3, 4, 6, 7]. Other notable FT-MIR research includes studies of individual fatty acids and milk proteins [8, 9], and studies of milk properties related to manufacturing, especially coagulation and other cheese-making properties [10–12]. Further studies have focussed on traits not directly measurable in milk, including those related to pregnancy [13, 14], energy status [15, 16], nitrogen outputs [17] and methane emissions [18–20]. Such applications demonstrate that FT-MIR spectra can be used to predict a

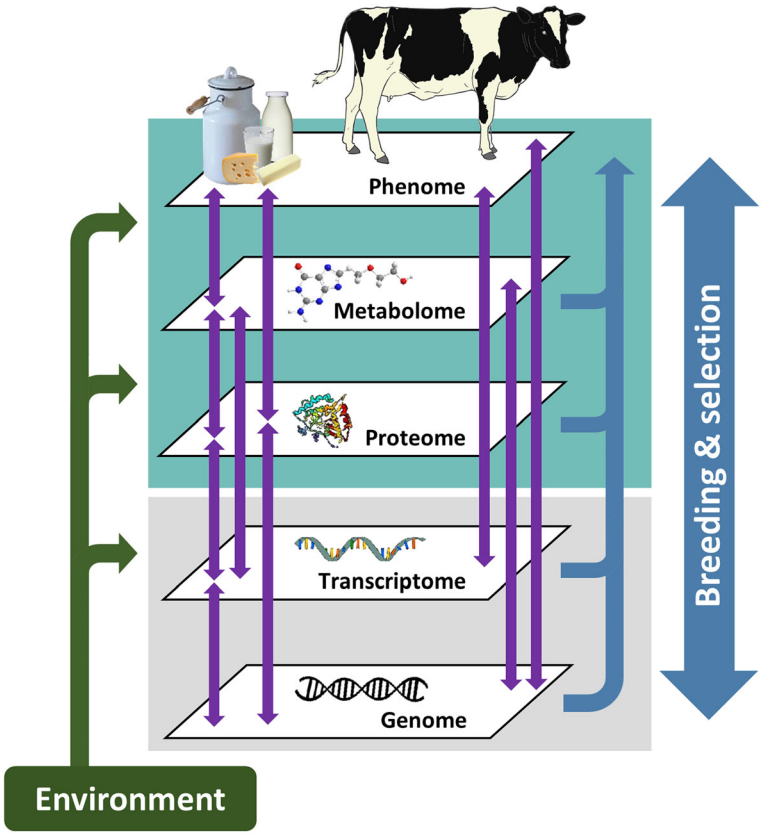

Fig. 1 Characterisation of the relationships between molecular and biological mechanisms underlying phenotypic trait expression

wide range of traits, including highly topical traits that are important to animal welfare and the environment. Whilst prediction accuracy is variable across these applications, a number of key principles and findings have been reported for improving spectra data quality and model prediction accuracy.

FT-MIR data quality and prediction model accuracy

Trait prediction using FT-MIR spectra requires development of a calibration model, typically using a modest set of samples that have corresponding trait values, measured by a benchmarked technique. The most widely used method for developing calibration models from FT-MIR spectra has been partial least squares (PLS) regression. Fewer studies have employed Bayesian methods to develop calibration models [13, 21, 22], but no consensus has been attained as to which methodology is best at providing prediction accuracy [21–23]. This is likely due to the unique set of characteristics of each dataset, and indicates that it is advisable to assess a number of different modelling approaches for any given study. Once a calibration model is developed, the trait of interest can be estimated for any existing spectral absorbance data, or any future milk sample where the FT-MIR spectra data is available. The performance of a FT-MIR calibration model is assessed by how well the model predicts the benchmarked trait measurements in an independent dataset, or within the development dataset using a cross-validation framework. The utility and accuracy of trait predictions from FT-MIR spectra can often be improved by increasing the number of observations used to develop the calibration equation, and by ensuring that a similar extent of the variation in the prediction population is represented in the calibration dataset [24–27]. Prediction accuracy may also be improved by modifying the scale of the trait. For example, higher prediction accuracies have been reported when evaluating fatty acids as a percentage of total milk volume, compared to as a percentage of total fat content [24, 28, 29]. Similar considerations are important for studies of the concentrations of individual casein and whey proteins [28, 30–33]; and in studies related to cheese-making efficiency [28, 34–36]. Other considerations that influence prediction accuracy include: pre-processing treatments to address scaling and baseline effects in spectra data; appropriate management of outliers; low

repeatability of sample measurement for specific regions of the infrared spectrum affected by the water content in milk; and managing systematic instrument variation due to factors such as temperature fluctuations and wavelength or detector intensity instability [37].

Pre-processing

Pre-processing treatments are commonly applied to FT-MIR spectra before generating a calibration model. The objective of pre-processing is to retain important discriminatory features of the spectra, but address baseline and scaling effects caused by light scattering, that can erode prediction accuracy. Baseline effects are additive and represent a baseline offset in the spectral response, whereas scaling effects are multiplicative and scale the spectral results by a given factor. One common group of methods for pre-processing are the multiplicative scatter methods [38, 39]. Multiplicative scatter correction is a normalization method that corrects spectra for scaling and baseline effects by comparing each spectrum to an expected spectral profile. Another family of techniques are the derivation methods, such as the Savitzky-Golay derivative [40]. Derivation methods are based on changes in the spectrum across specified window sizes, and are intended to smooth the spectrum whilst retaining key features of its shape.

Overall, there is no consensus about the best pre-processing treatment to apply to FT-MIR spectra. For example, some studies report that pre-processing spectra provides no significant gains to model prediction accuracy [35, 36], whilst others observe better predictions after pre-processing [27, 41], and several studies report mixed results [30, 33]. This is likely because of the unique characteristics of each dataset, indicating that in the development of a new calibration, it is advisable to compare a number of approaches to determine their effectiveness. Notably, even when different pre-processing strategies are examined in a study, authors often only report the best model, making it difficult to compare the effectiveness of other pre-processing strategies [3].

Outlier removal and removal of low signal-to-noise regions of the MIR spectrum

Outliers in FT-MIR datasets are often identified using a Mahalanobis distance (MD) metric, where the MD is a multivariate indicator of the distance between a spectral record and the average spectral response. Many studies are based on spectra from a single instrument, and are therefore not required to account for the different variance-covariance structures of measurements from different instruments. In a study of spectra from 66 instruments, Grelet et al. [42] showed considerable variability in the spectral responses of the instruments, while we have also observed that the distribution of MD values can be heterogeneous across instruments [43]. These results highlight the need to apply MD thresholds within instrument for the purpose of outlier removal.

Bands of the infrared spectrum with low repeatability of sample measurement due to the water content in milk are typically reported in the O–H bending (~ 1600 to 1700 cm^{-1}) and O–H stretching bands (>~ 3000 cm^{-1}). These regions have low signal-to-noise ratios, with varying boundaries reported across publications: 1600 to 1700 cm^{-1} and 3040 to 3470 cm^{-1} [30]; 1586 to 1698 cm^{-1} and 3052 to 3669 cm^{-1} [44]; 1600 to 1689 cm^{-1} and 3008 to 5010 cm^{-1} [42]. Although it is common practice to remove spectra from low signal-to-noise ratio regions, some studies indicate that there may be wavenumbers within these regions that carry valuable information. For example, Wang et al. [45, 46] identified wavenumbers in these regions that are affected by a polymorphism in the *DGAT1* gene, and Toledo-Alvarado et al. [13] identified a significant association between the 3683 cm^{-1} wavenumber and pregnancy status. More generally, Bittante and Cecchinato [44] showed that the transmittance of individual spectra wavenumbers had moderate to high heritability across most of the mid-infrared region and highlighted that absorbance peaks for non-water milk components were present in low signal-to-noise ratio regions and should be considered for investigation. The findings of these studies indicate that a prudent approach to removal of wavenumbers in low signal-to-noise ratio regions should be taken, retaining spectra from all regions in applications where the wavenumbers are considered independently, but removing them in applications where wavenumbers are considered in a multivariate manner [43].

Managing systematic instrument variation

The instrument calibration approach outlined by Lynch et al. [47] has been widely used to standardize instrument predictions for major milk composition traits and reduce the impact of systematic variation between and within instruments across time. With this approach, a small set of reference samples are analysed through the instrument, where the reference samples have also been measured for traits of interest using benchmarked standards, such as the Rose-Gottlieb method for fat determination and the Kjeldahl method for protein determination. For these samples, unadjusted trait predictions are made from the spectra data, and instrument-specific correction coefficients are evaluated by comparing the unadjusted predictions to the measured trait values according to the benchmarked standard. A limitation of this approach is that it can only be used to adjust predictions for traits with pre-evaluated correction coefficients. More recent standardisation strategies have instead proposed calibrating the individual wavenumbers [42, 43, 48, 49], allowing the correction of any trait predicted as

a function of the spectral wavenumbers. Studies have shown that standardising individual wavenumbers can effectively reduce prediction errors when transferring calibration models between instruments for fat composition traits [42, 48], as well as for calibration models for traits that are more difficult to predict reliably such as methane emissions and cheese yield [49].

Tiplady et al. [43] showed that the most consistent standardisation approach for reducing prediction errors relies on analysing identical reference samples across all instruments, as outlined by Grelet et al. [42]. Ideally, global reference sample sharing would be established, facilitating standardisation across instruments in different countries. That would enable the consolidation of spectral data collected on different instruments, and improve accuracy when applying calibration models developed on one instrument to spectral data collected on other instruments. Global reference sample sharing, however, is reliant on resolving issues related to sample preservation, and on adherence to the bio-security legislation of different countries. Instrument manufacturers such as Foss (Hillerød, Denmark) and Bentley (Chaska, MN) have started to offer alternative standardisation procedures. The Foss procedure uses a liquid equaliser with a known spectral response to adjust spectral results [50], whereas the Bentley procedure uses a polystyrene film to adjust for interferometer laser frequency shifts across time [51], and infrared flow cell information to adjust for shifts in absorbance measurement [52]. While these within-instrument standardisation procedures offer promise for automatic spectral standardisation, there have been no independent studies to validate their effectiveness for standardisation of milk samples collected across or within networks.

The genetics of FT-MIR predicted traits

Predictions of major milk composition traits from FT-MIR spectra are already widely incorporated into dairy improvement programs. Other FT-MIR predicted traits that could be of interest to industry improvement programs include milk fatty acids and protein fractions, and traits that form proxy indicators for milk processability properties, and animal health and environmental outcomes. The accuracy of FT-MIR predictions is an important indicator of their utility, but for breeding purposes, the critical parameters are the extent of genetic variation in the benchmarked trait, the heritability of the FT-MIR predictions, and the genetic correlations between the FT-MIR predictions and the benchmarked trait.

Milk fatty acid and protein composition traits

Heritability estimates for FT-MIR predicted individual and grouped fatty acids, and their genetic correlations with gas chromatography (GC) based measurements are shown in Table 1. Where available, standard errors are shown in brackets. For individual milk fatty acids, heritability estimates ranged from 0.05 to 0.54 [9, 53, 54]. Heritability estimates for grouped fatty acids ranged from 0.11 to 0.51 [9, 55–57], with the lowest heritability estimates reported by Hein et al. [55]. In the studies by Fleming et al. [56] and Narayana et al. [57], heritability estimates were consistently higher for saturated fat and short- and medium-chain fatty acid groups, compared to unsaturated fat and long-chain fatty acid groups. Rutten et al. [54] was the only study to report genetic correlations between the FT-MIR predicted and GC-based fatty acids. These genetic correlations were high, ranging from 0.82 to 0.99.

Fewer studies exist of the genetic parameter estimates of FT-MIR predicted individual milk proteins. Sanchez et al. [58] reported moderate to high heritability estimates (0.25 to 0.72) for a number of FT-MIR predicted milk protein contents/fractions (not shown), with especially high estimates for β-lactoglobulin (0.61 to 0.86). Moderate heritability estimates for FT-MIR predicted lactoferrin, ranging from 0.16 to 0.22 have also been reported [59–61]. These studies quantify the useful extent of genetic variation in FT-MIR predicted fatty acids and individual milk proteins, and suggest that these predicted traits could be incorporated into cattle improvement programs to change the fatty acid profile and the protein composition of bovine milk.

Milk processability traits

Heritability estimates and genetic correlations between measured and FT-MIR predicted milk processability traits are shown in Table 2. Where available, standard errors are shown in brackets. For coagulation traits, heritability estimates ranged from 0.16 to 0.43 [62–64]. Cecchinato et al. [63] was the only study reporting genetic correlations between FT-MIR predicted and measured coagulation traits (not shown). Those ranged from 0.91 to 0.96 for rennet coagulation time (RCT), and from 0.71 to 0.87 for curd firmness after 30 min (a_{30}). Heritability estimates for FT-MIR predicted minerals ranged from 0.32 to 0.56, with phosphorus having the highest estimated heritability, and sodium having the lowest estimated heritability in both studies presented [64, 65]. Heritability estimates for nutrient recovery traits were typically higher than for cheese yield traits [66, 67]. Bittante et al. [66] was the only study reporting genetic correlations between FT-MIR predicted and measured cheese yield and nutrient recovery traits. These ranged from 0.76 to 0.98 for cheese yield traits, and from 0.79 to 0.98 for nutrient recovery traits. Overall, these studies show that many FT-MIR predicted processability traits are heritable, and that sufficient variation exists to use

Table 1 Heritability estimates for FT-MIR predicted fatty acids (h^2), and their genetic correlations (r_a) with GC-based[a] fatty acids

Individual fatty acids[b]	Lopez-Villalobos et al. [9]	Soyeurt et al. [53]	Rutten et al. [54]	
	h^2 (SE)	h^2 (SE)	h^2 (SE)	r_a (SE)
C4:0	0.38 (0.03)	–	0.42 (0.09)	0.94 (0.03)
C6:0	0.32 (0.03)	–	0.35 (0.09)	0.97 (0.02)
C8:0	0.29 (0.03)	–	0.38 (0.09)	0.99 (0.01)
C10:0	0.17 (0.02)	–	0.46 (0.10)	0.98 (0.01)
C10:1	0.30 (0.02)	–	–	–
C12:0	0.16 (0.02)	0.29 (0.02)	0.54 (0.11)	0.97 (0.02)
C12:1	0.41 (0.03)	–	–	–
C14:0	0.19 (0.02)	0.31 (0.03)	0.50 (0.10)	0.99 (0.01)
C14:1	0.27 (0.01)	–	–	–
C15:0	0.22 (0.02)	–	–	–
C16:0	0.29 (0.02)	0.38 (0.02)	0.30 (0.09)	0.86 (0.07)
C16:1	0.30 (0.02)	–	–	–
C17:0	0.41 (0.03)	–	–	–
C17:1	0.14 (0.02)	–	–	–
C18:0	0.26 (0.02)	0.30 (0.02)	0.52 (0.10)	0.82 (0.08)
C18:1	0.43 (0.03)	0.05 (0.01)	–	–
C18:1 *cis*-9	0.22 (0.02)	–	0.25 (0.08)	0.93 (0.05)
C18:1 *trans*-11	0.27 (0.03)	–	–	–
C18:2 *cis*-9, *cis*-12	0.45 (0.03)	0.20 (0.02)	–	–
C18:2 *cis*-9, *trans*-11	0.41 (0.03)	–	–	–
C20:0	0.38 (0.03)	–	–	–
C20:1 *cis*-11	0.37 (0.03)	–	–	–
C22:0	0.35 (0.03)	–	–	–

Grouped fatty acids[c]	Lopez-Villalobos et al. [9]	Hein et al. [55]	Fleming et al. [56]	Narayana et al. [57]
	h^2 (SE)	h^2	h^2	h^2
SCFA	0.39 (0.03)	0.16	0.42	0.24
MCFA	0.30 (0.03)	0.12	0.50	0.32
LCFA	0.50 (0.03)	0.11	0.26	0.23
SFA	0.46 (0.03)	0.15	0.51	0.33
UFA	0.48 (0.03)	–	0.26	0.21
PUFA	0.42 (0.03)	–	–	–

[a]*GC-based* Gas chromatography based.
[b]All fatty acids expressed as a % of the total fatty acids.
[c]*SCFA* Short-chain fatty acids, *MCFA* Medium-chain fatty acids, *LCFA* Long-chain fatty acids, *SFA* Saturated fatty acids, *UFA* Unsaturated fatty acids, *PUFA* Polyunsaturated fatty acids.

FT-MIR predicted traits to change milk processing and cheese-making characteristics in cattle improvement programs.

Animal health traits

Health and fertility traits are valuable targets for breeding programs and selection for these traits would be considerably enhanced if they could be reliably predicted from FT-MIR. A recent review by Bastin et al. [68] across a wide range of FT-MIR predicted traits related to fertility, mastitis, ketosis and other disease traits highlighted that more research is required to understand the relationships between health and fertility indicators and FT-MIR predicted traits, and to estimate the genetic parameters of these traits. Since then, Belay et al. [69] have reported moderate heritability estimates for FT-MIR predicted blood β-hydroxybutyrate (BHB), ranging from 0.25 to 0.37 across different stages of lactation, and moderate genetic correlations between clinical ketosis and the FT-MIR predicted BHB (0.47). More research is

Table 2 Heritability estimates of FT-MIR predicted milk processability traits (h^2), and their genetic correlations (r_a) with measured traits

Trait[a]	Visentin et al. [62]	Cecchinato et al. [63]	Costa et al. [64]	Sanchez et al. [65]
	h^2 (SE)	h^2 range[b] (SE)	h^2 (SE)	h^2 (SE)
Coagulation traits				
RCT, min	0.28 (0.01)	0.30–0.34 (0.08)	0.35 (0.05)	–
k_{20}, min	0.43 (0.02)	–	0.43 (0.03)	–
a_{30}, mm	0.36 (0.02)	0.22–0.27 (0.07)	0.39 (0.03)	–
a_{60}, mm	0.27 (0.01)	–	–	–
HCT, min	0.16 (0.01)	–	–	–
CMS, nm	0.31 (0.02)	–	–	–
Acidity				
pH, units	0.27 (0.01)	–	–	0.37 (0.01)
Minerals, mg/kg milk				
Calcium	–	–	0.45 (0.02)	0.50 (0.01)
Phosphorus	–	–	0.53 (0.03)	0.56 (0.01)
Magnesium	–	–	0.47 (0.03)	0.52 (0.01)
Potassium	–	–	0.45 (0.03)	0.53 (0.01)
Sodium	–	–	0.38 (0.03)	0.32 (0.01)
	Sanchez et al. [65]	Bittante et al. [66]		Cecchinato et al. [67]
	h^2 (SE)	h^2 (SE)	r_a	h^2 range[c]
Cheese yield, %				
CY_{CURD}	0.38 (0.01)	0.21 (0.09)	0.97	0.18–0.33
CY_{SOLIDS}	0.39 (0.01)	0.22 (0.08)	0.98	0.18–0.28
CY_{WATER}	–	0.18 (0.05)	0.76	0.14–0.29
Nutrient recovery, %				
$REC_{PROTEIN}$	–	0.44 (0.09)	0.88	0.32–0.41
REC_{FAT}	–	0.28 (0.07)	0.79	0.15–0.33
REC_{ENERGY}	–	0.21 (0.07)	0.96	0.19–0.30
REC_{SOLIDS}	–	0.24 (0.08)	0.98	0.17–0.29

[a]*RCT* Rennet coagulation time; k_{20} = curd-firming time; a_{30} = curd firmness after 30 min; a_{60} = curd firmness after 60 min; HCT = heat coagulation time; CMS = casein micelle size; CY: weight of fresh curd, curd solids, and curd as a percentage of weight of milk processed; REC: protein, fat, energy and solids of the curd as a percentage of the protein, fat, energy and solids of the milk processed.
[b]Range of estimates from 4 subsets of data used to validate calibration equations.
[c]Range of estimates from 3 different breeds.

required in this area to realise the value that FT-MIR spectra might add to animal health breeding goals.

Environment traits

Despite increasing interest in FT-MIR predictions of environmental traits related to methane (CH_4) and nitrogen outputs from dairy systems, there have been few reports of the genetic parameter estimates of these FT-MIR predicted traits, or of the genetic correlations between measured and FT-MIR predicted trait values. Kandel et al. [70] report moderate heritability estimates, ranging from 0.22 to 0.25 for predicted daily CH_4 emission and 0.17 to 0.18 for log-transformed predicted CH_4 intensity. There is, therefore, some potential for the future incorporation of FT-MIR predicted methane traits into breeding programs. However, there are still issues to be resolved to address uncertainties and discrepancies in methane datasets and measurement methods, and to improve the accuracy and robustness of prediction equations to make them applicable across a broader range of production systems and environments [19, 20, 71, 72].

Milk urea nitrogen (MUN) concentrations are routinely predicted using FT-MIR spectroscopy [4], however, there are few studies of the genetic parameters of FT-MIR predicted MUN and its relationship with other production traits. Amongst those studies, moderate to high heritability estimates, ranging from 0.38 to 0.59

were reported by Wood et al. [73] and Miglior et al. [74], with lower estimates of 0.22 and 0.14 reported in studies by Mitchell et al. [75] and Stoop et al. [76], respectively. Mitchell et al. [75] was the only study reporting genetic correlations between wet-chemistry direct measurements of MUN and FT-MIR predicted MUN, which were 0.38 and 0.23 in lactations 1 and 2, respectively. These genetic correlations are significantly lower than those reported for fatty acids (0.82 to 0.99) [54] and milk processability traits (0.76 to 0.98) [66], and indicate that wet-chemistry measurements of MUN and FT-MIR predicted MUN are genetically different traits. Large differences in heritability estimates across studies of FT-MIR predicted MUN indicate that there may be underlying instability in prediction equations. This highlights the importance of developing prediction models that are robust across different breeds and production systems. Research is ongoing to determine the role that FT-MIR predicted MUN could have in reducing nitrogen outputs from dairy systems.

The genetics of individual FT-MIR wavenumbers

In contrast to the prevalence of studies reporting genetic parameter estimates of FT-MIR predicted traits, there are relatively few studies reporting genetic parameter estimates for the individual spectral wavenumbers. Nevertheless, the transmittance of FT-MIR spectra wavenumbers is moderately to highly heritable across a large proportion of the mid-infrared region [44, 45, 77–79]. Although heritability estimates were consistently low in water absorption regions across all studies, estimates > 0.2 were reported across most of the mid-infrared region in studies by Soyeurt et al. [77] and Wang et al. [45]. This indicates that genetic gain may be obtained by directly selecting on a linear function of estimated breeding values (EBV) for individual FT-MIR wavenumbers; rather than indirect selection as currently practised on EBV of composite indicator traits, like fat yield, which are linear functions individual FT-MIR wavenumber absorbances. Recent studies have confirmed this, showing that the accuracies of breeding value predictions estimated directly from FT-MIR spectra can be higher than for breeding value predictions estimated indirectly from the FT-MIR predicted composite traits [80–82]. Estimating breeding values directly from FT-MIR spectra requires that spectral data is routinely stored, rather than just the spectral based predictions of milk components, and that has not historically been the case in most dairy nations.

GWAS of individual FT-MIR wavenumbers

Many GWAS have been published in the last decade for FT-MIR predicted major milk production traits [83–87], and for fatty acid and protein fractions [88–92]. However, only two studies report GWAS results for individual FT-MIR wavenumbers. In a study of 1748 Dutch Holsteins across 50,688 SNP, Wang & Bovenhuis [46] conducted a GWAS on a subset of 50 wavenumbers, selected using a clustering approach to capture more than 95% of the phenotypic variation. In that study, significant associations between individual wavenumbers and over 20 genomic regions were identified. While most of these genomic regions had already been reported for having significant associations with other milk production traits, three new regions were identified. In a larger study of 5202 Holstein, Jersey and crossbred cows across 626,777 SNP, Benedet et al. [93] used a PLS approach to associate genotypes to spectral data, and showed that FT-MIR spectra could be used to increase the power of a GWAS, and assist with distinguishing milk composition QTL. The studies by Wang & Bovenhuis [46] and Benedet et al. [93] both demonstrate that there are genetic signals in the individual FT-MIR wavenumbers that we do not observe in the currently-used portfolio of composite FT-MIR predicted traits. This confirms that the individual FT-MIR wavenumbers can provide an additional layer of granularity to assist with establishing causal links between the genome and observed phenotypes. Notably, both studies use relatively low numbers of animals compared to recent GWAS published for other traits, and applying these methodologies to larger datasets, with higher genotype densities, promises to increase the power of these approaches. This should enable the discovery of QTL with smaller effect sizes in addition to novel QTL characterised by lower minor allele frequencies than those QTL discovered with datasets numbering only thousands of animals.

Computational challenges

Over the last two decades, the scope of genomic resources available for GWAS has increased, both in terms of the number of genotyped individuals, and in terms of variant density. Developing strategies for managing GWAS on large numbers of densely genotyped individuals is an active area of research, as we look to generate new, more efficient algorithms that will enable the processing of these datasets within acceptable timeframes and computational limits of RAM and CPU. The importance of efficient algorithms is further highlighted when we conduct GWAS across large numbers of FT-MIR predicted traits and the individual FT-MIR wavenumbers. Existing mixed-linear model-based methods for conducting GWAS, such as GCTA-MLMA [94] primarily run in $O\ (mn^2)$ or $O\ (m^2n)$ time per trait, where m is the number of variants and n is the number of animals. These models become prohibitively slow as the numbers of genotyped individuals and variants increase [95]. The ever-increasing cohort sizes of densely-genotyped individuals frequently requires subsampling

to use these methods within acceptable computation constraints. This has spurred the development of faster, more memory-efficient algorithms and software. One software package, Bolt-LMM [95, 96] runs in approximately O ($mn^{1.5}$) time; however, it makes assumptions that are valid only for larger sample sizes. Recent versions are capable of running the entire UK biobank data set ($n = 459$ k) in a few days on a single computational node [96]. Another algorithm, fastGWA [97], available as a recent enhancement of the GCTA software package, provides further reductions in algorithmic complexity, running in approximately O (mn) time. These improvements mean that it is capable of running $n = 400$ k UK biobank samples in around 20 min, compared to 22 h for BOLT-LMM on the same hardware. Developments such as this make GWAS across sizeable populations with large numbers of FT-MIR phenotypes feasible.

Consolidating FT-MIR spectra with other omics data sources for QTL mapping

After conducting a GWAS, it is useful to identify the candidate genes and mutations underlying genomic loci with signal for a trait of interest. This can aid marker-assisted selection and improve our understanding of the biological pathways regulating the trait. Moreover, it has been shown that genomic prediction can be improved by including variants close to the causative mutations [98]. Software such as Ensembl's Variant Effect Predictor [99] is commonly used to identify candidate causal variants that have protein-coding or loss-of-function effects, with the expectation that these variants are more likely to impact the trait than other variants. However, recent studies in both humans [100, 101] and dairy cattle [86, 102] have highlighted the prevalence of QTL underpinned by expression-based mechanisms, and demonstrate that the majority of variance for at least some traits can be explained by non-coding variants located in regulatory elements. These variants are typically identified by considering the expression levels of genes as phenotypes, and using these data for genetic mapping studies in an approach known as expression QTL (eQTL) analysis.

Expression-based phenotypes

Assuming a causality chain hypothesis, as illustrated in Fig. 1, observation of an eQTL, co-located with a QTL for an FT-MIR predicted trait can inform on the mechanism of the trait of interest. This methodology can also be used to identify mechanisms underlying QTL observed for individual FT-MIR wavenumbers. A strong correlation between the variant effects for the two QTL (expression and FT-MIR related) suggest a shared underlying genetic architecture regulating both, while a weak correlation suggests that the two QTL, though co-located, do not co-segregate, and therefore represent distinct genetic signals with different causal variants.

Similar to eQTL analysis, a range of additional omics data sources can be used for QTL mapping, and the resulting QTL could be applied to identify causative genes for FT-MIR predicted traits and individual FT-MIR wavenumbers. The factors yielding these omics data sources can occur before or after mRNA transcription. Factors acting before transcription, such as DNA methylation and chromatin accessibility, can help unravel causative regulatory variants by highlighting actively-transcribed regions of the genome, and the variants that sit within them. One of these factors is chromatin accessibility. Transcriptionally active genes, as well as active regulatory elements (such as enhancers), are found in regions of open chromatin (euchromatin); whereas inactive regions of the genome are typically much more densely compacted into a structure known as heterochromatin. Genome features found in euchromatin are therefore more accessible to transcription factors and other factors involved in gene expression, and so are more likely to influence traits of interest compared to factors located in inactive regions. Methods to assay chromatin accessibility include ChIP-seq [103], DNase-seq [104], and ATAC-seq [105].

Other factors acting during or after transcription provide intermediate phenotypes that can aid in understanding the underlying biological control of these traits [106]. One such factor is RNA-editing, i.e. direct enzymatic conversion of bases within the mRNA transcripts, with the most common form of editing in vertebrates being the conversion of adenosine nucleotides into inosine [107]. Biologically, RNA editing is involved in protection against dsRNA viruses [108] and in adaptation to different environmental conditions [109], and therefore has potential relevance to variation in animal health and in providing for animal adaptability to changing environments. RNA-editing QTL (edQTL) were initially identified in *Drosophila* [110], followed soon after by mice [111] and humans [112]. Recently, edQTL were reported for the first time within the bovine mammary gland [113], and subsequently used to characterise candidate causative genes underlying a milk yield QTL at the *CSF2RB/NCF4* locus [114]. That study highlighted the manner in which intermediate molecular phenotypes can be used to investigate mechanisms underlying FT-MIR predicted trait QTL, and exemplifies how other similarly novel molecular phenotypes can be applied.

Metabolomics

Absorbance levels at individual FT-MIR wavenumbers provide insights into the presence of particular chemical bonds in the sample and accordingly provide information

as to the chemical composition of a milk sample. Analysing the chemical composition of a sample in more detail, using methodologies such as nuclear magnetic resonance (NMR) spectroscopy or mass spectroscopy (MS), yields the metabolome, i.e., a more complete set of all small molecules present in a tissue sample. Metabolomics can provide detailed information about enzymatic activity in the pathways that exist between gene expression and FT-MIR predicted traits, providing a near-terminal link in the chain of causality. For example, rumen volatile fatty acid (VFA) levels can provide information on measuring and controlling methane production [115]. Levels of VFAs in the rumen could therefore provide a proxy measurement for methane production. Identifying QTL that underlie variation in the concentrations of these metabolites could complement genetic signals identified using FT-MIR wavenumbers and FT-MIR based methane trait predictions, and facilitate selection of low-methane emitting animals.

Conclusions

Over the last 100 years, milk composition phenotyping for dairy cattle has evolved from manual on-farm methods for determining yield and fat levels in milk, to high-tech analysis at centralised laboratories, with many novel FT-MIR predicted traits now being considered for incorporation into improvement programs. Multiple studies have demonstrated that the accuracy of FT-MIR predictions are strongly influenced by how well the variation in the prediction population is represented in the calibration population. Trait prediction accuracy is also strongly affected by how well instrument-specific measurement differences are accounted for, particularly when transferring calibration equations developed on one instrument to spectra collected on other instruments. Utilising FT-MIR data to generate proxies for novel traits has grown in popularity, however, compared to FT-MIR predictions of major milk components, there are relatively few studies of the genetics of other FT-MIR predicted traits, and even fewer of the genetics of the individual wavenumbers. This is despite the individual wavenumbers exhibiting additional genetic signal that is often not observed in FT-MIR predictions of major milk composition traits. Integrating results from GWAS applied to FT-MIR predicted traits and GWAS applied to individual wavenumbers with other molecular datasets could improve our understanding of the underlying biological systems controlling traits of interest. However, integration of these data sources also brings computational challenges due to the size and complexity of the datasets involved. Resolving the challenges of effectively integrating FT-MIR datasets with other omics data sources will require a mix of both bioinformatics and molecular biology approaches. Successfully consolidating these approaches promises to improve our knowledge of milk composition and enable the future enhancement of animal breeding programmes.

Abbreviations

a_{30}: Curd firmness after 30 min; a_{60}: Curd firmness after 60 min; BHB: Beta-hydroxybutyrate; CH_4: Methane; CMS: Casein micelle size; CY: Curd yield; EBV: Estimated breeding value; edQTL: RNA-editing QTL; eQTL: Expression quantitative trait loci; FT-MIR: Fourier-transform mid-infrared; GC: Gas chromatography; GWAS: Genome wide association study; HCT: Heat coagulation time; k_{20}: Curd-firming time; LCFA: Long-chain fatty acids; MCFA: Medium-chain fatty acids; MD: Mahalanobis distance; MS: Mass spectroscopy; MUN: Milk urea nitrogen; NMR: Nuclear magnetic resonance; PLS: Partial least squares; PUFA: Polyunsaturated fatty acids; RCT: Rennet coagulation time; REC: Nutrient recovery; SCFA: Short-chain fatty acids; SFA: Saturated fatty acids; UFA: Unsaturated fatty acids; VFA: Volatile fatty acid

Acknowledgements

We acknowledge staff in the Research & Development group of Livestock Improvement Corporation (LIC; Hamilton, NZ) and fellow students based at the Massey University, AL Rae campus (Hamilton, NZ) for their support throughout the time of writing this manuscript. The authors would also like to thank the reviewers for their constructive feedback, which helped to significantly improve the manuscript.

Authors' contributions

K.T. and T.L. wrote the main manuscript text. M.L. and D.G. provided design input and critical evaluation, and contributed to the main manuscript text. All authors have read and approved the final manuscript.

Funding

This research was funded by Livestock Improvement Corporation (LIC) and the New Zealand Ministry for Primary Industries, through the Sustainable Food & Fibre Futures programme.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

K.T, T. L and M. L are employees of Livestock Improvement Corporation (Hamilton, NZ), a commercial milk testing and animal breeding company. K.T. is a PhD candidate in the School of Agriculture and Environment at Massey University (AL Rae Centre, Hamilton, NZ). M.L. is an Adjunct Professor at Massey University. D.G. is the Chief Scientist and Director at the AL Rae Centre (Hamilton, NZ) and a Professor of Animal Breeding and Genetics (Massey University). The authors declare that they have no competing interests in the publication of this manuscript.

Received: 3 December 2019 Accepted: 9 March 2020
Published online: 17 April 2020

References

1. Bayly C. 100 years of herd testing. Newstead, Hamilton, NZ: Livestock Improvement Corporation Ltd; 2009.
2. Miglior F, Fleming A, Malchiodi F, Brito LF, Martin P, Baes CF. A 100-year review: identification and genetic selection of economically important traits in dairy cattle. J Dairy Sci. 2017;100(12):10251–71.
3. De Marchi M, Toffanin V, Cassandro M, Penasa M. Invited review: mid-infrared spectroscopy as phenotyping tool for milk traits. J Dairy Sci. 2014; 97(3):1171–86.
4. Gengler N, Soyeurt H, Dehareng F, Bastin C, Colinet F, Hammami H, et al. Capitalizing on fine milk composition for breeding and management of dairy cows. J Dairy Sci. 2016;99(5):4071–9.
5. Te Pas MFW, Madsen O, Calus MPL, Smits MA. The importance of Endophenotypes to evaluate the relationship between genotype and external phenotype. Int J Mol Sci. 2017;18(2):472.

6. De Marchi M, Penasa M, Zidi A, Manuelian CL. Invited review: use of infrared technologies for the assessment of dairy products—applications and perspectives. J Dairy Sci. 2018;101(12):10589–604.
7. Egger-Danner C, Cole JB, Pryce JE, Gengler N, Heringstad B, Bradley A, et al. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal. 2015;9(2):191–207.
8. Bonfatti V, Vicario D, Lugo A, Carnier P. Genetic parameters of measures and population-wide infrared predictions of 92 traits describing the fine composition and technological properties of milk in Italian Simmental cattle. J Dairy Sci. 2017;100(7):5526–40.
9. Lopez-Villalobos N, Spelman RJ, Melis J, Davis SR, Berry SD, Lehnert K, et al. Estimation of genetic and crossbreeding parameters of fatty acid concentrations in milk fat predicted by mid-infrared spectroscopy in New Zealand dairy cattle. J Dairy Res. 2014;81(3):340–9.
10. Toffanin V, De Marchi M, Lopez-Villalobos N, Cassandro M. Effectiveness of mid-infrared spectroscopy for prediction of the contents of calcium and phosphorus, and titratable acidity of milk and their relationship with milk quality and coagulation properties. Int Dairy J. 2015;41:68–73.
11. Visentin G, McDermott A, McParland S, Berry DP, Kenny OA, Brodkorb A, et al. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J Dairy Sci. 2015;98(9):6620–9.
12. Visentin G, Penasa M, Niero G, Cassandro M, Marchi MD. Phenotypic characterisation of major mineral composition predicted by mid-infrared spectroscopy in cow milk. Ital J Anim Sci. 2018;17(3):549–56.
13. Toledo-Alvarado H, Vazquez AI, de los CG, Tempelman RJ, Bittante G. Cecchinato a. diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J Dairy Sci. 2018;101(3):2496–505.
14. Lainé A, Bastin C, Grelet C, Hammami H, Colinet FG, Dale LM, et al. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. J Dairy Sci. 2017;100(4):2863–76.
15. McParland S, Kennedy E, Lewis E, Moore SG, McCarthy B, O'Donovan M, et al. Genetic parameters of dairy cow energy intake and body energy status predicted using mid-infrared spectrometry of milk. J Dairy Sci. 2015; 98(2):1310–20.
16. Luke TDW, Rochfort S, Wales WJ, Bonfatti V, Marett L, Pryce JE. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J Dairy Sci. 2019;102(2):1747–60.
17. Oliveira MCPP, Silva NMA, Bastos LPF, Fonseca LM, Cerqueira MMOP, Leite MO, et al. Fourier transform infrared spectroscopy (FTIR) for MUN analysis in normal and adulterated Milk. Arq Bras Med Veterinária E Zootec. 2012;64(5):1360–6.
18. Bittante G, Cipolat-Gotet C. Direct and indirect predictions of enteric methane daily production, yield, and intensity per unit of milk and cheese, from fatty acids and milk Fourier-transform infrared spectra. J Dairy Sci. 2018;101(8):7219–35.
19. van Gastelen S, Mollenhorst H, Antunes-Fernandes EC, Hettinga KA, van Burgsteden GG, Dijkstra J, et al. Predicting enteric methane emission of dairy cows with milk Fourier-transform infrared spectra and gas chromatography–based milk fatty acid profiles. J Dairy Sci. 2018;101(6):5582–98.
20. Vanlierde A, Soyeurt H, Gengler N, Colinet FG, Froidmont E, Kreuzer M, et al. Short communication: development of an equation for estimating methane emissions of dairy cows from milk Fourier transform mid-infrared spectra by using reference data obtained exclusively from respiration chambers. J Dairy Sci. 2018;101(8):7618–24.
21. Ferragina A, de los Campos G, Vazquez AI, Cecchinato A, Bittante G. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. J Dairy Sci. 2015;98(11):8133–51.
22. El Jabri M, Sanchez M-P, Trossat P, Laithier C, Wolf V, Grosperrin P, et al. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. J Dairy Sci. 2019;102(8):6943–58.
23. Bonfatti V, Tiezzi F, Miglior F, Carnier P. Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations. J Dairy Sci. 2017;100(9):7306–19.
24. Rutten MJM, Bovenhuis H, Hettinga KA, van Valenberg HJF, van Arendonk JAM. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. J Dairy Sci. 2009; 92(12):6202–9.
25. McParland S, Banos G, Wall E, Coffey MP, Soyeurt H, Veerkamp RF, et al. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J Dairy Sci. 2011;94(7):3651–61.
26. McParland S, Banos G, McCarthy B, Lewis E, Coffey MP, O'Neill B, et al. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. J Dairy Sci. 2012;95(12):7225–35.
27. Soyeurt H, Dehareng F, Gengler N, McParland S, Wall E, Berry DP, et al. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J Dairy Sci. 2011;94(4):1657–67.
28. Bonfatti V, Degano L, Menegoz A, Carnier P. Short communication: mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. J Dairy Sci. 2016;99(10):8216–21.
29. Soyeurt H, Dardenne P, Dehareng F, Lognay G, Veselko D, Marlier M, et al. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J Dairy Sci. 2006;89(9):3690–5.
30. Bonfatti V, Di Martino G, Carnier P. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. J Dairy Sci. 2011;94(12):5776–85.
31. De Marchi M, Bonfatti V, Cecchinato A, Di Martino G, Carnier P. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Ital J Anim Sci. 2009;8(sup2):399–401.
32. McDermott A, Visentin G, De Marchi M, Berry DP, Fenelon MA, O'Connor PM, et al. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J Dairy Sci. 2016;99(4):3171–82.
33. Rutten MJM, Bovenhuis H, Heck JML, van Arendonk JAM. Predicting bovine milk protein composition based on Fourier transform infrared spectra. J Dairy Sci. 2011;94(11):5683–90.
34. Dal Zotto R, De Marchi M, Cecchinato A, Penasa M, Cassandro M, Carnier P, et al. Reproducibility and repeatability of measures of Milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy. J Dairy Sci. 2008;91(10):4103–12.
35. De Marchi M, Fagan CC, O'Donnell CP, Cecchinato A, Zotto RD, Cassandro M, et al. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. J Dairy Sci. 2009;92(1):423–32.
36. De Marchi M, Toffanin V, Cassandro M, Penasa M. Prediction of coagulating and noncoagulating milk samples using mid-infrared spectroscopy. J Dairy Sci. 2013;96(7):4707–15.
37. Wang Y, Veltkamp DJ, Kowalski BR. Multivariate instrument standardization. Anal Chem. 1991;63(23):2750–6.
38. Geladi P, MacDougall D, Martens H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl Spectrosc. 1985;39(3):491–500.
39. Martens H, Nielsen JP, Engelsen SB. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. Anal Chem. 2003;75(3):394–404.
40. Abraham S, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36(8):1627–39.
41. De Marchi M, Penasa M, Cecchinato A, Mele M, Secchiari P, Bittante G. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk. Animal. 2011;5(10):1653–8.
42. Grelet C, Pierna JAF, Dardenne P, Baeten V, Dehareng F. Standardization of milk mid-infrared spectra from a European dairy network. J Dairy Sci. 2015; 98(4):2150–60.
43. Tiplady KM, Sherlock RG, Littlejohn MD, Pryce JE, Davis SR, Garrick DJ, et al. Strategies for noise reduction and standardization of milk mid-infrared spectra from dairy cattle. J Dairy Sci. 2019;102(7):6357–72.
44. Bittante G, Cecchinato A. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J Dairy Sci. 2013;96(9):5991–6006.
45. Wang Q, Hulzebosch A, Bovenhuis H. Genetic and environmental variation in bovine milk infrared spectra. J Dairy Sci. 2016;99(8):6793–803.
46. Wang Q, Bovenhuis H. Genome-wide association study for milk infrared wavenumbers. J Dairy Sci. 2018;101(3):2260–72.
47. Lynch JM, Barbano DM, Schweisthal M, Fleming JR. Precalibration evaluation procedures for mid-infrared Milk analyzers. J Dairy Sci. 2006;89(7):2761–74.
48. Bonfatti V, Fleming A, Koeck A, Miglior F. Standardization of milk infrared spectra for the retroactive application of calibration models. J Dairy Sci. 2017;100(3):2032–41.
49. Grelet C, Pierna JAF, Dardenne P, Soyeurt H, Vanlierde A, Colinet F, et al. Standardization of milk mid-infrared spectrometers for the transfer and use of multiple models. J Dairy Sci. 2017;100(10):7910–21.
50. Winning H, Mulawa KM, Selberg T. Standardization of FT-IR instruments. White Paper from Foss A/S. 2014;1(1):7.

51. Gupta D, Wang L, Hanssen LM, Hsia JJ, Datla RU. Standard reference materials: Polystyrene films for calibrating the wavelength scale of infraredspectrophotometers - SRM 1921. Boulder (CO): U.S. Department of Commerce; 1995. Report No.: NIST spec publ. 260-122.
52. Parsons C, Lyder H inventors; Bentley Instruments Inc, assignee. Determining a size of cell of a transmission spectroscopy device. United States patent US 9,829,378. 2017.
53. Soyeurt H, Gillon A, Vanderick S, Mayeres P, Bertozzi C, Gengler N. Estimation of heritability and genetic correlations for the major fatty acids in bovine milk. J Dairy Sci. 2007;90(9):4435–42.
54. Rutten MJM, Bovenhuis H, van Arendonk JAM. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data. J Dairy Sci. 2010;93(10):4872–82.
55. Hein L, Sørensen LP, Kargo M, Buitenhuis AJ. Genetic analysis of predicted fatty acid profiles of milk from Danish Holstein and Danish Jersey cattle populations. J Dairy Sci. 2018;101(3):2148–57.
56. Fleming A, Schenkel FS, Malchiodi F, Ali RA, Mallard B, Sargolzaei M, et al. Genetic correlations of mid-infrared-predicted milk fatty acid groups with milk production traits. J Dairy Sci. 2018;101(5):4295–306.
57. Narayana SG, Schenkel FS, Fleming A, Koeck A, Malchiodi F, Jamrozik J, et al. Genetic analysis of groups of mid-infrared predicted fatty acids in milk. J Dairy Sci. 2017;100(6):4731–44.
58. Sanchez MP, Ferrand M, Gelé M, Pourchet D, Miranda G, Martin P, et al. Short communication: genetic parameters for milk protein composition predicted using mid-infrared spectroscopy in the French Montbéliarde, Normande, and Holstein dairy cattle breeds. J Dairy Sci. 2017;100(8):6371–5.
59. Soyeurt H, Colinet FG, Arnould VM-R, Dardenne P, Bertozzi C, Renaville R, et al. Genetic variability of lactoferrin content estimated by mid-infrared spectrometry in bovine milk. J Dairy Sci. 2007;90(9):4443–50.
60. Arnould VM-R, Soyeurt H, Gengler N, Colinet FG, Georges MV, Bertozzi C, et al. Genetic analysis of lactoferrin content in bovine milk. J Dairy Sci. 2009;92(5):2151–8.
61. Lopez-Villalobos N, Davis SR, Beattie EM, Melis J, Berry S, Holroyd SE, et al. Breed effects for lactoferrin concentration determined by Fourier transform infrared spectroscopy. Proc N Z Soc Anim Prod. 2009;69:60–4.
62. Visentin G, McParland S, De Marchi M, McDermott A, Fenelon MA, Penasa M, et al. Processing characteristics of dairy cow milk are moderately heritable. J Dairy Sci. 2017;100(8):6343–55.
63. Cecchinato A, Marchi MD, Gallo L, Bittante G, Carnier P. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. J Dairy Sci. 2009;92(10):5304–13.
64. Costa A, Visentin G, Marchi MD, Cassandro M, Penasa M. Genetic relationships of lactose and freezing point with minerals and coagulation traits predicted from milk mid-infrared spectra in Holstein cows. J Dairy Sci. 2019;102(8):7217–25.
65. Sanchez MP, El Jabri M, Minéry S, Wolf V, Beuvier E, Laithier C, et al. Genetic parameters for cheese-making properties and milk composition predicted from mid-infrared spectra in a large data set of Montbéliarde cows. J Dairy Sci. 2018;101(11):10048–61.
66. Bittante G, Ferragina A, Cipolat-Gotet C, Cecchinato A. Comparison between genetic parameters of cheese yield and nutrient recovery or whey loss traits measured from individual model cheese-making methods or predicted from unprocessed bovine milk samples using Fourier-transform infrared spectroscopy. J Dairy Sci. 2014;97(10):6560–72.
67. Cecchinato A, Albera A, Cipolat-Gotet C, Ferragina A, Bittante G. Genetic parameters of cheese yield and curd nutrient recovery or whey loss traits predicted using Fourier-transform infrared spectroscopy of samples collected during milk recording on Holstein, Brown Swiss, and Simmental dairy cows. J Dairy Sci. 2015;98(7):4914–27.
68. Bastin C, Théron L, Lainé A, Gengler N. On the role of mid-infrared predicted phenotypes in fertility and health dairy breeding programs. J Dairy Sci. 2016;99(5):4080–94.
69. Belay TK, Svendsen M, Kowalski ZM, Ådnøy T. Genetic parameters of blood β-hydroxybutyrate predicted from milk infrared spectra and clinical ketosis, and their associations with milk production traits in Norwegian red cows. J Dairy Sci. 2017;100(8):6298–311.
70. Kandel PB, Vanrobays M-L, Vanlierde A, Dehareng F, Froidmont E, Gengler N, et al. Genetic parameters of mid-infrared methane predictions and their relationships with milk production traits in Holstein cattle. J Dairy Sci. 2017;100(7):5578–91.
71. Negussie E, de Haas Y, Dehareng F, Dewhurst RJ, Dijkstra J, Gengler N, et al. Invited review: large-scale indirect measurements for enteric methane emissions in dairy cattle: a review of proxies and their potential for use in management and breeding decisions. J Dairy Sci. 2017;100(4):2433–53.
72. Hristov AN, Kebreab E, Niu M, Oh J, Bannink A, Bayat AR, et al. Symposium review: uncertainties in enteric methane inventories, measurement techniques, and prediction models. J Dairy Sci. 2018;101(7):6655–74.
73. Wood GM, Boettcher PJ, Jamrozik J, Jansen GB, Kelton DF. Estimation of genetic parameters for concentrations of milk urea nitrogen. J Dairy Sci. 2003;86(7):2462–9.
74. Miglior F, Sewalem A, Jamrozik J, Bohmanova J, Lefebvre DM, Moore RK. Genetic analysis of milk urea nitrogen and lactose and their relationships with other production traits in Canadian Holstein cattle. J Dairy Sci. 2007;90(5):2468–79.
75. Mitchell RG, Rogers GW, Dechow CD, Vallimont JE, Cooper JB, Sander-Nielsen U, et al. Milk urea nitrogen concentration: heritability and genetic correlations with reproductive performance and disease. J Dairy Sci. 2005;88(12):4434–40.
76. Stoop WM, Bovenhuis H, van Arendonk JAM. Genetic parameters for Milk urea nitrogen in relation to Milk production traits. J Dairy Sci. 2007;90(4):1981–6.
77. Soyeurt H, Misztal I, Gengler N. Genetic variability of milk components based on mid-infrared spectral data. J Dairy Sci. 2010;93(4):1722–8.
78. Rovere G, de los Campos G, Tempelman RJ, Vazquez AI, Miglior F, Schenkel F, et al. A landscape of the heritability of Fourier-transform infrared spectral wavelengths of milk samples by parity and lactation stage in Holstein cows. J Dairy Sci. 2019;102(2):1354–63.
79. Zaalberg RM, Shetty N, Janss L, Buitenhuis AJ. Genetic analysis of Fourier transform infrared milk spectra in Danish Holstein and Danish Jersey. J Dairy Sci. 2019;102(1):503–10.
80. Dagnachew BS, Meuwissen THE, Ådnøy T. Genetic components of milk Fourier-transform infrared spectra used to predict breeding values for milk composition and quality traits in dairy goats. J Dairy Sci. 2013;96(9):5933–42.
81. Bonfatti V, Vicario D, Degano L, Lugo A, Carnier P. Comparison between direct and indirect methods for exploiting Fourier transform spectral information in estimation of breeding values for fine composition and technological properties of milk. J Dairy Sci. 2017;100(3):2057–67.
82. Belay TK, Dagnachew BS, Boison SA, Ådnøy T. Prediction accuracy of direct and indirect approaches, and their relationships with prediction ability of calibration models. J Dairy Sci. 2018;101(7):6174–89.
83. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, et al. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 2010;5(10):e13661.
84. Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, et al. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genet Sel Evol. 2015;47(1):29.
85. Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, et al. Sequence-based association analysis reveals an *MGST1* eQTL with pleiotropic effects on bovine Milk composition. Sci Rep. 2016;6:25376.
86. Lopdell TJ, Tiplady K, Struchalin M, Johnson TJJ, Keehan M, Sherlock R, et al. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. BMC Genomics BioMed Central. 2017;18(1):968.
87. Raven L-A, Cocks BG, Hayes BJ. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics. 2014;15(1):62.
88. Bouwman AC, Bovenhuis H, Visker MHPW, van Arendonk JAM. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet. 2011;12(1):43.
89. Buitenhuis B, Janss LL, Poulsen NA, Larsen LB, Larsen MK, Sørensen P. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics. 2014;15(1):1112.
90. Buitenhuis B, Poulsen NA, Gebreyesus G, Larsen LB. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 2016;17(1):114.

91. Li C, Sun D, Zhang S, Wang S, Wu X, Zhang Q, et al. Genome Wide Association Study Identifies 20 Novel Promising Genes Associated with Milk Fatty Acid Traits in Chinese Holstein. PLoS One. 2014;9(5):e96186.
92. Sanchez MP, Govignon-Gion A, Ferrand M, Gelé M, Pourchet D, Amigues Y, et al. Whole-genome scan to detect quantitative trait loci associated with milk protein composition in 3 French dairy cattle breeds. J Dairy Sci. 2016; 99(10):8203–15.
93. Benedet A, Ho PN, Xiang R, Bolormaa S, Marchi MD, Goddard ME, et al. The use of mid-infrared spectra to map genes affecting milk composition. J Dairy Sci. 2019;102(8):7189–203.
94. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82.
95. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47(3):284–90.
96. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nat Genet. 2018;50(7):906–8.
97. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. Nat Genet. 2019;51(12):1749–55.
98. van den Berg I, Boichard D, Guldbrandtsen B, Lund MS. Using Sequence Variants in Linkage Disequilibrium with Causative Mutations to Improve Across-Breed Prediction in Dairy Cattle: A Simulation Study. G3 Genes Genomes Genet. 2016, 6(8):2553–61.
99. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17(1):122.
100. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5.
101. Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, et al. A method to predict the impact of regulatory variants from DNA sequence. Nat Genet Nature Publishing Group. 2015;47:955–61.
102. Pausch H, Emmerling R, Schwarzenbacher H, Fries R. A multi-trait meta-analysis with imputed sequence variants reveals twelve QTL for mammary gland morphology in Fleckvieh cattle. Genet Sel Evol. 2016;48(1):14.
103. O'Neill LP, Turner BM. Immunoprecipitation of native chromatin: NChIP. Methods. 2003;31(1):76–82.
104. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311–22.
105. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015; 109(1):21–9.
106. Kemper KE, Littlejohn MD, Lopdell T, Hayes BJ, Bennett LE, Williams RP, et al. Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. BMC Genomics. 2016;17(1):858.
107. Savva YA, Rieder LE, Reenan RA. The ADAR protein family. Genome Biol. 2012;13(12):1.
108. Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science. 2015;349(6252):1115–20.
109. Garrett S, Rosenthal JJ. RNA editing underlies temperature adaptation in K+ channels from polar octopuses. Science. 2012;335(6070):848–51.
110. Ramaswami G, Deng P, Zhang R, Carbone MA, Mackay TF, Li JB. Genetic mapping uncovers cis-regulatory landscape of RNA editing. Nat Commun. 2015;6(1):1–9.
111. Gu T, Gatti DM, Srivastava A, Snyder EM, Raghupathy N, Simecek P, et al. Genetic architectures of quantitative variation in RNA editing pathways. Genetics. 2016;202(2):787–98.
112. Park E, Guo J, Lin L, Demirdjian L, Shen S, Xing Y, et al. Population and allelic variation of A-to-I RNA editing in human transcriptomes. Genome Biol. 2017;18(1):143.
113. Lopdell TJ, Hawkins V, Couldrey C, Tiplady K, Davis SR, Harris BL, et al. Widespread cis-regulation of RNA editing in a large mammal. RNA. 2019; 25(3):319–35.
114. Lopdell TJ, Tiplady K, Couldrey C, Johnson TJJ, Keehan M, Davis SR, et al. Multiple QTL underlie milk phenotypes at the CSF2RB locus. Genet Sel Evol. 2019;51(1):3.
115. Knapp J, Laur G, Vadas P, Weiss W, Tricarico J. Invited review: enteric methane in dairy cattle production: quantifying the opportunities and impact of reducing emissions. J Dairy Sci. 2014;97(6):3231–61.