High-Value Metabolic Phenotypes for Improved Dairy Cow Health

Submitted by

Timothy David William Luke

BVSc(Hons)

A thesis submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Applied Systems Biology

College of Science, Health and Engineering

La Trobe University

Victoria, Australia

July 2020

Table of Contents

List of Abbreviations

^{1}H NMR	Proton nuclear magnetic resonance
A:G	Albumin to globulin ratio
ABV	Australian breeding value
BHBA	β-hydroxybutyrate
Ca	Calcium
CM	Clinical mastitis
DIM	Days in milk
DMI	Dry matter intake
DNA	Deoxyribonucleic acid
EBV	Estimated breeding value
FTIR	Fourier-transform infrared
GC-MS	Gas chromatography mass spectroscopy
GEBV	Genomic estimated breeding value
GS	Genomic selection
GWAS	Genome-wide association
h^2	Heritability
Hp	Haptoglobin
LC-MS	Liquid chromatography mass spectroscopy
LDL	Low-density lipoprotein
Mg	Magnesium
mGWAS	Metabolite-based genome-wide association
MIR	Mid-infrared
N	Sample size
NEFA	Non-esterified fatty acids
PLS	Partial least squares
PLS-DA	Partial least squares discriminant analysis
PTH	Parathyroid hormone

QTL	Quantitative trait locus
R^2	Coefficient of determination
RFI	Residual feed intake
RMSE	Root mean square error
RNA	Ribonucleic acid
SCC	Somatic cell count
SNP	Single nucleotide polymorphism
USD	United States dollar
VLDL	Very-low-density lipoprotein

Abstract

Metabolic disorders in the periparturient period have significant negative impacts on the health, welfare and profitability of dairy cows. Most genetic evaluations of metabolic health traits have used health data phenotypes, which although widely available, are limited by under/inconsistent reporting, misdiagnoses and binary clinical disease definitions. The aim of the work described in this thesis was therefore to investigate metabolic phenotypes for use in genetic selection for improved metabolic health in early-lactation dairy cows. Firstly, genetic parameters and genomic prediction accuracies were estimated for nine serum metabolic profile biomarker traits associated with energy balance (β-hydroxybutyrate (BHBA) and non-esterified fatty acids (NEFA)), macro-mineral status (calcium and magnesium), protein nutritional status (urea and albumin) and immune status (globulins, albumin:globulin ratio and haptoglobin), using data collected from a genotyped female reference population (N = 1393). All traits except haptoglobin were heritable ($0.07 \leq h^2$ 0.46), and favorable genetic correlations between traits suggested that selection for overall metabolic resilience may be possible. Genomic prediction accuracies were consistent with heritability estimates and the small reference population size. Secondly, the use of mind-infrared (MIR) spectral data derived from routine milk recording to predict concentrations of the aforementioned serum biomarkers was investigated. Prediction accuracies were promising for BHBA, NEFA and urea (R^2 0.48, 0.61 and 0.91, respectively), but poor for the remaining biomarkers. Results suggested that MIR-predicted phenotypes could offer a high throughput and cost-effective way to increase the size of reference populations for genomic selection for metabolic resilience. Lastly, an untargeted proton nuclear magnetic resonance (^{1}H NMR) metabolomics approach was used to identify intermediate phenotypes associated with BHBA and NEFA. Sixteen metabolites were identified, including intermediates of energy, phospholipid, and/or methyl donor metabolism. Overall, the findings of this thesis provide further evidence that metabolic phenotypes are likely to be of great value in the development of more accurate breeding values for improved metabolic health.

Statement of Authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

The extent and nature of contributions by co-authors are specified in Appendix 1. Each statement of collaborative input has been approved by all co-authors and their approval verified by Prof Jennie Pryce in the Authority to Submit Form.

Timothy David William Luke

3rd of July 2020

Funding Acknowledgement

This project was funded by DairyBio, a joint venture between Dairy Australia (Melbourne, Australia), Agriculture Victoria (Melbourne, Australia) and the Gardiner Foundation (Melbourne, Australia). Timothy David William Luke was supported by a DairyBio Graduate Researcher Scholarship. This work was supported by an Australian Government Research Training Program Scholarship.

Animal Ethics Statement

All procedures undertaken in this study were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (National Health and Medical Research Council, 2013[1]). Approval to proceed was granted by the Agricultural Research and Extension Animal Ethics Committee of the Department of Jobs, Precincts and Regions Animal Ethics Committee (DJPR, 475 Mickleham Road, Attwood, Victoria 3049, Australia), and the Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIPWE Animal Biosecurity and Welfare Branch, 13 St Johns Avenue, New Town, Tasmania 7008, Australia).

[1] National Health and Medical Research Council. 2013. Australian code for the care and use of animals for scientific purposes. 8th ed. NHMRC Publications, Canberra, Australia.

Acknowledgements

To adequately thank all who have helped me during this PhD journey would require a document of comparable length to this thesis. To distill such a list down to just a few paragraphs has been a challenge - talk about leaving the hardest section until last!

Firstly, I am incredibly grateful to DairyBio, jointly funded by Agriculture Victoria, Dairy Australia and the Gardiner Foundation, for their generous financial support.

To my supervisors, Professor Jennie Pryce and Dr Simone Rochfort. Thank you both for your vision, guidance and humour, and for giving me the freedom and support to set and follow my own path. I reckon we've made a great team – it's been fun!

To Dr Bill Wales, thank you for including me as part of the Ellinbank team, and for generously sharing your experience and wisdom. And to Dr Thuy Nguyen, thank you for your invaluable guidance, support and friendship.

To the on-farm "A-team" of Di Mapleson, Brigid Ribaux, Erika Oakes, Jen Williams and Dr Joe Brady. My sincere thanks for your expert help, and for filling the often long and tiring on-farm days with friendship and humour. A special thanks also to the farmers who took part in this project.

I have had the great privilege of working alongside the Computational Biology, Molecular Phenomics and Dairy Production Science teams at Agriculture Victoria Research. To all my friends in these teams who have helped me throughout the course of this PhD, a heartfelt thank you!

And lastly, to my steadfast and loving support crew; my truly amazing parents, Lorna and Richard, and my four-legged mate, Charlie the dog, thank you.

Preface

This thesis consists of six chapters. Chapter 1 is a general introduction which provides an overview of the research topic and outlines the main objectives of the project. Chapters 2 to 5 present the findings of original research undertaken as part of this project. All four research chapters have been published in peer-reviewed scientific journals and appear in this thesis in the format of the journal of publication. Chapter 6 is a general discussion which places the main findings of this research in a broader context and offers suggestions for future research.

Chapter 1:

General Introduction

Chapter 1: General Introduction

The modern dairy cow is one of the success stories of 20th century animal breeding. Between 1919 and 2019, average per cow production in Australia increased from 1264 L/year (Commonwealth Bureau of Census and Statistics, 1925), to 6169 L/year (Dairy Australia, 2019); a staggering 488% increase. Much of this increase has occurred since the mid-1970s, and approximately 30% can be attributed to genetic improvement (Pryce et al., 2018) (Figure 1).

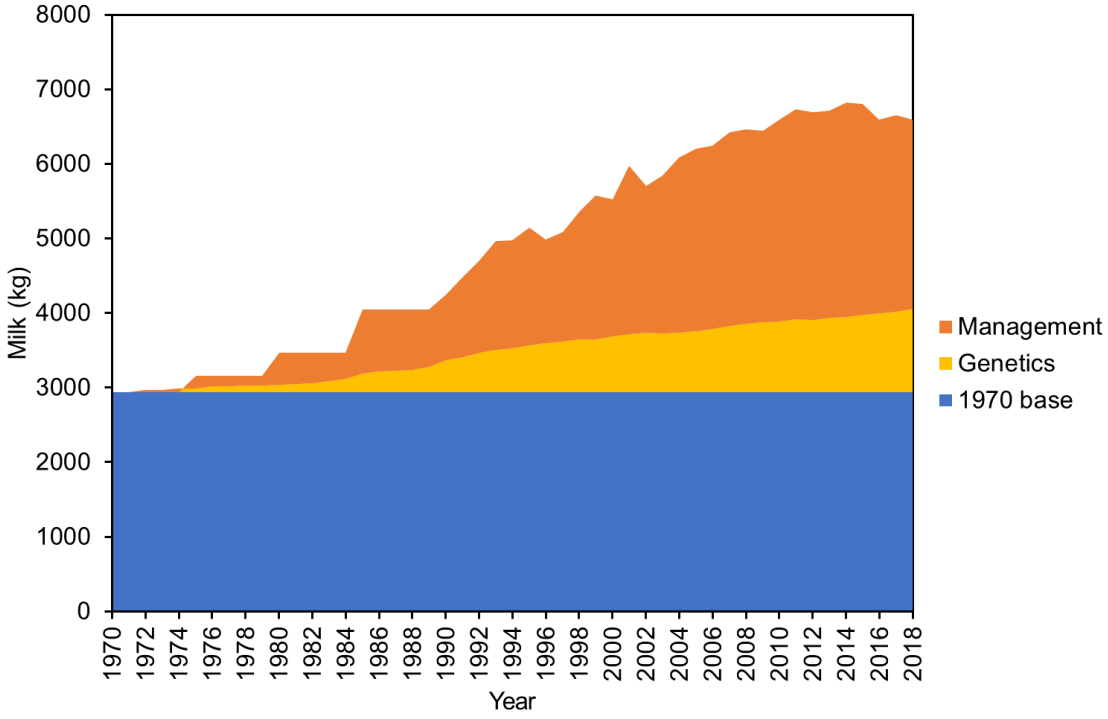


Figure1. Relative contributions of genetic improvement and management to increases in annual per cow milk production in Australia between 1970 and 2018. Adapted from Cole et al. (2020) using Australian data (Source: DataGene, Bundoora, Australia).

However, single trait selection for improved milk yield has also given rise to several significant challenges. Perhaps the most well-recognised example has been the marked reduction in dairy cow reproductive performance (Lucy, 2001), due in large part to antagonistic genetic correlations between fertility and production traits (Berry et al., 2014). In order to prevent continuation of this problem in the future, breeding objectives have been broadened to include functional traits such as udder health, longevity and fertility (Egger-Danner et al., 2015, Miglior et al., 2017). Moreover, improving animal health and resilience through genetic selection is seen as vital if the dairy industry is to maintain its social licence to operate (Boichard and Brochard, 2012).

The overwhelming majority of adverse health events affecting dairy cows occur during the periparturient or transition period (Drackley, 1999), which is defined as the three to four weeks before and after calving (Grummer, 1995, Drackley, 1999, Lean and Degaris, 2010). Many of these adverse events are associated with metabolic perturbations (Curtis et al., 1985, Ospina et al., 2010b, McArt et al., 2013, Rodríguez et al., 2017), and there is therefore increasing interest in improving metabolic health through genetic selection (Pryce et al., 2016). However, transition cow metabolism is incredibly complex (Drackley, 1999), and while modern breeding tools such as genomic selection are enabling genetic improvement in such traits (Boichard and Brochard, 2012, Egger-Danner et al., 2015), progress is limited by a dearth of phenotypes which accurately capture and describe this complexity (Coffey, 2020).

The aim of the work described in this thesis is therefore to explore novel, high-resolution metabolic phenotypes for use in genomic selection for improved dairy cow health. After all, *"in the age of the genotype, phenotype is king"* (Coffey, 2020).

1.1 The cost of success…

High levels of milk production come at an enormous metabolic cost. The metabolizable energy (ME) requirements of a 650 kg Holstein cow producing 40 litres/Day of milk (4.0% fat, 3.4% protein) are approximately 272 MJ ME/Day; more than four times her maintenance energy requirements of 65.5 MJ ME/Day (Chamberlain and Wilkinson, 1996). This is equivalent to the maximum energy expenditure of a cyclist competing in the hill stages of the Tour de France (Saris et al., 1989).

When compared to those of lactation, the metabolic demands of pregnancy are relatively small. Figure 2 shows the estimated requirements for metabolizable energy, metabolizable protein, calcium and magnesium, for maintenance, pregnancy and lactation; for a 650kg cow at drying off (32 weeks pregnant), at the point of calving (40 weeks pregnant), and at various levels of milk production (Chamberlain and Wilkinson, 1996). These figures clearly demonstrate the enormous metabolic demands of lactation. More specifically, Bell (1995) has shown that the requirements of the udder for glucose, amino acids, and fatty acids, at day four after calving, are approximately 2.7, 2.0, and 4.5 times those of the uterus/foetus at day 250 of pregnancy, respectively.

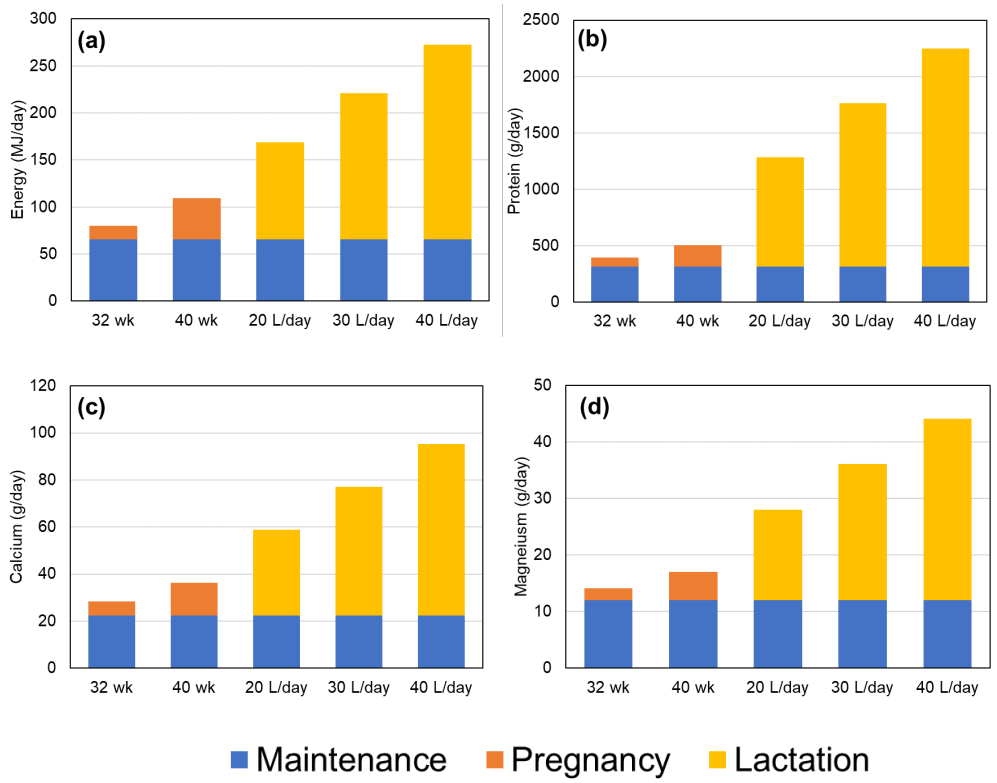


Figure 2. Estimated (a) metabolizable energy (MJ/day); (b) metabolizable protein (g/day); calcium (g/day); and magnesium (g/day) requirements for a 650 kg Holstein Friesian dairy cow (1) at drying off (32 weeks pregnant), (2) at the point of calving (40 weeks pregnant); (3) at different milk production levels. Figures based on 4.0 % milk fat, 3.4% milk protein, and ration quality 0.6. Source Chamberlain and Wilkinson (1996).

The meeting of metabolic demands for lactation is made more challenging by the rapid onset of milk production at, or immediately before calving (Tucker, 1981), and a concurrent decrease in feed intake (Grant and Albright, 1995). The result is a period of negative energy balance in the immediate post-calving period, during which body energy stores are mobilized to meet demand (Figure 3); which is associated with increased risk of health disorders (Cameron et al., 1998, Drackley, 1999, Ospina et al., 2010b). Cows also experience immune suppression around calving, which increases the risk of infectious diseases such as mastitis and metritis (Goff and Horst, 1997, Drackley, 1999).

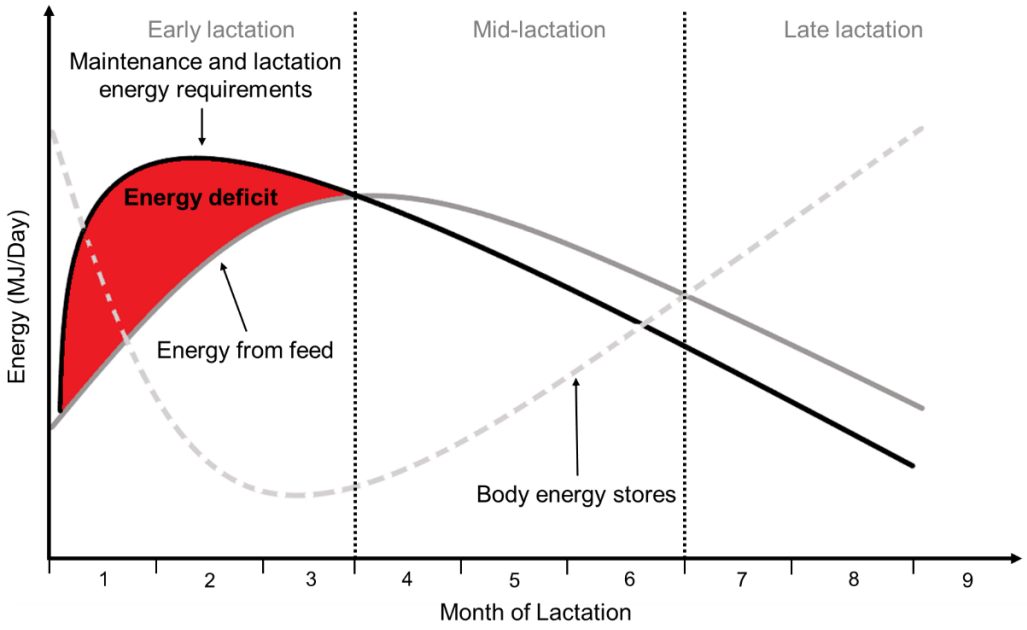


Figure 3. Generalized milk energy to feed energy relationship demonstrating the period of negative energy balance (energy deficit) immediately post-calving. Adapted from Hoffman et al. (2000).

1.2 Meeting the challenge…metabolic adaptations during the transition period

Overcoming the challenges of the transition period requires a series of complex and coordinated changes in metabolism and nutrient partitioning known as homeorhesis (Bauman and Currie, 1980). Unlike homeostasis, maintenance of physiological equilibrium, homeorhesis describes the orchestrated changes required to facilitate and support a change in physiological state (Bauman and Currie, 1980), change such as that from pregnancy to lactation.

Some of the most important metabolic pathways under homeorhetic control during the transition period are summarised in Figure 4. Changes in energy and protein metabolism include (1) increased lipolysis and decreased lipogenesis, (2) increased glucose production through gluconeogenesis and glycogenolysis, (3) increased use of lipids, and decreased use of glucose, for energy, and (4) increased mobilisation of protein reserves (Bauman and Currie, 1980). Changes to calcium metabolism are mediated by increased release of parathyroid hormone (PTH) and subsequent activation of vitamin D3 to 1,25-dihydroxyvitamin D3 in the kidney, which in turn leads to (1) increased intestinal absorption of calcium, (2) increased resorption of calcium from bone, and (3) decreased excretion of calcium by the kidney (Degaris and Lean, 2008).

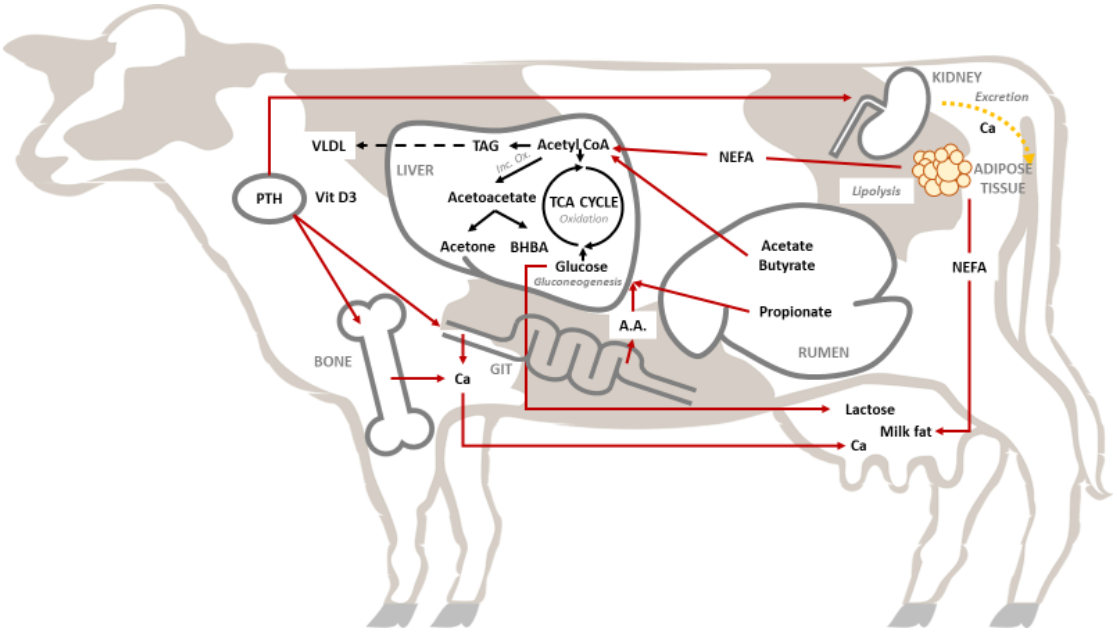


Figure 4. Summary of important metabolic processes under homeorhetic control during the transition period. Adapted from Parkinson et al. (2019). A.A. = amino acids; BHBA = β-hydroxybutyrate; Ca = calcium; GIT = gastrointestinal tract; Inc. Ox = incomplete oxidation; NEFA = non-esterified fatty acids; PTH = parathyroid hormone; TAG = triglycerides; Vit D3 = vitamin D3; VLDL = very-low-density lipoprotein.

Of these processes, mobilisation of stored energy from adipose tissue as non-esterified fatty acids (NEFA) is particularly important. Mobilised NEFA are transported via the blood stream to the mammary gland for milk fat synthesis, or to the liver where they undergo (1) complete oxidation via the TCA cycle to produce energy in the form of adenosine triphosphate (ATP), (2) partial oxidation to ketone bodies (β-hydroxybutyrate (BHBA), acetone and acetoacetate) which can be used as an energy source by peripheral tissues (heart, kidney, skeletal muscle, mammary gland and gastrointestinal tract) (Heitmann et al., 1987), or (3) re-esterification to form triglycerides, which can either be stored in the liver or exported as very-low-density lipoprotein (VLDL).

1.3 When things go wrong…metabolic disorders

Metabolic disorders occur when one or more homeorhetic controls fail. The most commonly described metabolic disorders in dairy cows are the result of perturbed energy and/or lipid metabolism (ketosis/acetonaemia/hyperketonaemia) and mineral metabolism (hypocalcaemia and hypomagnesaemia). The pathophysiology of these disorders is complex, and is the focus of several excellent reviews (Herdt, 2000, Degaris and Lean, 2008). It is well accepted that many metabolic processes are intricately linked, and that many metabolic disorders are inter-related and often occur simultaneously (Curtis et al., 1985, Lean and Degaris, 2010). There is also increasing evidence that the pathophysiologies of metabolic and infectious

diseases are closely linked (Degaris and Lean, 2008, Sordillo and Raphael, 2013, Sordillo, 2016).

Metabolic disorders can be either clinical (associated with observable signs of illness) or subclinical (effectively "invisible"). Returning to the analogy of the Tour de France (Figure 5), a cow that develops a clinical metabolic disorder is like a cyclist that falls off their bike (the red cow), while a cow that develops a subclinical metabolic disorder is like a cyclist who has an imperceptible wobble at the beginning of the race, and goes on to finish a shorter race over a lower mountain.

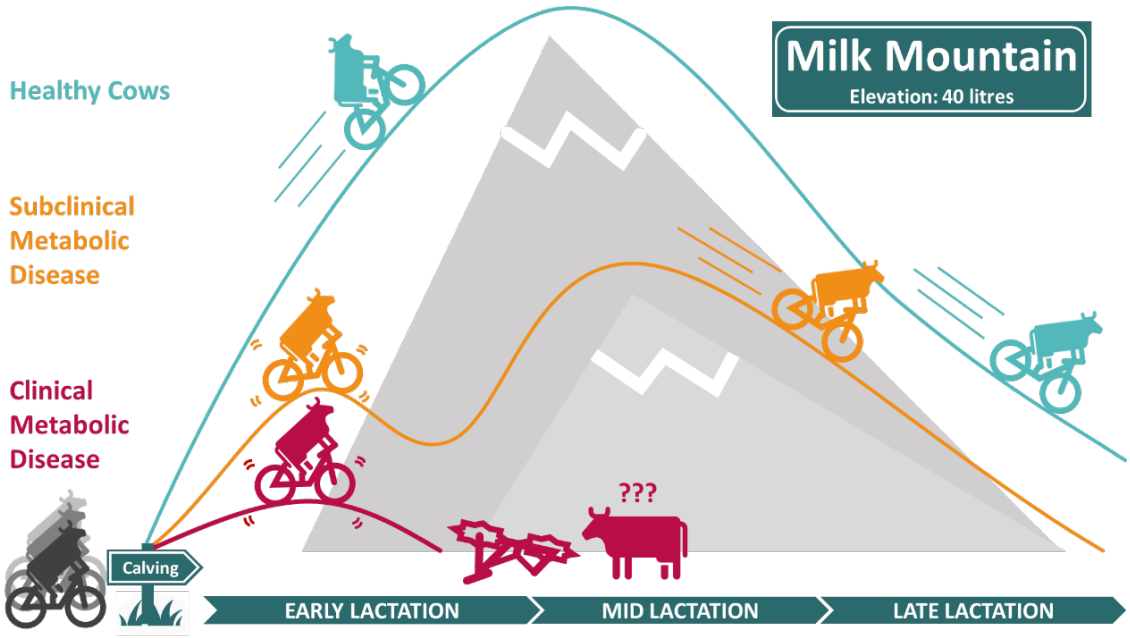


Figure 5. Differences in the health and performance of healthy cows (blue), cows that develop a subclinical metabolic disorder (orange), and cows that develop a clinical metabolic disorder (red) during the transition period. The analogy is based on the fact that during the early lactation period, dairy cows have energy requirements comparable to those of cyclists competing in the hill stages of the Tour de France.

1.3.1 Clinical metabolic disorders

Clinical metabolic disorders are relatively rare, and incidence rates vary significantly between farms (Pryce et al., 2016). The clinical signs and median incidences of ketosis, hypocalcaemia and hypomagnesaemia are summarised in Table 1. Many of the clinical signs associated with metabolic disorders (such as decreased appetite and milk production) are non-specific, making accurate diagnosis difficult, even for experienced veterinary practitioners. Definitive diagnosis is made more complicated by the fact that multiple metabolic disorders can occur simultaneously.

Table 1. Clinical signs and average incidences of the most common clinical metabolic disorders affecting dairy cows in early lactation. Clinical signs compiled from Scott et al. (2011), Parkinson et al. (2019) and the author's clinical experience.

Disorder	Alternative Name	Clinical Signs	Median incidence*
Ketosis	Acetonaemia Hyperketonaemia	↓ milk production, ↓ appetite, weight loss, acetone smell on breath, neurological signs (excitement, licking, chewing, aggression, incoordination)	3.30%
Hypocalcaemia	Milk fever	Stage 1: ↓ milk production, ↓ appetite, ↓ rumen motility, neurological signs (hyperexcitability, fine muscle tremors, ataxia/gait abnormalities), dry faeces and constipation Stage 2: sternal recumbency, depression, 'S'-bend in neck, dry muzzle and cold extremities, ↓ temperature, ↑ heart rate (weak), rumen stasis and bloat Stage 3: lateral recumbency, coma progressing to death	2.82%
Hypomagnesamia	Tetany Grass tetany Grass staggers	↑ heart rate (strong), ↑ temperature, neurological signs (hyperexcitability, aggression, teeth grinding, hypersalivation, muscle and ear twitching) progressing to ataxia, incoordination, recumbency, tetanic muscle spasm, seizures and death	0.15%

* Incidence reported as percent per cow per year. Source: Pryce et al. (2016).

1.3.2 Sub-clinical metabolic disorders

While not associated with obvious clinical signs, subclinical metabolic disorders still have significant negative effects on animal health, welfare and performance. Diagnosis of subclinical disorders is based on clinical pathology testing of biofluids such as blood, milk and/or urine. The prevalence of subclinical disorders is significantly higher than that of clinical disorders. For example, the herd-level prevalence of subclinical ketosis in western Europe has been reported to be 41.0%, compared with just 1.6% for clinical ketosis (Berge and Vertenten, 2014). Similarly, Roberts and McDougall (2019) recently estimated the mean herd-level prevalence of subclinical hypocalcaemia in New Zealand to be 52%, much higher than the 2% prevalence of clinical hypocalcaemia previously reported by the same research group (McDougall, 2001). For this reason, clinical metabolic diseases are often seen as being "the tip of the iceberg".

Arguably the most important aspect of subclinical disorders is their associations with other adverse health, fertility and production outcomes. Compared to normocalcaemic animals,

cows with subclinical hypocalcaemia have been shown to, (1) have reduced immune function (Martinez et al., 2012), (2) be at greater risk of developing clinical diseases such as ketosis, metritis, endometritis, retained foetal membranes and abomasal displacement (Martinez et al., 2012, Ribeiro et al., 2013, Rodríguez et al., 2017), (3) have reduced feed intake (Hansen et al., 2003) and rumen contractility (Jørgensen et al., 1998, Hansen et al., 2003), and (4) suffer an exaggerated degree of negative energy balance (Martinez et al., 2012, Ribeiro et al., 2013). Similarly, subclinical ketosis is associated with (1) reduced reproductive performance (Ospina et al., 2010a, Compton et al., 2014), (2) increased risk of clinical diseases such as abomasal displacement and metritis (LeBlanc et al., 2005, Ospina et al., 2010b, Compton et al., 2014), (3) decreased milk production (Duffield et al., 2009, Ospina et al., 2010a, McArt et al., 2012), and (4) an increased risk of culling (Ospina et al., 2010b, Seifi et al., 2011, McArt et al., 2012). Oetzel (2012) estimated that each case of subclinical hypocalcaemia costs producers approximately $125 USD, while the cost of a single case of subclinical hyperketonaemia has been estimated to be $289 USD (McArt et al., 2015).

1.4 Genetic Selection for improved metabolic health

Metabolic health traits are complex and are influenced by both genetic and environmental effects. For the purposes of animal breeding, we are mostly interested in the proportion of genetic effects that can be passed on from one generation to the next (known as the additive genetic effect). This is summarised by the equation

$$P = A + E$$

where P = phenotype, A = additive genetic effect and E = environmental effects.

Quantitative geneticists use mathematical models to estimate the additive genetic and environmental effects. The additive genetic effect is also known as the estimated breeding value (EBV) of an animal. The relative importance of genetic and environmental effects varies depending on both the trait and the population being studied. The proportion of phenotypic variation ($\sigma^2{}_P$) that can be explained by the additive genetic variance ($\sigma^2{}_A$) is termed the heritability (h^2).

$$h^2 = \frac{\sigma^2{}_A}{\sigma^2{}_P}$$

The mathematical model most commonly used to estimate breeding values is best linear unbiased prediction (BLUP); a linear mixed model which can be summarised by the equation

$$y = Xb + Zu + e$$

where y is a vector of phenotypic observations, b is a vector of fixed effects (e.g. herd-year-season, age, parity, stage of lactation etc.), u is a vector of random genetic effects, e is a vector of the random residual effects, and X and Z are design matrices allocating phenotypes to b and u, respectively. The solutions to u are the EBVs of individuals.

Breeding values for metabolic health traits are commercially available in many countries, including the USA (Council on Dairy Cattle Breeding, 2018), Germany and Austria (VIT, 2020), and the Scandinavian countries (Nordic Cattle genetic Evaluation, 2019). The phenotypes used in these genetic evaluations are predominantly producer and/or veterinarian recorded health data (Johansson et al., 2008, Egger-Danner et al., 2012, Parker Gaddis et al., 2014). Although widely available, the usefulness of these phenotypes is limited by (1) underreporting and inconsistent recording (Østerås et al., 2007), (2) the potential for misdiagnoses, and (3) the fact that trait definitions are often restricted to binary clinical disease events which inherently do not capture important information on subclinical disorders. Consequently, heritability estimates for metabolic health traits are generally low (Pryce et al., 2016) (Figure 6).

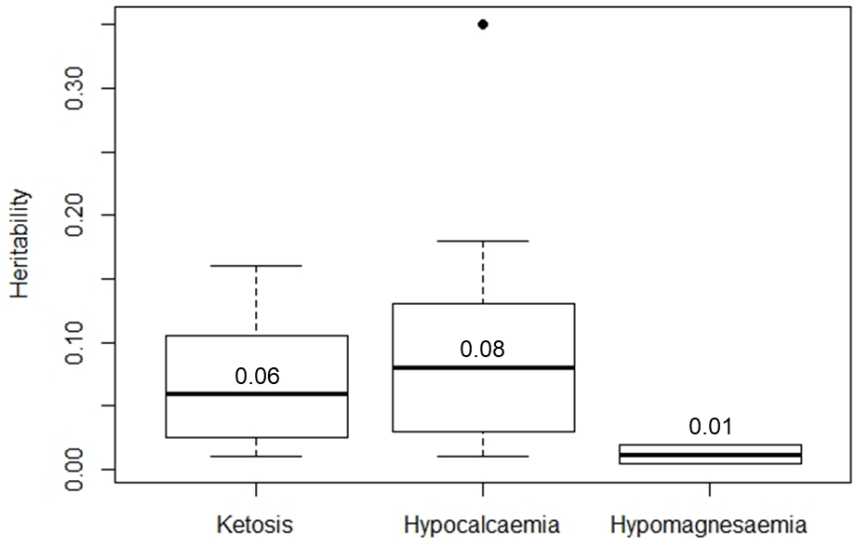


Figure 6. Heritability estimates for clinical metabolic diseases in dairy cows, reported in the literature between 1995 and 2015. Median figure labelled and shown as horizontal line, interquartile range as box, and range as whiskers. Source: Pryce et al. (2016).

There is therefore, much interest in identifying accurate and objective phenotypes for use in genetic evaluations of metabolic health. Such phenotypes are relatively expensive and difficult to measure, making them unsuitable for use in traditional genetic evaluations. They are, however, good candidates for genomic selection.

1.4.1 Genomic selection

Traditional quantitative genetic models rely on pedigree information to estimate the genetic variances of traits. This approach has been extremely successful, however large amounts of phenotypic and pedigree information from many thousands of individuals are required to accurately estimate genetic parameters. The reliability of a bull's breeding value (how close the estimated breeding value is to the animal's true breeding value) is a function of both the heritability of the trait and the number of progeny with phenotypic records (Gonzalez-Recio, 2014). Current Australian breeding value reliabilities for milk production (highly heritable) and fertility (lowly heritable) traits are shown in Table 2.

Table 2. Average reliabilities of Australian breeding values for milk production and fertility traits for first proof bulls (between 70 and 100 progeny with phenotypic data) and proven bulls (> 200 progeny with phenotypic data). (Source: Datagene, Bundoora, Australia)

Trait	First Proof: Small progeny group (70 – 100 daughters)	Proven: Large progeny group (> 200 daughters)
Milk production	92%	97%
Fertility	86%	93%

Genomic selection (GS) replaces or augments pedigree information with DNA markers spread across the genome to predict the genetic merit of an individual (Meuwissen et al., 2001). Firstly, a prediction equation is created using a reference population for which both genotypic and phenotypic information are known. This prediction equation can then be used to estimate genomic breeding values for selection candidates, for which only genomic information is known (Figure 7). One of the most exciting benefits of GS over traditional pedigree-based genetic evaluations is that significant genetic improvement can be achieved in lowly-heritable traits, even with a relatively small reference population (Calus et al., 2013). This makes it an ideal breeding tool for expensive and difficult-to-measure traits, such as residual feed intake (Pryce et al., 2012), heat tolerance (Nguyen et al., 2016) and potentially, metabolic health (Chesnais et al., 2016).

The accuracy of GS is affected by several factors, including (1) the effective population size, (2) the size of the reference population, (3) the heritability of the trait, and (4) the genetic architecture of the trait (Daetwyler et al., 2008, Meuwissen Theo, 2009, Hayes et al., 2010, Gonzalez-Recio et al., 2014).

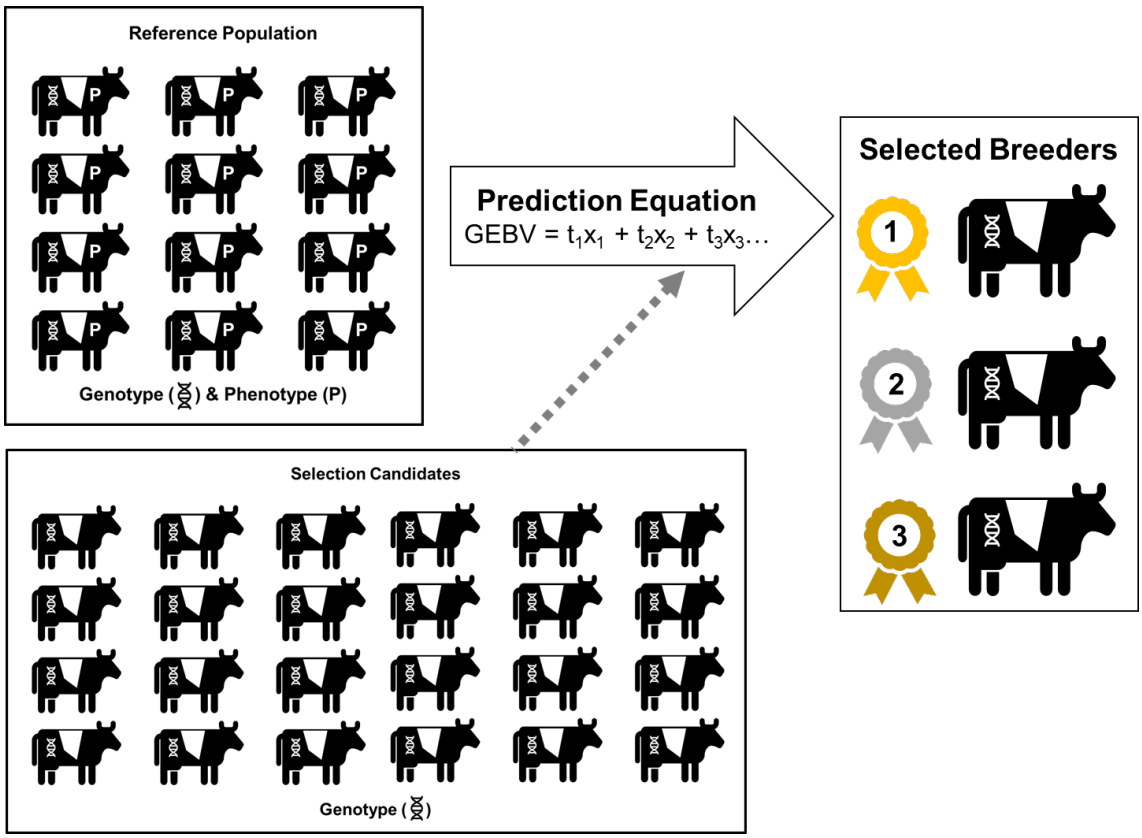


Figure 7. In genomic selection, a prediction equation is constructed using data from a genotyped reference population with high quality phenotypic information. This prediction equation can then be used to estimate genomic breeding values for selection candidates for which only genomic information is available.

1.5 Metabolic phenotypes of metabolic health

1.5.1 Traditional serum metabolic profiles

Serum metabolic profiles were first proposed in the early 1970s as an early warning system for metabolic perturbations in dairy cows (Payne et al., 1970). Metabolic profiling involves measuring the concentrations of a suite of biomarkers in serum which provide objective information on the nutritional status and metabolic health of an animal (Macrae et al., 2006). A biomarker is an objective and measurable indicator and/or predictor of any biological state or process, including normal biological processes, pathological processes, responses to environmental exposure, and/or pharmacologic responses to a therapeutic intervention (National Institutes of Health, 2001). Biomarkers can include amongst other things, DNA, RNA, proteins (including enzymes), and/or small metabolites.

Commonly used biomarkers in metabolic profiling include those associated with energy balance (BHBA and NEFA), macro-mineral status (Ca and Mg), protein nutritional status (urea and albumin) and immune status (globulins and albumin to globulin ratio (**A:G**))

(Whitaker, 2004, Anderson, 2009). Quantification of these biomarkers is most commonly done using colourimetric assays, based on reagents that undergo a measurable and proportional colour change in the presence of the analyte.

Biomarker concentration thresholds (i.e. biomarker concentrations, above or below which adverse health, reproductive and/or production outcomes are more likely to be seen) are determined using epidemiological studies. In clinically healthy animals, subclinical ketosis has been defined as a blood BHBA concentration > 1.2 mmol/L (Ospina et al., 2010b, Compton et al., 2014); subclinical hypocalcaemia as blood calcium concentration < 2.00 mmol/L (Degaris and Lean, 2008); and subclinical hypomagnesaemia as blood magnesium concentration < 0.62 mmol/L (Anderson, 2009). Optimal concentration ranges for some of the most common serum metabolic profile biomarkers are shown in Table 3.

Table 3. Upper and lower concentration thresholds for serum biomarkers commonly used for metabolic profile analyses.

		Optimum Concentration of Serum Metabolites	
Metabolite	Reference	Lower Threshold	Upper Threshold
BHBA	(McArt et al., 2012, Compton et al., 2014)	-	1.2 mmol/L
NEFA	(Ospina et al., 2010a)	-	0.7 mmol/L
Ca	(Degaris and Lean, 2008)	2.0 mmol/L	-
Mg	(Anderson, 2009)	0.62 mmol/L	-
Urea	(Butler et al., 1996, Macrae et al., 2006)	1.7 mmol/L	6.78 mmol/L
Albumin	(Whitaker, 2004)	30 g/L	-
Globulin	(Whitaker, 2004)	-	50 g/L

Heritability estimates for biomarker traits tend to be higher than those for traditional disease record traits. For example, heritability estimates for BHBA range from 0.07 (Tsiamadis et al., 2016) to 0.40 (Oikonomou et al., 2008), compared to 0.01 (Kadarmideen et al., 2000) to 0.16 (Heringstad et al., 2005) for clinical ketosis records. Promisingly, Rius-Vilarrasa et al. (2018) have reported favourable genetic correlations between BHBA and acetone measurements, and clinical ketosis records. Furthermore, inclusion of BHBA and acetone measurements as indicator traits for metabolic disorders, in the Nordic Cattle Genetic Evaluation, has led to breeding values with increased reliability (Rius-Vilarrasa et al., 2018).

1.5.2 Fourier transform mid-infrared spectroscopy of milk

Fourier transform mid-infrared (MIR) spectroscopy is a rapid, non-destructive analytical technique which exploits inherent differences in the way electromagnetic waves interact with matter. Briefly, when a beam of radiation in the mid-infrared region (2,500-25,00 nm) is shone through a substance, different functional groups within the molecules in that substance absorb different wavelengths of radiation. Therefore, depending on the molecules present in a sample, some frequencies of MIR radiation are absorbed, others are partially absorbed, and others are not absorbed at all. The intensity of absorption (y-axis) versus wavelength (x-axis) makes up the absorption spectra of the substance (Pasquini, 2003), which can be considered its "molecular fingerprint".

MIR spectroscopy is the worldwide method of choice for determining the quality and composition (i.e. fat, protein and lactose percentage) of liquid milk, performed as part of routine milk recording (Grelet et al., 2015). A representative MIR absorbance spectrum from liquid milk is shown in Figure 8.

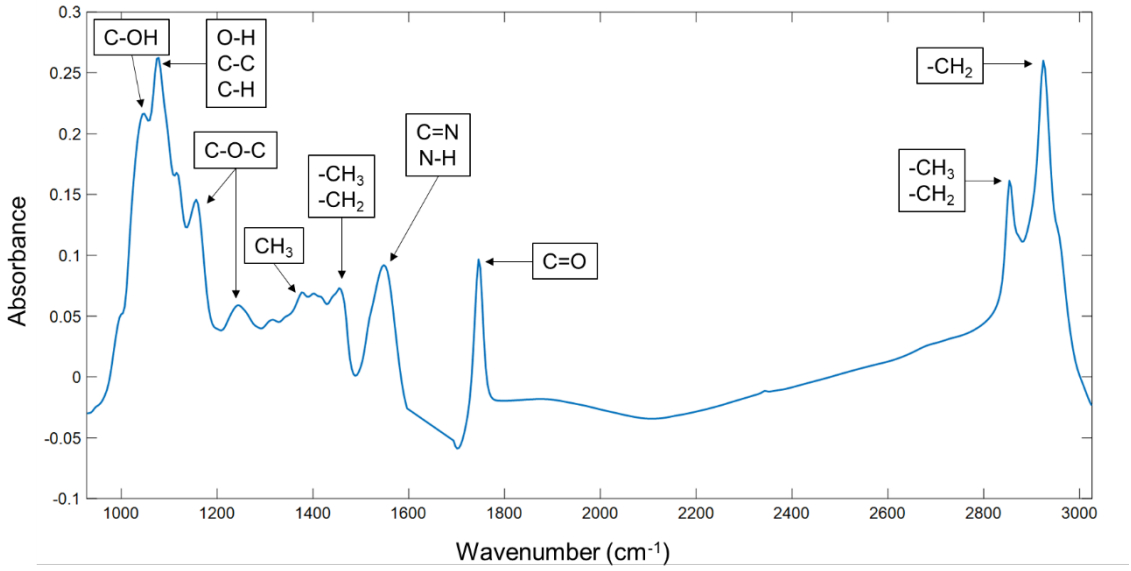


Figure 8. Representative Fourier transform mid-infrared (MIR) absorbance spectrum of liquid milk. Regions associated with water absorption have been removed. Annotations show molecular bonds which absorb MIR radiation at given wavelengths. Adapted from Grelet et al. (2015).

In addition to information about major milk components, milk MIR spectral data has been shown to contain valuable latent biochemical information about fine milk composition (Gengler et al., 2016). This information has shown promise as the basis for novel, milk-based phenotypes for use in genetic evaluations for improved product quality, animal health, and reduced environmental impact (De Marchi et al., 2014, Gengler et al., 2016). In the context of this thesis, of particular interest are MIR spectral predictions of measured energy balance

(McParland et al., 2011, McParland et al., 2012), and the concentration of ketone bodies in milk (de Roos et al., 2007, van Knegsel et al., 2010, Grelet et al., 2016) and serum (Gelé et al., 2015, Belay et al., 2017, Pralle et al., 2018).

The approach for MIR-based phenotyping is similar to that described previously for genomic selection. Firstly, a prediction equation is created using detailed phenotypic and MIR spectral data collected from a reference population. This equation is then used to predict phenotypes for other animals using spectra obtained from milk recording. In Australia, approximately 46% of cows participate in routine milk recording (Pryce et al., 2018), however, since 2001 there has been a decreasing trend in herd participation (Gonzalez-Recio et al., 2014). As more milk-recording laboratories upgrade to MIR technology, MIR spectral data could become a cost-effective way to "scale-up" high-value phenotypes for complex traits such as metabolic health. Furthermore, if sufficiently accurate, MIR-predicted phenotypes could also be a valuable source of animal health data for producers, nutritionists and veterinarians, offering a potential paradigm shift for milk-recording into the future.

1.5.3 Metabolomics: next generation metabolic profiling

Metabolomics involves the high-throughput, synchronous characterisation of the small metabolites in a biological matrix (Wishart, 2008). These metabolites are known collectively as the metabolome, and provide a snapshot of the metabolic state of an organism at a given point in time. In dairy cows, the metabolome is the end-product of complex interactions between host genetics, the rumen microbiome, and the environment.

The most commonly used analytical techniques in metabolomics are proton nuclear magnetic resonance spectroscopy (^{1}H NMR), and mass spectroscopy coupled with either gas or liquid chromatography (GC-MS and LC-MS, respectively). While mass spectrometry-based techniques are more sensitive and specific, NMR methods have the advantage of being (1) non-destructive, (2) reproducible, (3) higher-throughput and (4) inherently quantitative.

The principles of ^{1}H NMR spectroscopy are summarised in Figures 9 and 10. Briefly, all atomic nuclei possess both charge and spin, and subsequently each has their own magnetic moment (i.e. each atomic nucleus acts as a tiny magnet). Under normal conditions, the orientations of these nuclear magnetic fields are randomly distributed (Figure 9a). However, when placed in an external magnetic field (designated B_0), these individual nuclear magnetic fields become aligned in either the same or opposite direction as B_0 (Figure 9b).

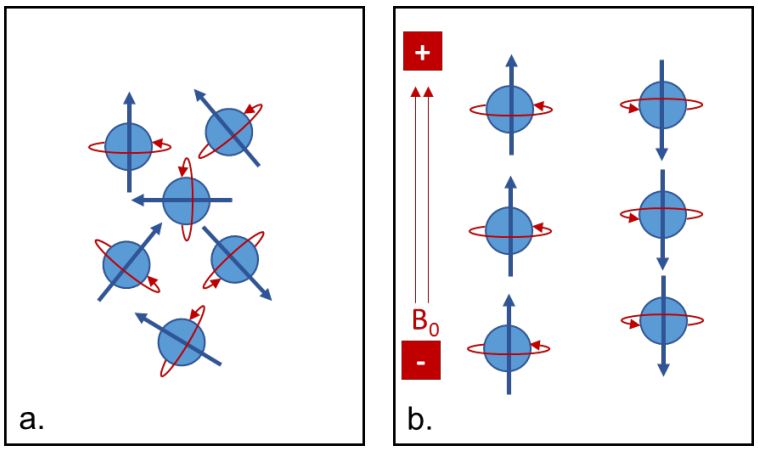


Figure 9. Representation of the spin (red arrow) and direction of magnetic field (blue arrow) of atomic nuclei in the absence (a) and presence (b) and an external magnetic field (B_0). Adapted from diagrams produced by Dr A. Maher (2016).

The aligned nuclei can exist in 2 distinct energy states; a low-energy state (α) and a high-energy state (β). Applying an external energy source (in NMR spectroscopy this is a radiofrequency (RF) pulse) causes nuclei to jump from their low energy state to their high-energy state; a phenomenon known as resonance (Figures 10a and 10b). The amount of energy required to achieve resonance depends on the electromagnetic environment around the nucleus. For example, nuclei that are surrounded by electrons are said to be "shielded" and require more energy to reach resonance than so-called "de-shielded" nuclei that are not surrounded by electrons. When the external energy source is removed, nuclei relax and return to their low-energy state via a spinning motion (Figure 10d). This, in turn, generates a detectable electromagnetic signal known as free induction decay (FID), the magnitude of which is directly proportional to the energy required to reach resonance.

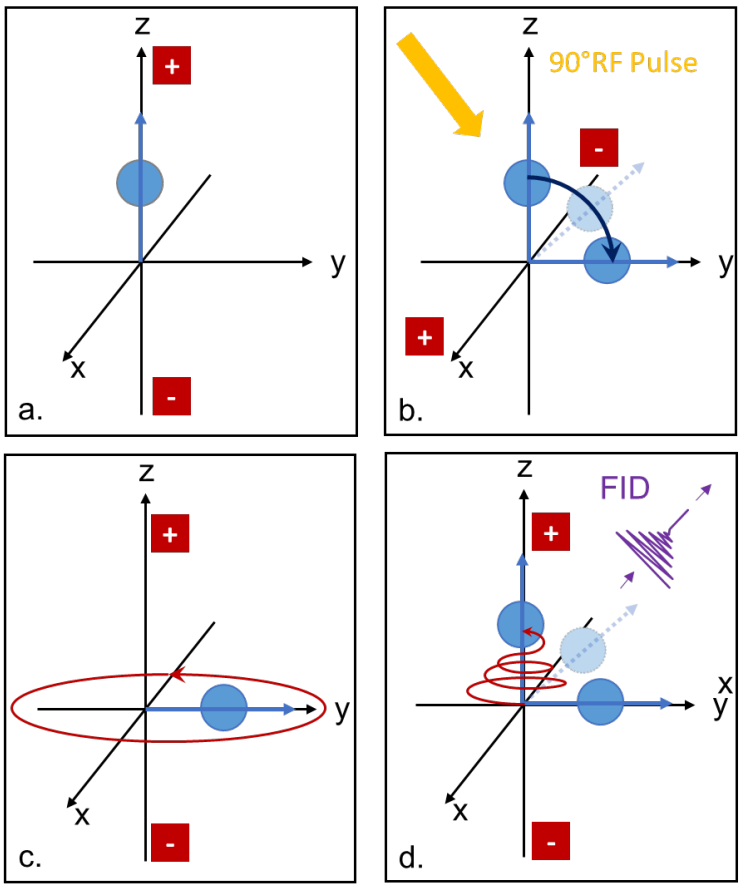
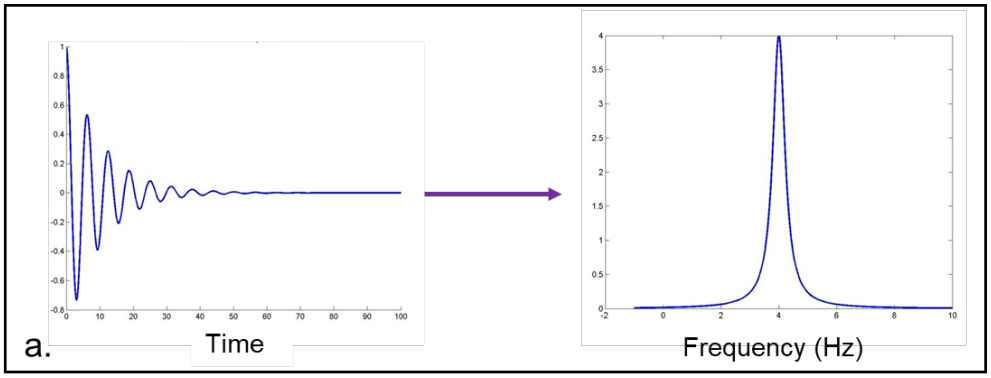
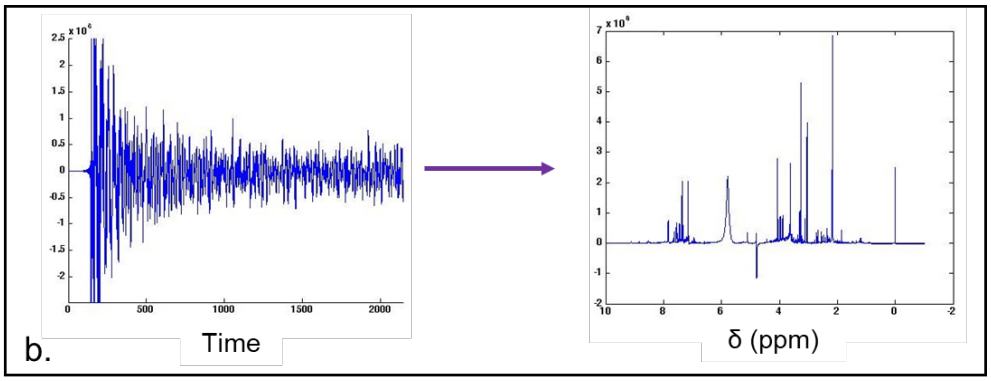
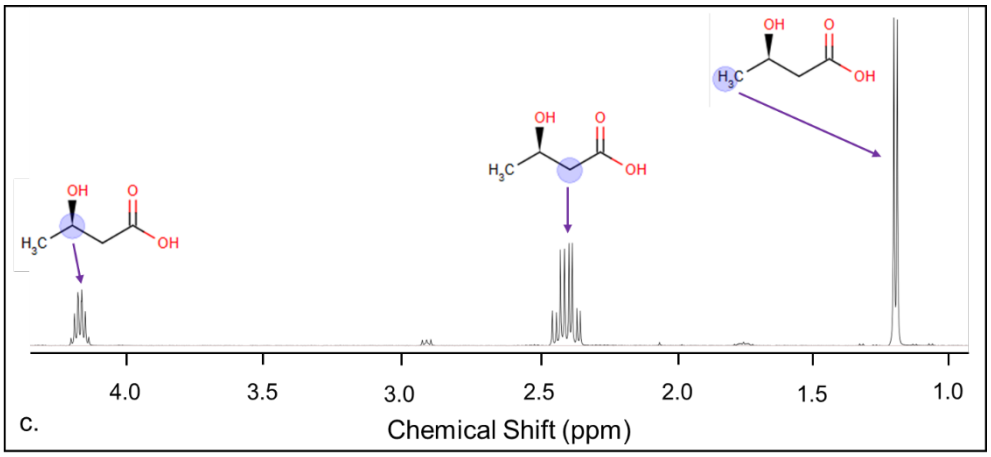


Figure 10. Representation of nuclear magnetic resonance; (a) an atomic nucleus aligned with an external magnetic field and in a low energy state (α), (b) application of an external radiofrequency (RF) pulse at 90° leading to excitation to (c) an excited or high-energy (β) state, and (d) relaxation back to the low-energy state leading to generation of free induction decay (FID) signal. Adapted from diagrams produced by Dr A. Maher (2016).

In ^{1}H NMR, each proton in a sample produces a signal and the overall FID from a sample is a summation of all of the signals emitted by all protons present. The FID can then be Fourier-transformed to produce a spectrum (Figure 11). The number, shape and location of peaks on an ^{1}H NMR spectrum depends on both the number of protons present in the sample, and the molecular environments around these protons (e.g. the number of protons attached to adjacent carbon atoms). A ^{1}H NMR spectrum therefore contains detailed information about the molecular structure of hydrogen-containing compounds in the sample being analysed. This makes it an ideal analytical technique for the study of organic compounds. The intensity of peaks is directly proportional to abundance, meaning NMR spectroscopy is inherently quantitative. Examples of FIDs and corresponding NMR spectra derived from both simple and complex samples are shown in Figure 11.


a.
Time
Frequency (Hz)


b.
Time
δ (ppm)


OH
O
H3C
OH
4.0
3.5
3.0
2.5
2.0
1.5
1.0
c.
Chemical Shift (ppm)

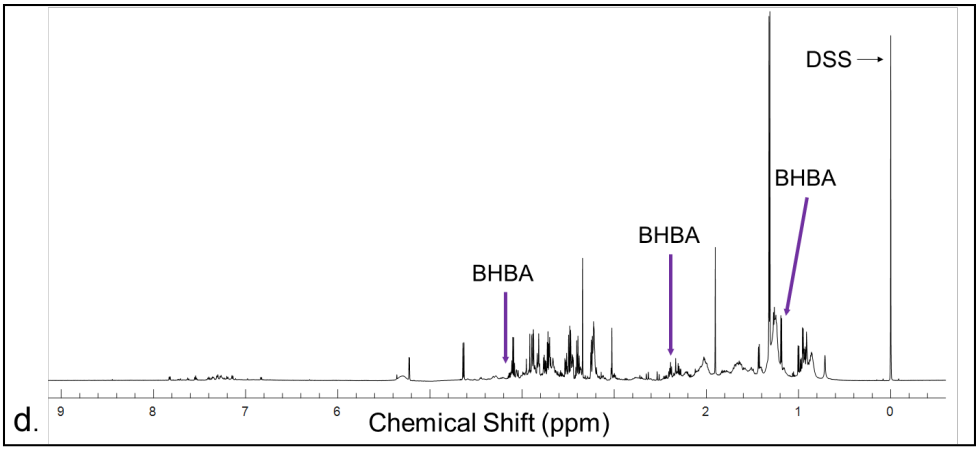

DSS
BHBA
BHBA
BHBA
9
8
7
6
2
1
0
d.
Chemical Shift (ppm)

Figure 11. Examples of free induction decay (FID) signals, and corresponding spectra following Fourier-transformation, derived from nuclear magnetic resonance spectroscopy of (a) simple and (b) complex samples. Figure (c) shows the ^{1}H NMR spectrum of β-hydroxybutyrate (BHBA), including peak assignments for different protons, and figure (d) shows the location of BHBA peaks relative to an internal standard (sodium trimethylsilylpropanesulfonate (DSS)) in a representative ^{1}H NMR spectrum derived from bovine serum. Adapted from diagrams produced by Dr A. Maher (2016), raw BHBA spectral file obtained from HMDB (Wishart et al. 2009), and peak annotations obtained from the Chenomx NMR suite software v.8.4 (Chenomx Inc., Edmonton, AB, Canada).

While a traditional metabolic profile consists of approximately 10 biomarkers, modern metabolomic techniques provide quantitative or semi-quantitative characterisation of tens (NMR) to thousands (LC-MS) of metabolites. Such high-resolution data offers exciting opportunities to better understand and characterise the complex physiological and biochemical processes taking place during the transition period (Kenéz et al., 2016, Ceciliani et al., 2018). This, in turn, can facilitate identification of novel metabolic phenotypes (metabotypes) associated with existing health phenotypes. For example, Sun et al. (2014) used ^{1}H NMR metabolomics to identify 25 metabolites in serum that were differentially expressed in cows with and without clinical and sub-clinical ketosis. Furthermore, given the limitations of existing phenotypes, metabolomic studies could be used to define completely novel metabolic phenotypes which better characterize a successful and/or failed transition from pregnancy to lactation. For example, Hailemariam et al. (2014) identified a three-metabolite biomarker panel (carnitine, propionyl carnitine, and lysophosphatidylcholine acyl C14:0) that could predict the occurrence of one or more periparturient disorders (including metritis, mastitis, laminitis, or retained placenta) with 87% sensitivity and 85% specificity.

1.6 This thesis

This thesis contains four research chapters which have been published in peer-reviewed scientific journals. **Chapter two** presents a study of the genetic parameters and genomic prediction accuracies of "traditional" serum metabolic profile biomarkers. **Chapter three** contains description of the use of MIR spectroscopy of milk as a high-throughput phenotyping tool to predict biomarkers of metabolic health; and **chapters four and five** present studies on the use of ^{1}H NMR metabolomics to (1) better characterize existing metabolic health biomarkers, and (2) identify novel metabolic health biomarkers. Finally, **chapter six** presents a general discussion of the findings of this research, and how they might be integrated in order to develop a breeding value for metabolic resilience which is sufficiently accurate to be used by the Australian dairy industry.

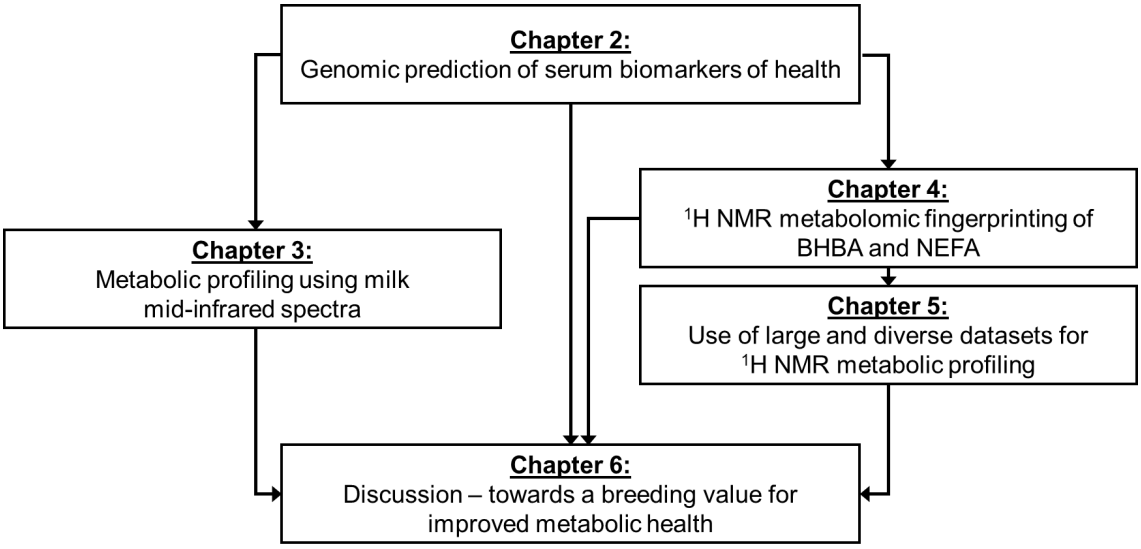


Figure 12. Outline of this thesis, including chapters published in peer-reviewed journals (chapters 2 to 5, inclusive) and the general discussion (chapter 6).

References

Anderson, D. E. 2009. Current Veterinary Therapy: Food Animal Practice. Saunders Elsevier, St. Louis, MO.

Bauman, D. E. and W. B. Currie. 1980. Partitioning of nutrients during pregnancy and lactation: a review of mechanisms involving homeostasis and homeorhesis. J. Dairy Sci. 63:1514-1529. 10.3168/jds.S0022-0302(80)83111-0

Belay, T. K., B. S. Dagnachew, Z. M. Kowalski, and T. Ådnøy. 2017. An attempt at predicting blood β-hydroxybutyrate from Fourier-transform mid-infrared spectra of milk using multivariate mixed models in Polish dairy cattle. J. Dairy Sci. 100:6312-6326. 10.3168/jds.2016-12252

Bell, A. W. 1995. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. J. Anim. Sci. 73:2804-2819. 10.2527/1995.7392804x

Berge, A. C. and G. Vertenten. 2014. A field study to determine the prevalence, dairy herd management systems, and fresh cow clinical conditions associated with ketosis in western European dairy herds. J. Dairy Sci. 97:2145-2154. 10.3168/jds.2013-7163

Berry, D. P., E. Wall, and J. E. Pryce. 2014. Genetics and genomics of reproductive performance in dairy and beef cattle. Animal. 1:105-121. 10.1017/S1751731114000743

Boichard, D. and M. Brochard. 2012. New phenotypes for new breeding goals in dairy cattle. Animal. 6:544-550. 10.1017/s1751731112000018

Butler, W. R., J. J. Calaman, and S. W. Beam. 1996. Plasma and milk urea nitrogen in relation to pregnancy rate in lactating dairy cattle. J. Anim. Sci. 74:858-865. 10.2527/1996.744858x

Calus, M. P. L., Y. de Haas, M. Pszczola, and R. F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. 7:183-191. 10.1017/S1751731112001450

Cameron, R. E. B., P. B. Dyk, T. H. Herdt, J. B. Kaneene, R. Miller, H. F. Bucholtz, J. S. Liesman, M. J. Vandehaar, and R. S. Emery. 1998. Dry cow diet, management, and energy balance as risk factors for displaced abomasum in high producing dairy herds. J. Dairy Sci. 81:132-139. 10.3168/jds.S0022-0302(98)75560-2

Ceciliani, F., C. Lecchi, C. Urh, and H. Sauerwein. 2018. Proteomics and metabolomics characterizing the pathophysiology of adaptive reactions to the metabolic challenges during

the transition from late pregnancy to early lactation in dairy cows. J. Proteom. 178:92-106. 10.1016/j.jprot.2017.10.010

Chamberlain, A. T. and J. M. Wilkinson. 1996. Feeding the dairy cow. Chalcombe Publishers, Shedfield, Hampshire, UK.

Chesnais, J. P., T. A. Cooper, G. R. Wiggans, M. Sargolzaei, J. E. Pryce, and F. Miglior. 2016. Using genomics to enhance selection of novel traits in North American dairy cattle. J. Dairy Sci. 99:2413-2427. 10.3168/jds.2015-9970

Coffey, M. 2020. Dairy cows: in the age of the genotype, #phenotypeisking. Anim. Front. 10:19-22. 10.1093/af/vfaa004

Cole, J. B., S. A. E. Eaglen, C. Maltecca, H. A. Mulder, and J. E. Pryce. 2020. The future of phenomics in dairy cattle breeding. Anim. Front. 10:37-44. 10.1093/af/vfaa007

Commonwealth Bureau of Census and Statistics. 1925. Year Book Australia. Melbourne, Australia.

Compton, C., L. Young, and S. McDougall. 2014. Subclinical ketosis in post-partum dairy cows fed a predominantly pasture-based diet: defining cut-points for diagnosis using concentrations of beta-hydroxybutyrate in blood and determining prevalence. N. Z. Vet. J. 63:241-248. 10.1080/00480169.2014.999841

Council on Dairy Cattle Breeding. 2018. CDCB Health Traits. Accessed 16th June 2020. https://www.uscdcb.com/cdcb-health-traits-2/

Curtis, C. R., H. N. Erb, C. J. Sniffen, R. D. Smith, and D. S. Kronfeld. 1985. Path analysis of dry period nutrition, postpartum metabolic and reproductive disorders, and mastitis in holstein cows(1). J. Dairy Sci. 68:2347-2360. 10.3168/jds.S0022-0302(85)81109-7

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395. 10.1371/journal.pone.0003395

Dairy Australia. 2019. The Australian dairy industry at a glance 2018-19. Accessed 17th June 2020. https://www.dairyaustralia.com.au/industry/farm-facts/dairy-at-a-glance

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171-1186. 10.3168/jds.2013-6799

de Roos, A. P. W., H. J. C. M. van Den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for subclinical ketosis in dairy cattle by fourier transform infrared spectrometry. J. Dairy Sci. 90:1761-1766. 10.3168/jds.2006-203

Degaris, P. J. and I. J. Lean. 2008. Milk fever in dairy cows: A review of pathophysiology and control principles. Vet. J. 176:58-69. 10.1016/j.tvjl.2007.12.029

Drackley, J. K. 1999. Biology of dairy cows during the transition period: the final frontier? J. Dairy Sci. 82:2259-2273. 10.3168/jds.S0022-0302(99)75474-3

Duffield, T. F., K. D. Lissemore, B. W. McBride, and K. E. Leslie. 2009. Impact of hyperketonemia in early lactation dairy cows on health and production. J. Dairy Sci. 92:571-580. 10.3168/jds.2008-1507

Egger-Danner, C., J. B. Cole, J. E. Pryce, N. Gengler, B. Heringstad, A. Bradley, and K. F. Stock. 2015. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal. 9:191-207. 10.1017/s1751731114002614

Egger-Danner, C., B. Fuerst-Waltl, W. Obritzhauser, C. Fuerst, H. Schwarzenbacher, B. Grassauer, M. Mayerhofer, and A. Koeck. 2012. Recording of direct health traits in Austria—Experience report with emphasis on aspects of availability for breeding purposes. J. Dairy Sci. 95:2765-2777. 10.3168/jds.2011-4876

Gelé, M., M. Ferrand-Calmels, L. Brun-Lafleur, A. Werner, and F. Gollé-Leidreiter. 2015. Predicting the risk of ketosis using mid infrared spectrometry. Pages 19-24 in Proc. ICAR Tech. Meet. ICAR, Krakow, Poland.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M. L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071-4079. 10.3168/jds.2015-10140

Goddard, M.E., B.J. Hayes, and T.H.E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409–421. doi:10.1111/j.1439-0388.2011.00964.x.

Goff, J. P. and R. L. Horst. 1997. Physiological changes at parturition and their relationship to metabolic disorders(1, 2). J. Dairy Sci. 80:1260-1268. 10.3168/jds.S0022-0302(97)76055-7

Gonzalez-Recio, O., M. P. Coffey, and J. E. Pryce. 2014. On the value of the phenotypes in the genomic era. J. Dairy Sci. 97:7905-7915. 10.3168/jds.2014-8125

Grant, R. J. and J. L. Albright. 1995. Feeding behavior and management factors during the transition period in dairy cattle. J. Anim. Sci. 73:2791-2803. doi: 10.2527/1995.7392791x

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816-4825. 10.3168/jds.2015-10477

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150-2160. 10.3168/jds.2014-8764

Grummer, R. R. 1995. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. J. Anim. Sci. 73:2820-2833. 10.2527/1995.7392820x

Hailemariam, D., R. Mandal, F. Saleem, S. M. Dunn, D. S. Wishart, and B. N. Ametaj. 2014. Identification of predictive biomarkers of disease state in transition dairy cows. J. Dairy Sci. 97:2680-2693. 10.3168/jds.2013-6803

Hansen, S. S., P. Nørgaard, C. Pedersen, R. J. Jørgensen, L. S. B. Mellau, and J. D. Enemark. 2003. The effect of subclinical hypocalcaemia induced by Na_2EDTA on the feed intake and chewing activity of dairy cows. Vet. Res. Commun. 27:193-205. 10.1023/A:1023340506782

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139. 10.1371/journal.pgen.1001139

Heitmann, R. N., D. J. Dawes, and S. C. Sensenig. 1987. Hepatic ketogenesis and peripheral ketone body utilization in the ruminant. J. Nutr. 117:1174-1180. 10.1093/jn/117.6.1174

Herdt, T. H. 2000. Ruminant adaptation to negative energy balance: influences on the etiology of ketosis and fatty liver. Vet. Clin. North Am. Food Anim. Pract. 16:215-230. 10.1016/S0749-0720(15)30102-X

Heringstad, B., Y. M. Chang, D. Gianola, and G. Klemetsdal. 2005. Genetic analysis of clinical mastitis, milk fever, ketosis, and retained placenta in three lactations of norwegian red cows. J. Dairy Sci. 88:3273-3281. 10.3168/jds.S0022-0302(05)73010-1

Hoffman, K., R. DeClue, and D. L. Emmick. 2000. Prescribed grazing and feeding management of lactating dairy cows. NYS grazing lands conservation initiative. USDA-National Resources Conservation Service, Syracuse, NY, USA. Accessed May 20, 2020. http://grazingguide.net/pdfs/cow-feeding-mgt.pdf

Johansson, K., J. Pösö, U. S. Nielsen, J.-Å. Eriksson, and G. P. Aamand. 2008. Joint genetic evaluation of other disease traits in Denmark, Finland and Sweden. Interbull Bull. 38:107-112.

Jørgensen, R. J., N. R. Nyengaard, S. Hara, J. M. Enemark, and P. H. Andersen. 1998. Rumen motility during induced hyper- and hypocalcaemia. Acta Vet. Scand. 39:331-338.

Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 27:83-84. 10.1017/S1463981500040577

Kenéz, Á., S. Dänicke, U. Rolle-Kampczyk, M. Von Bergen, and K. Huber. 2016. A metabolomics approach to characterize phenotypes of metabolic transition from late pregnancy to early lactation in dairy cows. Metabolomics. 12:1-11. 10.1007/s11306-016-1112-8

Lean, I. J. and P. J. Degaris. 2010. Transition Cow Management. A review for nutritional professionals, veterinarians and farm advisers. Dairy Australia, Melbourne, Australia.

LeBlanc, S. J., K. E. Leslie, and T. F. Duffield. 2005. Metabolic predictors of displaced abomasum in dairy cattle. J. Dairy Sci. 88:159-170. 10.3168/jds.S0022-0302(05)72674-6

Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? J. Dairy Sci. 84:1277-1293. 10.3168/jds.S0022-0302(01)70158-0

Macrae, A. I., D. A. Whitaker, E. Burrough, A. Dowell, and J. M. Kelly. 2006. Use of metabolic profiles for the assessment of dietary adequacy in UK dairy herds. Vet. Rec. 159:655. 10.1136/vr.159.20.655

Martinez, N., C. A. Risco, F. S. Lima, R. S. Bisinotto, L. F. Greco, E. S. Ribeiro, F. Maunsell, K. Galvão, and J. E. P. Santos. 2012. Evaluation of peripartal calcium status, energetic

profile, and neutrophil function in dairy cows at low or high risk of developing uterine disease. J. Dairy Sci. 95:7158-7172. 10.3168/jds.2012-5812

McArt, J. A. A., D. V. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. J. Dairy Sci. 95:5056-5066. 10.3168/jds.2012-5443

McArt, J. A. A., D. V. Nydam, G. R. Oetzel, T. R. Overton, and P. A. Ospina. 2013. Elevated non-esterified fatty acids and β-hydroxybutyrate and their association with transition dairy cow performance. Vet. J. 198:560-570. 10.1016/j.tvjl.2013.08.011

McArt, J. A. A., D. V. Nydam, and M. W. Overton. 2015. Hyperketonemia in early lactation dairy cattle: a deterministic estimate of component and total cost per case. J. Dairy Sci. 98:2043-2054. 10.3168/jds.2014-8740

McDougall, S. 2001. Effects of periparturient diseases and conditions on the reproductive performance of New Zealand dairy cows. N. Z. Vet. J. 49:60-67. 10.1080/00480169.2001.36204

McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'neill, M. O'donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. J. Dairy Sci. 95:7225-7235. 10.3168/jds.2012-5406

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651-3661. 10.3168/jds.2010-3965

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Meuwissen Theo, H. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. 10.1186/1297-9686-41-35

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251-10271. 10.3168/jds.2017-12968

Biomarkers Definitions Working Group (NIH). 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 69:89-95. 10.1067/mcp.2001.113989

Nguyen, T. T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. 2016. Genomic selection for tolerance to heat stress in Australian dairy cattle. J. Dairy Sci. 99:2849-2862. 10.3168/jds.2015-9685

Nordic Cattle genetic Evaluation. 2019. Genetic Evaluation of Other Diseases. Accessed June 16, 2020. https://www.nordicebv.info/wp-content/uploads/2015/05/Genetic-Evaluation-of-Other-Diseases.pdf

Oetzel, G. R. 2012. An Update on Hypocalcemia on Dairy Farms. Accessed June 16, 2020. http://livestocktrail.illinois.edu/uploads/dairynet/papers/Update%20Hypocalcemia%20Oetzel1.pdf

Oikonomou, G., G. E. Valergakis, G. Arsenos, N. Roubies, and G. Banos. 2008. Genetic profile of body energy and blood metabolic traits across lactation in primiparous holstein cows. J. Dairy Sci. 91:2814-2822. 10.3168/jds.2007-0965

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010a. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. J. Dairy Sci. 93:1596-1603. 10.3168/jds.2009-2852

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010b. Evaluation of nonesterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. J. Dairy Sci. 93:546-554. 10.3168/jds.2009-2277

Østerås, O., H. Solbu, A. O. Refsdal, T. Roalkvam, O. Filseth, and A. Minsaas. 2007. Results and evaluation of thirty years of health recordings in the norwegian dairy cattle population. J. Dairy Sci. 90:4483-4497. 10.3168/jds.2007-0030

Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and C. Maltecca. 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97:3190-3199. 10.3168/jds.2013-7543

Parkinson, T. J., J. J. Vermunt, J. Malmo, and R. Laven. 2019. Diseases of Cattle in Australasia. 2nd ed. Massey University Press, Auckland, New Zealand.

Pasquini, C. 2003. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. J. Braz. Chem. Soc. 14:198-219. 10.1590/S0103-50532003000200006.

Payne, J., S. M. Dew, R. Manston, and M. Faulks. 1970. The use of a metabolic profile test in dairy herds. Vet. Rec. 87:150. 10.1136/vr.87.6.150

Pralle, R. S., K. W. Weigel, and H. M. White. 2018. Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. J. Dairy Sci. 10.3168/jds.2017-14076

Pryce, J. E., J. Arias, P. J. Bowman, S. R. Davis, K. A. Macdonald, G. C. Waghorn, W. J. Wales, Y. J. Williams, R. J. Spelman, and B. J. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95:2108-2119. 10.3168/jds.2011-4628

Pryce, J. E., T. T. T. Nguyen, M. Axford, G. Nieuwhof, and M. Shaffer. 2018. Symposium review: Building a better cow—The Australian experience and future perspectives(1). J. Dairy Sci. 101:3702-3713. 10.3168/jds.2017-13377

Pryce, J. E., K. L. Parker Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B. Heringstad, C. Egger-Danner, K. F. Stock, A. J. Bradley, and J. B. Cole. 2016. Invited review: Opportunities for genetic improvement of metabolic diseases. J. Dairy Sci. 99:6855-6873. 10.3168/jds.2016-10854

Ribeiro, E. S., F. S. Lima, L. F. Greco, R. S. Bisinotto, A. P. A. Monteiro, M. Favoreto, H. Ayres, R. S. Marsola, N. Martinez, W. W. Thatcher, and J. E. P. Santos. 2013. Prevalence of periparturient diseases and effects on fertility of seasonally calving grazing dairy cows supplemented with concentrates. J. Dairy Sci. 96:5682-5697. 10.3168/jds.2012-6335

Rius-Vilarrasa, E., W. F. Fikse, E. Carlén, J.-Å. Eriksson, J. Pöso, U. S. Nielsen, and G. P. Aamand. 2018. Improved genetic evaluation of health traits using metabolic biomarkers in Nordic dairy cattle. No 53 in Proc. Interbull Meet., Auckland, NZ. Interbull, Uppsala, Sweden.

Roberts, K. I. and S. McDougall. 2019. Risk factors for subclinical hypocalcaemia, and associations between subclinical hypocalcaemia and reproductive performance, in pasture-based dairy herds in New Zealand. N. Z. Vet. J. 67:12-19. 10.1080/00480169.2018.1527732

Rodríguez, E. M., A. Arís, and A. Bach. 2017. Associations between subclinical hypocalcemia and postparturient diseases in dairy cows. J. Dairy Sci. 100:7427-7434. 10.3168/jds.2016-12210

Saris, W. H., M. A. van Erp-Baart, F. Brouns, K. R. Westerterp, and F. ten Hoor. 1989. Study on food intake and energy expenditure during extreme sustained exercise: the Tour de France. Int. J. Sports Med. 10 Suppl 1:S26-31. 10.1055/s-2007-1024951

Scott, P. R., C. D. Penny, and A. I. MacRae. 2011. Cattle Medicine. Manson Publishing, London, United Kingdom.

Seifi, H. A., S. J. Leblanc, K. E. Leslie, and T. F. Duffield. 2011. Metabolic predictors of post-partum disease and culling risk in dairy cattle. Vet. J. 188:216-220. 10.1016/j.tvjl.2010.04.007

Sordillo, L. M. 2016. Nutritional strategies to optimize dairy cattle immunity. J. Dairy Sci. 99:4967-4982. 10.3168/jds.2015-10354

Sordillo, L. M. and W. Raphael. 2013. Significance of metabolic stress, lipid mobilization, and inflammation on transition cow disorders. Vet. Clin. North Am. Food Anim. Pract. 29:267-278. 10.1016/j.cvfa.2013.03.002

Sun, L. W., H. Y. Zhang, L. Wu, S. Shu, C. Xia, C. Xu, and J. S. Zheng. 2014. ^{1}H-nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. J. Dairy Sci. 97:1552-1562. 10.3168/jds.2013-6757

Tsiamadis, V., G. Banos, N. Panousis, M. Kritsepi-Konstantinou, G. Arsenos, and G. E. Valergakis. 2016. Genetic parameters of subclinical macromineral disorders and major clinical diseases in postparturient Holstein cows. J. Dairy Sci. 99:8901-8914. 10.3168/jds.2015-10789

Tucker, H. A. 1981. Physiological control of mammary growth, lactogenesis, and lactation(1). J. Dairy Sci. 64:1403-1421. 10.3168/jds.S0022-0302(81)82711-7

van Knegsel, A. T. M., S. G. A. van Der Drift, M. Horneman, A. P. W. de Roos, B. Kemp, and E. A. M. Graat. 2010. Short communication: Ketone body concentration in milk determined by Fourier transform infrared spectroscopy: Value for the detection of hyperketonemia in dairy cows. J. Dairy Sci. 93:3065-3069. 10.3168/jds.2009-2847

VIT. 2020. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed June 23, 2020. https://www.vit.de/fileadmin/DE/Zuchtwertschaetzung/Zws_Bes_eng.pdf

Whitaker, D. A. 2004. Metabolic profiles. Pages 804-817 in Bovine Medicine: Diseases and Husbandry of Cattle. 2nd ed. Blackwell Science, Oxford, UK.

Wishart, D. S. 2008. Metabolomics: applications to food science and nutrition research. Trends Food Sci. Technol. 19:482-493. doi.org/10.1016/j.tifs.2008.03.003

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I: HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009 Jan;37(Database issue):D603-10. doi: 10.1093/nar/gkn810

Chapter 2:

Genomic Prediction of Serum Biomarkers of Health in Early Lactation

J. Dairy Sci. 102:11142–11152; doi:10.3168/jds.2019-17127

T. D. W. Luke, T. T. T. Nguyen, S. Rochfort, W. J. Wales, C. M. Richardson, M. Abdelsayed and J. E. Pryce

Published: December 2019 (open access)

J. Dairy Sci. 102:11142–11152
https://doi.org/10.3168/jds.2019-17127

Genomic prediction of serum biomarkers of health in early lactation

T. D. W. Luke,[1,2] T. T. T. Nguyen,[1] S. Rochfort,[1,2] W. J. Wales,[3] C. M. Richardson,[1,2] M. Abdelsayed,[4] and J. E. Pryce[1,2]*
[1]Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia
[3]Agriculture Victoria Research, Ellinbank Centre, 1301 Hazeldean Rd., Ellinbank, Victoria 3820, Australia
[4]Datagene Ltd., AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia

ABSTRACT

In this study, we estimated genetic parameters and genomic prediction accuracies of serum biomarkers of health in early-lactation dairy cows. A single serum sample was taken from 1,393 cows, located on 14 farms in southeastern Australia, within 30 d after calving. Sera were analyzed for biomarkers of energy balance (β-hydroxybutyrate and fatty acids), macromineral status (Ca and Mg), protein nutritional status (urea and albumin), and immune status (globulins, albumin-to-globulin ratio, and haptoglobin). After editing, 47,162 SNP marker genotypes were used to estimate genomic heritabilities and breeding values (GEBV) for these traits in ASReml. Heritabilities were low for β-hydroxybutyrate, fatty acids, Ca, Mg, and urea (0.09 ± 0.04, 0.18 ± 0.05, 0.07 ± 0.04, 0.19 ± 0.06, and 0.18 ± 0.05, respectively), and moderate for albumin, globulins, and albumin-to-globulin ratio (0.27 ± 0.06, 0.46 ± 0.06, and 0.41 ± 0.06, respectively). The heritability of haptoglobin concentration was close to 0. The magnitude of genetic correlations between traits (estimated using bivariate models) varied considerably (0.01 to 0.96), and standard errors of these correlations were high (0.02 to 0.44). Interestingly, the direction of most genetic correlations was favorable, suggesting that selecting for more optimal concentrations of one biomarker may result in more optimal concentrations of other biomarkers. Correlations between biomarker GEBV and existing breeding values for survival, somatic cell count, and daughter fertility were small to moderate (0.07 to 0.45) and favorable, whereas correlations with breeding values for milk production traits were small (≤0.15). Accuracies of GEBV were evaluated by using 5-fold cross validation, and by calculating accuracies from prediction error variances associated with the GEBV. Accuracies of GEBV predicted using 5-fold cross validation were low (0.05 to 0.27), whereas the means of individual accuracies were greater, ranging from 0.31 to 0.51. Although increasing the size of the reference population should theoretically improve accuracies, our results suggest that genomic prediction of health biomarkers may allow identification of cows that are less susceptible to diseases in early lactation.
Key words: biomarker, health, metabolic stability, immune response, energy balance

INTRODUCTION

Improved animal health and resilience are increasingly important breeding objectives for the dairy industry (Boichard and Brochard, 2012). Most disease events affecting dairy cows occur in the first 30 d after calving (LeBlanc et al., 2006). Many of these diseases are associated with metabolic disorders such as ketosis and hypocalcemia (DeGaris and Lean, 2008; Ospina et al., 2010; McArt et al., 2013), which can have deleterious effects on animal health and welfare and farm profitability (Suthar et al., 2013; McArt et al., 2015). Although heritability estimates of metabolic disorders are generally low (Uribe et al., 1995; Pryce et al., 2016), sufficient genetic variation exists to indicate that improvement in metabolic health can be achieved through genetic selection. Furthermore, several authors have reported favorable genetic correlations between different metabolic disorders (Heringstad et al., 2005; Jamrozik et al., 2016) and between metabolic disorders and diseases such as mastitis and reproductive disorders (Lyons et al., 1991; Oikonomou et al., 2008a). These relationships suggest that selecting for improvements in metabolic health may lead to improvements in overall animal health.

Phenotypes used to investigate the genetic parameters of health traits include producer- or veterinarian-recorded health data (Neuenschwander et al., 2012; Parker Gaddis et al., 2014; Egger-Danner et al., 2015) and biomarkers of health measured in blood or milk

Received June 18, 2019.
Accepted August 21, 2019.
*Corresponding author: jennie.pryce@agriculture.vic.gov.au

(Koeck et al., 2014; Tsiamadis et al., 2016a; Cecchinato et al., 2018). Health data have the advantage of being widely available, but their usefulness is often limited by underreporting or inconsistent recording, and by the fact that trait definitions are often restricted to binary clinical disease events (Østerås et al., 2007). In contrast, biomarker concentrations provide more accurate and objective phenotypes and have continuous distributions that enable the identification of both clinical and subclinical health disorders. Subclinical health disorders are important because of their relatively high prevalence and significant negative effects on animal welfare and performance (Macrae et al., 2006; McArt et al., 2012; Suthar et al., 2013). Commonly used biomarkers include those associated with energy balance (BHB and fatty acids), macromineral status (Ca and Mg), protein nutritional status (urea and albumin), and immune status (globulins and albumin-to-globulin ratio, **A:G**) (Whitaker, 2004; Anderson, 2009). Although extremely valuable, collecting such phenotypes is time consuming and costly, and the collection process is invasive to the animal, making their use impractical in traditional large-scale genetic evaluations, yet good candidates for genomic prediction.

Genomic selection offers exciting potential application for achieving genetic improvement in economically important but difficult-to-measure and lowly heritable health traits, by using data obtained from relatively small genotyped reference populations with high-quality phenotypic data (Boichard and Brochard, 2012; Egger-Danner et al., 2015; Abdelsayed et al., 2017). Examples include genomic selection for residual feed intake (Pryce et al., 2012), tolerance to heat (Nguyen et al., 2016), resistance to bovine tuberculosis (Tsairidou et al., 2014), and enhanced immune response (Thompson-Crispi et al., 2012a).

The aims of this study were to estimate (1) the genetic parameters of serum biomarkers of health in early-lactation dairy cows using data collected from a genotyped female reference population, and (2) the accuracy of genomic predictions of serum biomarker concentrations. If sufficiently accurate, genomic selection for improved metabolic health offers the potential to provide permanent and cumulative improvements in dairy cow health and resilience, thereby increasing animal welfare and farm profitability.

MATERIALS AND METHODS

All procedures undertaken in this study were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (NHMRC, 2013). Approval to proceed was granted by the Agricultural Research and Extension Animal Ethics Committee of the Department of Jobs, Precincts and Resources Animal Ethics Committee (Attwood, Victoria, Australia), and the Tasmanian Department of Primary Industries, Parks, Water and Environment (Animal Biosecurity and Welfare Branch, New Town, Tasmania, Australia).

Phenotypes: Serum Biomarkers

A single serum sample (approximately 4 mL) was taken from 1,393 early-lactation Holstein-Friesian cows from 14 farms in southeastern Australia between August 2017 and October 2018, according to the protocol described in Luke et al. (2019). All cows were between 0 and 30 DIM at the time of sampling. This DIM range was chosen because (1) it is the period in which 75% of disease events affecting dairy cattle occur (LeBlanc et al., 2006), and (2) all farms involved in this study operated a seasonal calving pattern, with large numbers of animals calving in a short period. This made making multiple visits impractical, and the aim of our experimental design was to maximize both the number of animals in the immediate postcalving period and the total number of animals that could be sampled in a single visit.

Sera were analyzed for biomarkers of energy balance (BHB and fatty acids), mineral status (Ca and Mg), protein nutritional status (urea and albumin), and immune status (globulins, A:G, and haptoglobin) by Regional Laboratory Services (Benalla, Victoria, Australia). The concentrations of biomarkers were determined using the following assays: enzymatic kinetic assays for BHB (McMurray et al., 1984) and urea (Wilcox et al., 1966); enzymatic end-point assay for fatty acids (proprietary formulation, Randox Laboratories, Crumlin, UK); arsenazo III for Ca (Janssen and Helbing, 1991); xylidyl blue for Mg (Svoboda and Chromý, 1971); bromocresol green for albumin (Dumas et al., 1997); biuret for total protein (Gornall et al., 1949); and peroxidase activity for haptoglobin (Makimura and Suzuki, 1982). All assays were performed using a Kone 20 XT clinical chemistry analyzer (Thermo Fisher Scientific, Waltham, MA) with reagents supplied by Randox Laboratories for fatty acids, Ca, Mg, and urea, and by Regional Laboratory Services for BHB, albumin, total protein, and haptoglobin (Makimura and Suzuki, 1982). Globulin concentrations were calculated as total protein concentration minus albumin concentration, and A:G as albumin concentration divided by globulin concentration. Descriptive statistics of phenotypes, including the number of samples analyzed and optimal concentrations ranges for each biomarker, are shown in Table 1. Preliminary data analysis was undertaken using R version 3.6.0 (R Core Team, 2019).

Genomic Analysis

Genotypes and Population Structure. Genotypes for the 1,393 animals used in this study were provided by DataGene Ltd. (AgriBio, Bundoora, Victoria, Australia). After editing using the method described in Erbe et al. (2012), 47,162 SNP markers were available for genomic analyses. A genomic relationship matrix (**GRM**) was constructed according to Yang et al. (2010). Principal component analysis of the GRM was performed to examine the population structure of the data set. Plots of the first 5 principal components, which summed to a total of 37% of the variation within the GRM, were visually examined and no subpopulations were identified, confirming that the population was predominantly Holstein.

Genetic Parameters. First, variance components were estimated for each biomarker trait using single linear mixed animal models in ASReml version 4.1 (Gilmour et al., 2015). In matrix notation, the model used was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ is a vector of biomarker concentrations (BHB, fatty acids, Ca, Mg, urea, albumin, globulins, A:G, haptoglobin); $\mathbf{b}$ is a vector of fixed effects of DIM at time of sampling (covariate, from 0 to 30 d, either as a linear, linear + quadratic, third-order orthogonal polynomial, or fourth-order orthogonal polynomial), herd (14 levels, with a range of 10 to 254 cows per herd), parity (4 levels, defined as 1, 2, 3, or 4+), and date of sample collection (class variable with 20 levels); $\mathbf{u}$ is a vector of random genetic effects; $\mathbf{e}$ is a vector of the random residual effects; and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices for $\mathbf{b}$ and $\mathbf{u}$, respectively. It is assumed that $\text{var}(\mathbf{u}) = \mathbf{GRM}\sigma_u^2$, $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$, where σ_u^2 is the genetic variance, σ_e^2 is the residual variance, and $\mathbf{I}$ is an identity matrix.

Estimated variance components were then used to calculate the genomic heritability of each biomarker. The distributions of the residuals of each model were checked for normality using frequency histograms. Residuals of models for BHB, fatty acids, and haptoglobin were positively skewed. To fulfil the assumption of normality for subsequent genetic analyses, a $\log_{10}$ transformation was applied to BHB and haptoglobin concentrations, and a square root transformation was applied to fatty acid concentrations.

Second, a bivariate model was used to estimate the genetic correlation between each pair of serum biomarker traits. For a given pair of biomarkers, the bivariate model was

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & 0\\ 0 & \mathbf{X}_2\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_1 & 0\\ 0 & \mathbf{Z}_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix}, \qquad [2]$$

where $\mathbf{y}_i$ = vector of observations for the ith trait, $\mathbf{b}_i$ = vector of fixed effects for the ith trait, $\mathbf{u}_i$ = vector of random animal effects for the ith trait, $\mathbf{e}_i$ = residual effects for the ith trait, and $\mathbf{X}_i$ and $\mathbf{Z}_i$ = incidence matrices relating records of the ith trait to fixed and random animal effects, respectively. It was assumed that

$$\text{var}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \mathbf{e}_1\\ \mathbf{e}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{g}_{11}\mathbf{GRM} & \boldsymbol{g}_{12}\mathbf{GRM} & 0 & 0\\ \boldsymbol{g}_{21}\mathbf{GRM} & \boldsymbol{g}_{22}\mathbf{GRM} & 0 & 0\\ \mathbf{0} & 0 & \boldsymbol{r}_{11} & \boldsymbol{r}_{12}\\ 0 & 0 & \boldsymbol{r}_{21} & \boldsymbol{r}_{22}\end{bmatrix},$$

where $\boldsymbol{\Sigma g}$ = additive genetic (co)variance matrix for animal effect with each element, defined as $\boldsymbol{g}_{11}$ = additive genetic variance for direct effects for trait 1; $\boldsymbol{g}_{12}$

Table 1. Number of samples (n), phenotypic means, standard deviations, coefficients of variation (CV), and optimal concentration ranges of serum biomarkers of metabolic health of dairy cattle in the first 30 d of lactation

					Optimal concentration of serum metabolites		
Phenotype	n	Mean	SD	CV	Lower threshold	Upper threshold	Reference
BHB	1,393	0.48	0.22	0.46	—	1.2 mmol/L	Compton et al. (2015)
Fatty acids	1,393	0.55	0.33	0.60	—	0.7 mmol/L	Ospina et al. (2010)
Ca	1,327	2.31	0.18	0.08	2.0 mmol/L	—	DeGaris and Lean (2008)
Mg	1,294	0.98	0.14	0.14	0.62 mmol/L	—	Anderson (2009)
Urea	1,393	5.23	0.17	0.03	1.7 mmol/L	—	Macrae et al. (2006)
Albumin	1,294	32.79	2.95	0.09	30 g/L	—	Whitaker (2004)
Globulin	1,294	38.36	6.04	0.16	—	50 g/L	Whitaker (2004)
Albumin:globulin	1,294	0.88	0.17	0.19	0.84	—	Kaneko et al. (2008)
Haptoglobin	779	0.27	0.30	1.11	—	1.4 g/L	Pohl et al. (2015)

$= \boldsymbol{g}_{21}$ = additive genetic covariance between trait 1 and trait 2; $\boldsymbol{g}_{22}$ = additive genetic variance for direct effects for trait 2; **GRM** is the genomic relationship matrix among animals; and **R** = (co)variance matrix for residual effects, where $\boldsymbol{r}_{11}$ = residual variance for trait 1; $\boldsymbol{r}_{12} = \boldsymbol{r}_{21}$ = residual covariance between trait 1 and trait 2; and $\boldsymbol{r}_{22}$ = residual variance for trait 2.

Correlations Between Biomarker Genomic EBV and Breeding Values for Health, Fertility, and Production Traits. Genomic EBV (**GEBV**) for each biomarker trait were predicted using genomic BLUP (**GBLUP**) in ASReml (Gilmour et al., 2015), using variance components estimated from the univariate model (model [1]). Published EBV for health (survival and SCC), daughter fertility, and milk production (milk, fat, and protein yields), calculated using BLUP from pedigree, cow data, and genomics, were obtained from the Australian routine national genetic evaluations performed by DataGene Ltd. (Bundoora, Australia). We attempted to use the method described in Calo et al. (1973) to correct for the reliability of breeding values; however, this method led to unrealistically high correlations. For this reason, simple Pearson correlations between GEBV and EBV were calculated. Any individual breeding value (GEBV or EBV) with a reliability <0.1 was excluded from the analysis.

Genomic Predictions. The accuracy of genomic predictions was assessed in 2 ways. First, empirical prediction accuracy (r_e) was evaluated using 5-fold cross-validation as proposed by Legarra et al. (2008). This involved randomly dividing the total population into 5 equally sized groups or folds. Data from 1 fold (approximately 20% of the population) were set aside as a testing set. Data from the remaining 4 folds (approximately 80% of the population) formed the training set, which was used for model development. The resulting model was then used to predict GEBV for animals in the testing set. This was repeated 5 times, so that all animals appeared in the testing set once. Empirical accuracy was then calculated as the Pearson correlation between the predicted GEBV and actual phenotypic values, corrected for the fixed effects described in model [1]. The corrected phenotypes will include the additive genetic and residual components associated with each phenotype. To calculate predicted accuracies of the true breeding values (r_p) (as opposed to the EBV), the mean correlations between GEBV and corrected phenotypes for each cross-validation fold were divided by the square root of the heritability of the trait (Su et al., 2012).

Second, individual accuracy (r_i) of individual i was calculated as

$$r_i = \sqrt{1 - \frac{SE_i^2}{\sigma_g^2 \mathbf{GRM}_{ii}}}$$

where SE_i is the standard error of GEBV of individual i, and σ_g^2 is the genetic variance of each trait estimated from model [1], adjusted for inbreeding by multiplying by the corresponding diagonal elements in the **GRM** for each individual ($\mathbf{GRM}_{ii}$).

RESULTS

Model Selection and Genetic Parameters

We tested 4 functions of DIM as fixed effects in model [1]: linear, linear + quadratic, third-order orthogonal polynomial, or fourth-order orthogonal polynomial. The Akaike information criteria resulting from these models suggested that fatty acids was best fitted with linear DIM, whereas the rest of the traits were best fitted with the third-order orthogonal polynomial function of DIM.

Estimated genomic heritabilities obtained from model [1] are shown in Table 2. Heritability estimates for serum BHB, fatty acids, Ca, Mg, and urea concentrations were low, at 0.09, 0.18, 0.07, 0.19, and 0.18, respectively. Heritabilities of albumin, globulins, and A:G were higher at 0.27, 0.46, and 0.41, respectively. The estimated heritability of haptoglobin in our data set was close to zero. Standard errors for all heritability estimates were low (0.04 to 0.06).

Genetic and phenotypic correlations between biomarkers, estimated from model [2], are shown in Table 2. The magnitude of estimated genetic correlations varied considerably, ranging from close to 0 (BHB and globulin, and Ca and globulins) to −0.96 (globulins and A:G). With the exception of correlations between albumin and globulin and A:G, standard errors of all correlations were relatively high (0.12 to 0.44). We observed significant positive genetic correlations between Ca and albumin (0.54 ± 0.31), Mg and urea (0.44 ± 0.22), Mg and albumin (0.29 ± 0.17), and urea and albumin (0.79 ± 0.16). Significant negative genetic correlations were observed between fatty acids and Ca (−0.82 ± 0.44), fatty acids and albumin (−0.29 ± 0.18), and albumin and globulins (−0.50 ± 0.12). The trend in the direction of genetic correlations was generally favorable; toward lower concentrations of BHB, fatty acids, and globulins, and higher concentrations of Ca, Mg, urea, albumin, and A:G. This was true for all genetic correlations, except for those between BHB and Mg, urea and albumin, and fatty acids and globulins.

Table 2. Genetic parameter estimates of serum biomarkers of metabolic health in dairy cattle in the first 30 d of lactation, including heritability (diagonal), genetic (above diagonal), and phenotypic (below diagonal) correlations ± standard errors

Item	BHB_{Log10}[1]	Fatty $acids_{SQRT}$[2]	Ca	Mg	Urea	Albumin	Globulin	Albumin:globulin
BHB_{Log10}[1]	**0.09 ± 0.04**	0.24 ± 0.26	−0.06 ± 0.42	0.38 ± 0.29	0.21 ± 0.28	0.11 ± 0.24	0.01 ± 0.22	−0.07 ± 0.23
Fatty $acids_{SQRT}$[2]	0.20 ± 0.03	**0.18 ± 0.05**	−0.82 ± 0.44	−0.20 ± 0.21	−0.17 ± 0.21	−0.29 ± 0.18	−0.03 ± 0.16	−0.05 ± 0.16
Ca	−0.09 ± 0.03	−0.05 ± 0.03	**0.07 ± 0.04**	0.21 ± 0.33	0.48 ± 0.31	0.54 ± 0.22	−0.01 ± 0.25	0.12 ± 0.25
Mg	−0.02 ± 0.03	−0.01 ± 0.03	0.08 ± 0.03	**0.19 ± 0.06**	0.44 ± 0.22	0.29 ± 0.17	−0.21 ± 0.16	0.25 ± 0.16
Urea	0.16 ± 0.03	−0.06 ± 0.03	0.07 ± 0.03	0.06 ± 0.03	**0.18 ± 0.05**	0.79 ± 0.16	−0.16 ± 0.16	0.38 ± 0.16
Albumin	0.08 ± 0.03	0.10 ± 0.03	0.44 ± 0.02	0.34 ± 0.03	0.25 ± 0.03	**0.27 ± 0.06**	−0.50 ± 0.12	0.70 ± 0.08
Globulin	−0.14 ± 0.03	−0.05 ± 0.03	0.03 ± 0.03	−0.06 ± 0.03	−0.13 ± 0.03	−0.31 ± 0.03	**0.46 ± 0.06**	−0.96 ± 0.02
Albumin:globulin	0.12 ± 0.03	0.07 ± 0.03	0.15 ± 0.03	0.18 ± 0.03	0.19 ± 0.03	0.63 ± 0.02	−0.87 ± 0.01	**0.41 ± 0.06**

[1]Log_{10}-transformed BHB concentration.
[2]Square root-transformed fatty acid concentration.

The magnitude of phenotypic correlations varied from 0.01 (between fatty acids and Mg) to 0.44 (between Ca and albumin). Standard errors for all phenotypic correlations were small (≤0.03). The direction of phenotypic correlations was the same as the direction of genetic correlations for all trait pairs except BHB and Mg, BHB and A:G, fatty acids and albumin, fatty acids and A:G, and globulins and Ca. Of these, the only statistically significant genetic correlation was between fatty acids and albumin.

Correlations Between Biomarker GEBV and Breeding Values for Health, Fertility, and Production Traits

Pearson correlations between biomarker GEBV and EBV for survival, SCC, daughter fertility, and milk production traits are shown in Table 3. The magnitudes of correlations between biomarker GEBV and health and fertility EBV were low to moderate: 0.15 to 0.45 for survival, 0.07 to 0.32 for SCC, and 0.11 to 0.37 for daughter fertility. The direction of correlations was favorable for all pairs of breeding values; BHB, fatty acids, and globulins GEBV were negatively correlated with health and fertility EBV, whereas Ca, Mg, urea, albumin, and A:G GEBV were positively correlated with health and fertility EBV.

The magnitude of correlations between biomarker GEBV and EBV for production traits (milk, fat, and protein yields) were small (0.01 to 0.15). We observed small positive correlations between breeding values for milk yield and albumin (0.11), milk yield and urea (0.10), milk fat yield and BHB (0.15), and milk fat yield and urea (0.10). Small negative correlations were observed between the breeding values for milk yield and BHB (−0.10), milk yield and Mg (−0.06), milk protein yield and Mg (−0.13), and milk fat yield and Ca (−0.10).

Accuracy of Genomic Predictions

Empirical accuracies and means of individual accuracies of genomic predictions resulting from univariate models are shown in Table 4. Empirical accuracies for all models were low, ranging from 0.05 to 0.27, and increased with increasing trait heritability. Expected empirical accuracies of true breeding values, estimated by correcting empirical accuracies for trait heritabilities, were between 0.20 and 0.40. Mean individual accuracies were greater than empirical accuracies for all traits (0.31 to 0.51), but the results of the 2 methods were in agreement (i.e., the correlation between the 2 sets of accuracies was 0.89).

Table 3. Pearson correlations between genomic EBV for serum biomarkers of early lactation health, and Australian breeding values for survival, SCC, daughter fertility, and milk, fat, and protein yields

		EBV					
Biomarker	N	Survival	SCC	Daughter fertility	Milk yield	Milk protein yield	Milk fat yield
BHB_{Log10}[1]	848	−0.15	−0.08	−0.11	−0.10	0.06	0.15
Fatty $acids_{SQRT}$[2]	1,176	−0.27	−0.16	−0.20	−0.07	0.01	0.04
Ca	719	0.15	0.07	0.24	0.00	0.07	−0.10
Mg	1,129	0.21	0.15	0.25	−0.06	−0.13	0.03
Albumin	1,228	0.45	0.32	0.37	0.11	0.03	0.07
Globulin	1,313	−0.25	−0.19	−0.20	−0.02	0.06	0.03
Urea	1,161	0.38	0.23	0.26	0.10	0.05	0.10
Albumin:globulin	1,321	0.36	0.26	0.30	0.07	−0.01	0.01

[1]Log_{10}-transformed BHB concentration.

[2]Square root-transformed fatty acid concentration.

DISCUSSION

The genetic parameters of metabolic disorders in early-lactation dairy cows, as defined by producer or veterinarian-recorded health data, have been studied extensively. Few studies have investigated the genetic parameters of serum biomarkers of health, and to the best of the authors' knowledge, this is the first study to investigate and report the genetic parameters of a metabolic profile that covers a range of biomarkers of energy balance, macromineral status, protein nutritional status, and immune status. Furthermore, we believe this is the first study to report the accuracies of genomic predictions of these traits and marks the start of an emerging area for genomic prediction to reduce early-lactation disease in dairy cows.

Heritability Estimates of Serum Biomarkers of Health

Our results indicate that genetic variation exists for all biomarkers studied except haptoglobin. Heritability estimates were consistent with the literature for Mg (Tsiamadis et al., 2016a), albumin, globulins, and A:G (Cecchinato et al., 2018). The estimated heritability of fatty acid concentration in our study was consistent with the findings of Oikonomou et al. (2008b); however, it should be noted that our heritability estimate is for square root-transformed fatty acid concentrations, not raw concentrations. No reports of the heritability of serum urea concentration were found in the literature. However, our results are consistent with the reported heritability of MUN concentration (Mitchell et al., 2005), which is linearly correlated with serum urea concentration (Moore and Varga, 1996).

The genetic parameters of BHB have been more widely reported than those of the other biomarkers investigated in this study. Reported heritabilities of BHB concentration vary considerably, from 0.073 ± 0.77 (Tsiamadis et al., 2016b) to 0.40 ± 0.06 (Oikonomou et al., 2008b). However, care must be taken when interpreting results because of significant differences in study design (in particular, stage of lactation), math-

Table 4. Accuracies of genomic EBV for serum metabolic biomarkers including empirical accuracies for each of 5 cross-validation folds, mean empirical accuracy (μ), predicted accuracy of the true breeding value (μ/h), and the mean of individual accuracies calculated from predicted error variance (r_t)

	Cross-validation fold							
Trait	1	2	3	4	5	μ	μ/h	r_t
BHB_{Log10}[1]	0.04	0.08	0.14	0.09	0.08	0.09	0.29	0.34
Fatty $acids_{SQRT}$[2]	0.15	0.11	0.14	0.16	0.19	0.15	0.36	0.41
Calcium	0.02	0.09	0.12	−0.09	0.13	0.05	0.20	0.31
Magnesium	0.09	0.01	0.13	0.17	0.22	0.12	0.28	0.41
Urea	0.24	0.13	0.18	0.02	0.06	0.13	0.30	0.41
Albumin	0.18	0.26	0.25	0.15	0.14	0.20	0.38	0.44
Globulin	0.24	0.28	0.30	0.29	0.23	0.27	0.40	0.51
Albumin:globulin	0.24	0.26	0.30	0.28	0.19	0.25	0.40	0.49

[1]Log_{10}-transformed BHB concentration.

[2]Square root-transformed fatty acid concentration.

ematical transformations of metabolite concentrations, and the genetic models used. Our study and results are most comparable to those of Weigel et al. (2017) and van der Drift et al. (2012), who reported heritabilities of 0.093 ± 0.045 for square root-transformed BHB, and 0.17 ± 0.06 for $\log_{10}$-transformed BHB, respectively. Oikonomou et al. (2008b) demonstrated that the heritability of serum BHB concentration is greatest in the week immediately after calving and decreases rapidly over the first 7 wk of lactation. In our study, only 209 cows were in the first week of lactation at the time of sampling, and we expect that increasing the number of animals sampled during the suggested high-risk period will improve heritabilities.

The heritability of Ca in our data set was significantly less than that reported by Tsiamadis et al. (2016a), who found that the heritability of serum Ca at d 1, 2, 4, and 8 postpartum ranged from 0.23 ± 0.02 to 0.32 ± 0.03. In adult cows, homeostatic mechanisms maintain serum Ca concentrations between 2.1 and 2.5 mmol/L (Goff, 2008). However, serum Ca concentrations decline in the periparturient period and reach their nadir 12 to 24 h postcalving before rapidly returning to normal physiological levels once homeostatic mechanisms are restored (Kimura et al., 2006). It is likely that our low heritability estimate is the result of having sampled only 14 cows during this period of greatest phenotypic variability. Subclinical hypocalcemia, defined as serum Ca concentrations between 1.38 and 2.1 mmol/L and which increases the risk of other metabolic and infectious diseases, occurs most commonly in this 12-h window (Goff, 2008). We therefore plan to collect many more samples from animals during this period in future investigations.

Our study also differed from others that report higher heritabilities, in that we took only a single sample from each animal. It is likely that taking serial samples from individual animals across the early-lactation period and using random regression models would increase heritability estimates. This was not possible in the current study because all farms operated a seasonal calving system with large numbers of cows calving in a short period, making multiple visits impractical. These results demonstrate the importance of careful trait definition when investigating genetic parameters of health traits in the transition period. We plan to collect more samples during the periods of highest phenotypic and genetic variation in the future (e.g., 0 to 1 DIM for Ca, 0 to 7 DIM for BHB).

The heritability of haptoglobin in our data set was close to 0. Haptoglobin is a positive acute phase protein produced by the liver (Morimatsu et al., 1991), which is used as an indicator of inflammation in cattle (Horadagoda et al., 1999). The low genetic variance of haptoglobin concentration in our research was surprising, especially given the relatively high heritability of albumin, another acute phase protein produced by the liver (Jain et al., 2011), and of globulins and A:G, other indicators of inflammation (Burke et al., 2010). One possible reason for this result is the small sample size (n = 779), and we consider that more data are required to validate this finding. An alternative to haptoglobin may be A:G, which has been used as a nonspecific indicator of immune status (Piccinini et al., 2004). Our results indicate that A:G may be a promising biomarker for genetic selection of improved immune competence, similar to selecting for humoral immune response, which has been demonstrated to improve overall animal health and resilience (De La Paz, 2008; Thompson-Crispi et al., 2012b).

Genetic Correlations Between Serum Biomarkers of Health

The trend in the direction of genetic correlations between biomarkers was mostly favorable, suggesting that selection for more optimal concentrations of one biomarker may result in improvements in the concentrations of others. Optimal concentration ranges for health biomarkers, based on epidemiological associations between biomarker concentrations and health, production, and fertility outcomes, have been studied and reported extensively (Whitaker, 2004; Overton et al., 2017) and are summarized in Table 1. Our results are consistent with previous studies that reported favorable genetic correlations between the occurrence of clinical metabolic diseases such as milk fever and ketosis (Heringstad et al., 2005). The most favorable correlations were those of fatty acids with albumin and Ca, and albumin with Ca, Mg, urea, and globulins. These results suggest that albumin and fatty acids may be potential biomarkers for early lactation health. Albumin is an abundant protein synthesized in the liver that is responsible for several important biological functions, such as maintaining circulating blood volume; transporting metabolites, hormones, fatty acids, and other nutrients; and controlling biologically active concentrations of Ca in the bloodstream (Majorek et al., 2012). Given these diverse biological functions, it follows that higher concentrations of albumin are genetically associated with improved animal health and fertility. Alternatively, it may be that a higher albumin concentration is an indicator of good liver function, which is known to be important for early lactation health (Bionaz et al., 2007). Fatty acid concentration is an indicator of negative energy balance, and elevated (>0.57 mmol/L)

concentrations in the first 2 wk of lactation are associated with an increased risk of subsequent negative health events (Ospina et al., 2010; Chapinal et al., 2012; Sordillo and Raphael, 2013). Selecting for animals that mobilize less body fat may therefore lead to a decrease in other early-lactation diseases.

We could find relatively few reports in the scientific literature of genetic correlations between metabolic profile biomarkers in serum. Our results did not support reports of unfavorable correlations between serum BHB and albumin and A:G (Cecchinato et al., 2018). This may be because data in that study were not restricted to early lactation. Although the magnitude of correlations in our data set ranged considerably, standard error estimates for many correlations were relatively large, and many correlations were not significantly different from zero. This is consistent with the findings of Tsiamadis et al. (2016a), who found that genetic correlations between Ca, P, Mg, and K concentrations in the first 8 d of lactation were not significantly different from zero. The large standard errors we report are likely due, in part, to the small sample size, and we consider that more data are required to better understand these genetic relationships. Despite this, the trends in our data support previous findings that some early-lactation metabolic disorders share some common genetic basis, and that there may be potential to select for increased resilience in the form of improved metabolic stability in early lactation.

In addition to larger data sets, alternative approaches may be required to better understand the complex genetic relationships between metabolic diseases in early lactation. One approach is that taken by Ha et al. (2015), who used gene-based mapping and pathway analysis to demonstrate that BHB, fatty acid, and glucose concentrations share genetic pathways associated with steroid and lipid metabolism. Another approach could be to use metabolomic techniques, such as liquid chromatography-mass spectroscopy and nuclear magnetic resonance spectroscopy, in conjunction with genome-wide association studies, to identify novel biomarkers and genes associated with early-lactation health or disease.

Correlations Between Biomarker GEBV and Health, Fertility, and Milk Production Breeding Values

Correlations between biomarker GEBV and health (survival and SCC) and fertility EBV were low to moderate, offering further evidence that complex genetic relationships exist between the etiopathology of metabolic, inflammatory, and fertility disorders in early lactation. Interestingly, all correlations were favorable, with more-optimal biomarker concentrations being associated with higher breeding values for survival, SCC, and daughter fertility.

Correlations between breeding values should be viewed as indicative of genetic correlations, and it should be noted that phenotypic data from some of the cows in our study may also have been used in the DataGene evaluation. However, it was encouraging that our results were consistent with the findings of Oikonomou et al. (2008a), who reported favorable genetic correlations between fatty acid concentrations and fertility traits such as calving interval and absence of reproductive problems, and those of Koeck et al. (2014), who reported favorable correlations between milk BHB and SCC, fertility, and herd life. Cecchinato et al. (2018) reported only nonsignificant correlations between SCC and BHB (-0.081 ± 0.49) and SCC and globulins (-0.108 ± 0.54), and nonsignificant correlations between SCC and albumin (0.423 ± 0.52) and SCC and A:G (0.350 ± 0.51). Again, this could be attributed to differences in study design and lactation period. Of particular interest were the moderate correlations we observed between health and fertility EBV, and GEBV for fatty acids, albumin, globulins, and A:G. These results are consistent with favorable correlations among these biomarkers. Multi-trait genome-wide association studies may help improve our understanding of these complex genetic relationships.

In contrast to correlations with health and fertility traits, correlations between biomarker GEBV and EBV for production traits were low (≤ 0.15), suggesting that selection for improved metabolic health in early lactation may not have a large effect on milk production. In contrast, Cecchinato et al. (2018) reported significant genetic correlations between BHB and milk yield (0.716 ± 0.21), total protein and fat percentage (-0.854 ± 0.18), and albumin and fat percentage (-0.894 ± 0.17), although it should be reiterated that this study was not restricted to biomarker concentrations in early lactation. Belay et al. (2017) found positive genetic correlations between BHB predicted from milk mid-infrared spectra, and milk yield (0.277 ± 0.016), milk protein (0.107 ± 0.017), and milk fat (0.248 ± 0.016). These differing results are consistent with the variability in reported genetic correlations between producer- or veterinarian-recorded metabolic diseases traits and milk production traits (Pryce et al., 2016). Koeck et al. (2014) reported a moderate genetic correlation between milk BHB concentration and producer-recorder clinical ketosis occurrence (0.48 ± 0.35). Such favorable genetic correlations between biomarker concentrations and clinical disease data should enable the assembly of much larger data sets, which are required to better understand these relationships.

Accuracy of Genomic Predictions of Serum Biomarkers

The accuracies of genomic predictions observed are commensurate with a small female reference population and low-to-moderate trait heritabilities (Gonzalez-Recio et al., 2014). Weigel et al. (2017) reported greater empirical genomic prediction accuracies for BHB concentrations (0.29 to 0.36); however, those results are not comparable to results of the current study because their predicted phenotypes included model solutions for the herd-year-season contemporary group and parity, in addition to the GEBV; that is, their prediction was of the expressed phenotype. We expect that increasing the size of the reference population and refining trait definitions to maximize heritabilities should improve accuracies. Given the cost and logistical challenges of blood sampling large numbers of cows, one method to dramatically increase the number of phenotypes may be to use mid-infrared spectroscopy of milk to predict serum biomarker concentrations (Belay et al., 2017). Additionally, this may be a trait area that is suitable for sharing of data sets. A comparable example is genomic prediction of DMI using multi-country populations; for example, de Haas et al. (2015). They assembled a data set of DMI records from ~9,000 cows in 10 populations to show that genomic prediction accuracies were always higher when multi-country reference sets were used compared with within-country data sets. Finally, high-throughput metabolomic methods such as nuclear magnetic resonance spectroscopy and liquid chromatography-mass spectroscopy may help to identify novel and more accurate biomarkers of health and disease; if these data can be used to facilitate identification of important genetic variants, this may also help to improve genomic prediction accuracies.

CONCLUSIONS

We investigated genetic parameters of serum biomarkers of health using data collected from a genotyped female reference population. We found that biomarkers of energy balance (BHB and fatty acids), protein nutritional status (albumin and urea), macromineral status (Ca and Mg), and immune status (globulins and A:G) are heritable traits, and that genomic selection to improve the concentrations of these biomarkers should be possible. Of the biomarkers investigated, fatty acids, albumin, and A:G are of particular interest because of their (1) significant phenotypic associations with other early-lactation diseases, (2) moderate to high heritabilities, (3) promising genomic prediction accuracies, (4) favorable estimated genetic correlations with other biomarkers, and (5) promising correlations between GEBV and existing EBV for health, fertility, and production traits.

ACKNOWLEDGMENTS

We thank DairyBio, jointly funded by Dairy Australia (Melbourne, Australia) and Agriculture Victoria (Melbourne, Australia), for funding this project and T. Luke's PhD project. The authors also thank Di Mapleson, Brigid Ribaux, and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance; Erika Oakes, Miranda Clark, and the staff at Datagene (Bundoora, Australia) for their work coordinating this study; and the farmers who took part in this project.

REFERENCES

Abdelsayed, M., M. Haile-Mariam, and J. E. Pryce. 2017. Genetic parameters for health traits using data collected from genomic information nucleus herds. J. Dairy Sci. 100:9643–9655. https://doi.org/10.3168/jds.2017-12960.

Anderson, D. E. 2009. Current Veterinary Therapy: Food Animal Practice. Saunders/Elsevier, St. Louis, MO.

Belay, T. K., M. Svendsen, Z. M. Kowalski, and T. Ådnøy. 2017. 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. 100:6298–6311. https://doi.org/10.3168/jds.2016-12458.

Bionaz, M., E. Trevisi, L. Calamari, F. Librandi, A. Ferrari, and G. Bertoni. 2007. Plasma paraoxonase, health, inflammatory conditions, and liver function in transition dairy cows. J. Dairy Sci. 90:1740–1750. https://doi.org/10.3168/jds.2006-445.

Boichard, D., and M. Brochard. 2012. New phenotypes for new breeding goals in dairy cattle. Animal 6:544–550. https://doi.org/10.1017/S1751731112000018.

Burke, C. R., S. Meier, S. McDougall, C. Compton, M. Mitchell, and J. R. Roche. 2010. Relationships between endometritis and metabolic state during the transition period in pasture-grazed dairy cows. J. Dairy Sci. 93:5363–5373. https://doi.org/10.3168/jds.2010-3356.

Calo, L. L., R. E. McDowell, L. Dale Van Vleck, and P. D. Miller. 1973. Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. J. Anim. Sci. 37:676–682. https://doi.org/10.2527/jas1973.373676x.

Cecchinato, A., T. Bobbo, P. L. Ruegg, L. Gallo, G. Bittante, and S. Pegolo. 2018. Genetic variation in serum protein pattern and blood β-hydroxybutyrate and their relationships with udder health traits, protein profile, and cheese-making properties in Holstein cows. J. Dairy Sci. 101:11108–11119. https://doi.org/10.3168/jds.2018-14907.

Chapinal, N., S. J. Leblanc, M. E. Carson, K. E. Leslie, S. Godden, M. Capel, J. E. Santos, M. W. Overton, and T. F. Duffield. 2012. Herd-level association of serum metabolites in the transition period with disease, milk production, and early lactation reproductive performance. J. Dairy Sci. 95:5676–5682. https://doi.org/10.3168/jds.2011-5132.

Compton, C. W., L. Young, and S. McDougall. 2015. Subclinical ketosis in post-partum dairy cows fed a predominantly pasture-based diet: Defining cut-points for diagnosis using concentrations of beta-hydroxybutyrate in blood and determining prevalence. N. Z. Vet. J. 63:241–248. https://doi.org/10.1080/00480169.2014.999841.

de Haas, Y., J. E. Pryce, M. P. L. Calus, E. Wall, D. P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsegge, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North

America, and Australasia. J. Dairy Sci. 98:6522–6534. https://doi.org/10.3168/jds.2014-9257.
De La Paz, J. M. 2008. Using humoral and cellular response to novel antigens in periparturient dairy cows as a measure of genetic disease resistance in dairy cows. MSc Thesis. Veterinary Medicine Dept., University of Florida, Gainesville.
DeGaris, P. J., and I. J. Lean. 2008. Milk fever in dairy cows: A review of pathophysiology and control principles. Vet. J. 176:58–69. https://doi.org/10.1016/j.tvjl.2007.12.029.
Dumas, B. T., W. A. Watson, and H. G. Biggs. 1997. Albumin standards and the measurement of serum albumin with bromocresol green. Clin. Chim. Acta 258:21–30. https://doi.org/10.1016/s0009-8981(96)06447-9.
Egger-Danner, C., J. B. Cole, J. E. Pryce, N. Gengler, B. Heringstad, A. Bradley, and K. F. Stock. 2015. Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. Animal 9:191–207. https://doi.org/10.1017/S1751731114002614.
Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114–4129. https://doi.org/10.3168/jds.2011-5019.
Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2015. ASReml User Guide. Release 4.1 Structural Specification. VSN International Ltd., Hemel Hempstead, UK. www.vsni.co.uk.
Goff, J. P. 2008. The monitoring, prevention, and treatment of milk fever and subclinical hypocalcemia in dairy cows. Vet. J. 176:50–57. https://doi.org/10.1016/j.tvjl.2007.12.020.
Gonzalez-Recio, O., M. P. Coffey, and J. E. Pryce. 2014. On the value of the phenotypes in the genomic era. J. Dairy Sci. 97:7905–7915. https://doi.org/10.3168/jds.2014-8125.
Gornall, A. G., C. J. Bardawill, and M. M. David. 1949. Determination of serum proteins by means of the biuret reaction. J. Biol. Chem. 177:751–766.
Ha, N.-T., J. J. Gross, A. van Dorland, J. Tetens, G. Thaller, M. Schlather, R. Bruckmaier, and H. Simianer. 2015. Gene-based mapping and pathway analysis of metabolic traits in dairy cows. PLoS One 10:e0122325. https://doi.org/10.1371/journal.pone.0122325.
Heringstad, B., Y. M. Chang, D. Gianola, and G. Klemetsdal. 2005. Genetic analysis of clinical mastitis, milk fever, ketosis, and retained placenta in three lactations of Norwegian red cows. J. Dairy Sci. 88:3273–3281. https://doi.org/10.3168/jds.S0022-0302(05)73010-1.
Horadagoda, N. U., K. M. Knox, H. A. Gibbs, S. W. Reid, A. Horadagoda, S. E. Edwards, and P. D. Eckersall. 1999. Acute phase proteins in cattle: Discrimination between acute and chronic inflammation. Vet. Rec. 144:437–441. https://doi.org/10.1136/vr.144.16.437.
Jain, S., V. Gautam, and S. Naseem. 2011. Acute-phase proteins: As diagnostic tool. J. Pharm. Bioallied Sci. 3:118–127. https://doi.org/10.4103/0975-7406.76489.
Jamrozik, J., A. Koeck, G. J. Kistemaker, and F. Miglior. 2016. Multiple-trait estimates of genetic parameters for metabolic disease traits, fertility disorders, and their predictors in Canadian Holsteins. J. Dairy Sci. 99:1990–1998. https://doi.org/10.3168/jds.2015-10505.
Janssen, J. W., and A. R. Helbing. 1991. Arsenazo III: An improvement of the routine calcium determination in serum. Eur. J. Clin. Chem. Clin. Biochem. 29:197–201.
Kaneko, J. J., J. W. Harvey, and M. L. Bruss. 2008. Normal concentrations of blood constituents in domestic animals. Pages 873–904 in Clinical Biochemistry of Domestic Animals. 6th ed. Academic Press/Elsevier, Burlington, MA.
Kimura, K., T. A. Reinhardt, and J. P. Goff. 2006. Parturition and hypocalcemia blunts calcium signals in immune cells of dairy cattle. J. Dairy Sci. 89:2588–2595. https://doi.org/10.3168/jds.S0022-0302(06)72335-9.
Koeck, A., J. Jamrozik, F. S. Schenkel, R. K. Moore, D. M. Lefebvre, D. F. Kelton, and F. Miglior. 2014. Genetic analysis of milk β-hydroxybutyrate and its association with fat-to-protein ratio, body condition score, clinical ketosis, and displaced abomasum in early first lactation of Canadian Holsteins. J. Dairy Sci. 97:7286–7292. https://doi.org/10.3168/jds.2014-8405.
LeBlanc, S. J., K. D. Lissemore, D. F. Kelton, T. F. Duffield, and K. E. Leslie. 2006. Major advances in disease prevention in dairy cattle. J. Dairy Sci. 89:1267–1279. https://doi.org/10.3168/jds.S0022-0302(06)72195-6.
Legarra, A., C. Robert-Granie, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618. https://doi.org/10.1534/genetics.108.088575.
Luke, T. D. W., S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett, and J. E. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.
Lyons, D. T., A. E. Freeman, and A. L. Kuck. 1991. Genetics of health traits in Holstein cattle. J. Dairy Sci. 74:1092–1100. https://doi.org/10.3168/jds.S0022-0302(91)78260-X.
Macrae, A. I., D. A. Whitaker, E. Burrough, A. Dowell, and J. M. Kelly. 2006. Use of metabolic profiles for the assessment of dietary adequacy in UK dairy herds. Vet. Rec. 159:655–661. https://doi.org/10.1136/vr.159.20.655.
Majorek, K. A., P. J. Porebski, A. Dayal, M. D. Zimmerman, K. Jablonska, A. J. Stewart, M. Chruszcz, and W. Minor. 2012. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. Mol. Immunol. 52:174–182. https://doi.org/10.1016/j.molimm.2012.05.011.
Makimura, S., and N. Suzuki. 1982. Quantitative determination of bovine serum haptoglobin and its elevation in some inflammatory diseases. Nihon Juigaku Zasshi 44:15–21.
McArt, J. A. A., D. V. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. J. Dairy Sci. 95:5056–5066. https://doi.org/10.3168/jds.2012-5443.
McArt, J. A. A., D. V. Nydam, G. R. Oetzel, T. R. Overton, and P. A. Ospina. 2013. Elevated non-esterified fatty acids and β-hydroxybutyrate and their association with transition dairy cow performance. Vet. J. 198:560–570. https://doi.org/10.1016/j.tvjl.2013.08.011.
McArt, J. A. A., D. V. Nydam, and M. W. Overton. 2015. Hyperketonemia in early lactation dairy cattle: A deterministic estimate of component and total cost per case. J. Dairy Sci. 98:2043–2054. https://doi.org/10.3168/jds.2014-8740.
McMurray, C. H., W. J. Blanchflower, and D. A. Rice. 1984. Automated kinetic method for D-3-hydroxybutyrate in plasma or serum. Clin. Chem. 30:421–425.
Mitchell, R. G., G. W. Rogers, C. D. Dechow, J. E. Vallimont, J. B. Cooper, U. Sander-Nielsen, and J. S. Clay. 2005. Milk urea nitrogen concentration: Heritability and genetic correlations with reproductive performance and disease. J. Dairy Sci. 88:4434–4440. https://doi.org/10.3168/jds.S0022-0302(05)73130-1.
Moore, D. A., and G. Varga. 1996. BUN and MUN: Urea nitrogen testing in dairy cattle. Compend. Contin. Educ. Vet. 18:712–720.
Morimatsu, M., B. Syuto, N. Shimada, T. Fujinaga, S. Yamamoto, M. Saito, and M. Naiki. 1991. Isolation and characterization of bovine haptoglobin from acute phase sera. J. Biol. Chem. 266:11833–11837.
NHMRC (National Health and Medical Research Council). 2013. Australian code for the care and use of animals for scientific purposes. 8th ed. NHMRC Publications, Canberra, Australia.
Neuenschwander, T. F., F. Miglior, J. Jamrozik, O. Berke, D. F. Kelton, and L. R. Schaeffer. 2012. Genetic parameters for producer-recorded health data in Canadian Holstein cattle. Animal 6:571–578. https://doi.org/10.1017/S1751731111002059.
Nguyen, T. T. T., P. J. Bowman, M. Haile-Mariam, J. E. Pryce, and B. J. Hayes. 2016. Genomic selection for tolerance to heat stress in Australian dairy cattle. J. Dairy Sci. 99:2849–2862. https://doi.org/10.3168/jds.2015-9685.
Oikonomou, G., G. Arsenos, G. E. Valergakis, A. Tsiaras, D. Zygoyiannis, and G. Banos. 2008a. Genetic relationship of body energy

and blood metabolites with reproduction in Holstein cows. J. Dairy Sci. 91:4323–4332. https://doi.org/10.3168/jds.2008-1018.
Oikonomou, G., G. E. Valergakis, G. Arsenos, N. Roubies, and G. Banos. 2008b. Genetic profile of body energy and blood metabolic traits across lactation in primiparous Holstein cows. J. Dairy Sci. 91:2814–2822. https://doi.org/10.3168/jds.2007-0965.
Ospina, P. A., D. Nydam, T. Stokol, and T. Overton. 2010. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. J. Dairy Sci. 93:1596–1603. https://doi.org/10.3168/jds.2009-2852.
Østerås, O., H. Solbu, A. O. Refsdal, T. Roalkvam, O. Filseth, and A. Minsaas. 2007. Results and evaluation of thirty years of health recordings in the Norwegian dairy cattle population. J. Dairy Sci. 90:4483–4497. https://doi.org/10.3168/jds.2007-0030.
Overton, T. R., J. A. A. McArt, and D. V. Nydam. 2017. A 100-Year Review: Metabolic health indicators and management of dairy cattle. J. Dairy Sci. 100:10398–10417. https://doi.org/10.3168/jds.2017-13054.
Parker Gaddis, K. L., J. B. Cole, J. S. Clay, and C. Maltecca. 2014. Genomic selection for producer-recorded health event data in US dairy cattle. J. Dairy Sci. 97:3190–3199. https://doi.org/10.3168/jds.2013-7543.
Piccinini, R., E. Binda, M. Belotti, G. Casirani, and A. Zecconi. 2004. The evaluation of non-specific immune status of heifers in field conditions during the periparturient period. Vet. Res. 35:539–550. https://doi.org/10.1051/vetres:2004030.
Pohl, A., O. Burfeind, and W. Heuwieser. 2015. The associations between postpartum serum haptoglobin concentration and metabolic status, calving difficulties, retained fetal membranes, and metritis. J. Dairy Sci. 98:4544–4551. https://doi.org/10.3168/jds.2014-9181.
Pryce, J. E., J. Arias, P. J. Bowman, S. R. Davis, K. A. Macdonald, G. C. Waghorn, W. J. Wales, Y. J. Williams, R. J. Spelman, and B. J. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95:2108–2119. https://doi.org/10.3168/jds.2011-4628.
Pryce, J. E., K. L. Parker Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B. Heringstad, C. Egger-Danner, K. F. Stock, A. J. Bradley, and J. B. Cole. 2016. Invited review: Opportunities for genetic improvement of metabolic diseases. J. Dairy Sci. 99:6855–6873. https://doi.org/10.3168/jds.2016-10854.
R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Sordillo, L. M., and W. Raphael. 2013. Significance of metabolic stress, lipid mobilization, and inflammation on transition cow disorders. Vet. Clin. North Am. Food Anim. Pract. 29:267–278. https://doi.org/10.1016/j.cvfa.2013.03.002.
Su, G., O. F. Christensen, T. Ostersen, M. Henryon, and M. S. Lund. 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7:e45293. https://doi.org/10.1371/journal.pone.0045293.
Suthar, V. S., J. Canelas-Raposo, A. Deniz, and W. Heuwieser. 2013. Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. J. Dairy Sci. 96:2925–2938. https://doi.org/10.3168/jds.2012-6035.
Svoboda, V., and V. Chromý. 1971. A spectrophotometric study of the magnesium-xylidyl blue II complex: Optimal conditions for the determination of magnesium. Anal. Chim. Acta 54:121–131. https://doi.org/10.1016/S0003-2670(01)81862-6.
Thompson-Crispi, K., R. Ventura, F. Schenkel, F. Miglior, and B. Mallard. 2012a. Genomic selection for enhanced immune response to improve dairy cattle health. Genome 55:721.
Thompson-Crispi, K. A., A. Sewalem, F. Miglior, and B. A. Mallard. 2012b. Genetic parameters of adaptive immune response traits in Canadian Holsteins. J. Dairy Sci. 95:401–409. https://doi.org/10.3168/jds.2011-4452.
Tsairidou, S., J. A. Woolliams, A. R. Allen, R. A. Skuce, S. H. McBride, D. M. Wright, M. L. Bermingham, R. Pong-Wong, O. Matika, S. W. J. McDowell, E. J. Glass, and S. C. Bishop. 2014. Genomic prediction for tuberculosis resistance in dairy cattle. PLoS One 9:e96728. https://doi.org/10.1371/journal.pone.0096728.
Tsiamadis, V., G. Banos, N. Panousis, M. Kritsepi-Konstantinou, G. Arsenos, and G. E. Valergakis. 2016a. Genetic parameters of calcium, phosphorus, magnesium, and potassium serum concentrations during the first 8 days after calving in Holstein cows. J. Dairy Sci. 99:5535–5544. https://doi.org/10.3168/jds.2015-10787.
Tsiamadis, V., G. Banos, N. Panousis, M. Kritsepi-Konstantinou, G. Arsenos, and G. E. Valergakis. 2016b. Genetic parameters of subclinical macromineral disorders and major clinical diseases in postparturient Holstein cows. J. Dairy Sci. 99:8901–8914. https://doi.org/10.3168/jds.2015-10789.
Uribe, H. A., B. W. Kennedy, S. W. Martin, and D. F. Kelton. 1995. Genetic parameters for common health disorders of Holstein cows. J. Dairy Sci. 78:421–430. https://doi.org/10.3168/jds.S0022-0302(95)76651-6.
van der Drift, S. G., K. J. van Hulzen, T. G. Teweldemedhn, R. Jorritsma, M. Nielen, and H. C. Heuven. 2012. Genetic and nongenetic variation in plasma and milk β-hydroxybutyrate and milk acetone concentrations of early-lactation dairy cows. J. Dairy Sci. 95:6781–6787. https://doi.org/10.3168/jds.2012-5640.
Weigel, K. A., R. S. Pralle, H. Adams, K. Cho, C. Do, and H. M. White. 2017. Prediction of whole-genome risk for selection and management of hyperketonemia in Holstein dairy cattle. J. Anim. Breed. Genet. 134:275–285. https://doi.org/10.1111/jbg.12259.
Whitaker, D. A. 2004. Metabolic profiles. Pages 804–817 in Bovine Medicine: Diseases and Husbandry of Cattle. 2nd ed. Blackwell Science, Oxford, UK.
Wilcox, A. A., W. E. Carroll, R. E. Sterling, H. A. Davis, and A. G. Ware. 1966. Use of the Berthelot reaction in the the automated analysis of serum urea nitrogen. Clin. Chem. 12:151–157.
Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569. https://doi.org/10.1038/ng.608.

ORCIDS

T. D. W. Luke https://orcid.org/0000-0001-5070-2332
T. T. T. Nguyen https://orcid.org/0000-0002-9880-6859
S. Rochfort https://orcid.org/0000-0001-8442-6081
J. E. Pryce https://orcid.org/0000-0002-1397-1282

Chapter 3:

Metabolic Profiling of Early Lactation Dairy Cows Using Milk Mid-Infrared Spectra

J. Dairy Sci. 2019. 102:1-14; doi:10.3168/jds.2018-15103

T. D. W. Luke, S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett and J. E. Pryce

Published: February 2019 (open access)

J. Dairy Sci. 102:1747–1760
https://doi.org/10.3168/jds.2018-15103

Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra

T. D. W. Luke,[1,2] S. Rochfort,[1,2] W. J. Wales,[3] V. Bonfatti,[4] L. Marett,[3] and J. E. Pryce[1,2]*
[1]Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia
[2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia
[3]Department of Economic Development, Jobs, Transport and Resources, Ellinbank Centre, 1301 Hazeldean Rd., Ellinbank, Victoria, 3820 Australia
[4]Department of Comparative Biomedicine and Food Science, University of Padova, viale dell' Università 16 35020, Legnaro, PD, Italy

ABSTRACT

Metabolic disorders in early lactation have negative effects on dairy cow health and farm profitability. One method for monitoring the metabolic status of cows is metabolic profiling, which uses associations between the concentrations of several metabolites in serum and the presence of metabolic disorders. In this cross-sectional study, we investigated the use of mid-infrared (MIR) spectroscopy of milk for predicting the concentrations of these metabolites in serum. Between July and October 2017, serum samples were taken from 773 early-lactation Holstein Friesian cows located on 4 farms in the Gippsland region of southeastern Victoria, Australia, on the same day as milk recording. The concentrations in sera of β-hydroxybutyrate (BHB), fatty acids, urea, Ca, Mg, albumin, and globulins were measured by a commercial diagnostic laboratory. Optimal concentration ranges for each of the 7 metabolites were obtained from the literature. Animals were classified as being either affected or unaffected with metabolic disturbances based on these ranges. Milk samples were analyzed by MIR spectroscopy. The relationships between serum metabolite concentrations and MIR spectra were investigated using partial least squares regression. Partial least squares discriminant analyses (PLS-DA) were used to classify animals as being affected or not affected with metabolic disorders. Calibration equations were constructed using data from a randomly selected subset of cows (n = 579). Data from the remaining cows (n = 194) were used for validation. The coefficient of determination (R^2) of serum BHB, fatty acids, and urea predictions were 0.48, 0.61, and 0.90, respectively. Predictions of Ca, Mg, albumin, and globulin concentrations were poor ($0.06 \leq R^2 \leq 0.17$). The PLS-DA models could predict elevated fatty acid and urea concentrations with an accuracy of approximately 77 and 94%, respectively. A second independent validation data set was assembled in March 2018, comprising blood and milk samples taken from 105 autumn-calving cows of various breeds. The accuracies of BHB and fatty acid predictions were similar to those obtained using the first validation data set. The PLS-DA results were difficult to interpret due to the low prevalence of metabolic disorders in the data set. Our results demonstrate that MIR spectroscopy of milk shows promise for predicting the concentration of BHB, fatty acids, and urea in serum; however, more data are needed to improve prediction accuracies.
Key words: mid-infrared spectral prediction, metabolic profile, ketosis, energy balance

INTRODUCTION

Metabolic disorders in early lactation have significant negative effects on dairy cow health and welfare as well as farm profitability (Suthar et al., 2013; McArt et al., 2015). The most commonly described metabolic disorders are ketosis, hypocalcemia, and hypomagnesemia. Subclinical metabolic disorders, which are not associated with obvious clinical signs, are of particular interest due to their relatively high prevalence and significant effects on animal welfare and performance (Macrae et al., 2006; McArt et al., 2012; Suthar et al., 2013). Identification of subclinical disorders can also allow for timely management interventions to prevent the development of clinical disease.

One way of monitoring the metabolic health and nutritional status of dairy cows is serum metabolic profile testing, which employs well-established epidemiological associations between the concentrations of several metabolites in serum and the presence of both subclinical and clinical metabolic disorders (Payne et al., 1970; Ospina et al., 2010a). The metabolites evaluated in metabolic profile testing vary, but often include BHB and fatty acids as indicators of energy balance, albumin and BUN as indicators of protein status, globulins as an

Received May 23, 2018.
Accepted October 31, 2018.
*Corresponding author: jennie.pryce@ecodev.vic.gov.au

indicator of chronic inflammatory disease, and Ca and Mg as indicators of macromineral status (Whitaker, 2004; Anderson, 2009). Epidemiological studies have demonstrated that elevated concentrations of BHB and fatty acids in early lactation are associated with an increased risk of subsequent negative health events and reduced production (Ospina et al., 2010b; Chapinal et al., 2012; Sordillo and Raphael, 2013). Blood urea nitrogen concentration is of increasing interest, as it (1) gives an indication of RDP intake and the ratio of RDP to energy in the ration, and (2) has been demonstrated to be a useful indicator of an animal's nitrogen utilization efficiency and excretion (Kohn et al., 2005; Kume et al., 2008).

Critical concentration thresholds are used to define the optimum concentration range for of each metabolite employed in a metabolic profile test. Concentrations falling outside this range are associated with detrimental downstream health and production outcomes (Ospina et al., 2010b). Herd prevalence thresholds are similarly defined as the proportion of animals with metabolite concentrations outside the optimum range, above which detrimental herd-level health and production outcomes are seen (Ospina et al., 2010b; Chapinal et al., 2012). The aim of metabolic profiling is therefore not necessarily to identify individual sick animals, but to gain objective information on the nutritional status and metabolic health of a herd by estimating the prevalence of metabolic disorders.

Despite the advantages of metabolic profile testing, blood testing animals on a regular basis is invasive, logistically challenging, and costly. Given the ready availability of milk, its use as a biofluid to monitor the health and nutritional status of dairy cows has been widely investigated (Hamann and Krömker, 1997). In early lactation a milk fat-to-protein ratio of greater than 1.4 (Schcolnik, 2016) and 2.0 (Toni et al., 2011) have been described as indicators of negative energy balance and subclinical ketosis, respectively, and changes in milk fat-to-lactose and milk fat-to-protein ratios in early lactation have been suggested as early indicators of disease (Paudyal et al., 2016). Milk urea nitrogen is routinely used by nutritionists to monitor and optimize protein nutrition (Jonker et al., 2002; Nousiainen et al., 2004). More recently, mid-infrared (**MIR**) spectroscopy of milk has shown promise for assessing more complex animal health traits (Gengler et al., 2016). Several authors have demonstrated that MIR spectral data can be used to screen for subclinical ketosis through identification of ketone bodies in milk (de Roos et al., 2007; van Knegsel et al., 2010; Grelet et al., 2016) and to estimate energy balance in early lactation (McParland et al., 2011). Attempts have also been made to estimate the concentration of serum biomarkers of energy balance using milk MIR spectra (Gelé et al., 2015; Belay et al., 2017a; Pralle et al., 2018).

The aim of our study was to determine if MIR spectral data, obtained from routine milk recording in commercial dairy herds, could be used to predict the concentration of metabolites routinely employed in serum metabolic profiling, with sufficient accuracy to provide useful information on the metabolic health of early-lactation dairy cows. We also aimed to assess the robustness of MIR prediction equations by validating our results with data collected from a herd managed under a different production system and in a different season. If sufficiently accurate, milk MIR predictions of serum biomarkers may help to improve the health, welfare, and productivity dairy cattle by (1) allowing early identification of metabolic disease and (2) providing high throughput and cost-effective phenotypes for genetic evaluation of complex animal health traits.

MATERIALS AND METHODS

All procedures were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (National Health and Medical Research Council, 2013). Approval to proceed was obtained from the Agricultural Research and Extension Animal Ethics Committee (Department of Economic Development, Jobs, Transport and Resources, Attwood, Victoria, Australia).

Sample Collection

Data Set 1. A single blood sample was taken from 773 spring-calving Holstein-Friesian cows in early lactation (between 5 and 49 DIM) on the same day as milk recording, between July and October 2017. The cows were located on 4 farms (farms A, B, C, and D) in the Gippsland region of southeastern Australia. All 4 farms operated a seasonal calving system, with the majority of cows calving in a short period of time to align the peak nutritional demands of the herd with maximal pasture availability. The farms implemented a feeding system reliant on grazed pasture plus other forages, with more than 1 t of a cereal grain per cow per year fed in the parlor at milking time. Two of the farms (farms C and D) operated rotary milking platforms, which allowed blood samples to be collected during milking. Samples were taken immediately after milking on the other 2 farms. Samples were collected after the morning milking on farm A, after the afternoon milking on farm B, and during the afternoon milking on farms C and D.

Blood was collected from the coccygeal vein into 10-mL serum clot activator vacutainer tubes (Becton

Dickinson, Franklin Lakes, NJ). Samples were allowed to clot for a minimum of 1 h at room temperature before centrifugation at 1,200 × *g* for 10 min at 18°C. All samples were processed within 6 h of collection. Serum samples were refrigerated at 4°C then transported on ice to Regional Laboratory Services (Benalla, Victoria, Australia) within 24 h of collection. Samples were analyzed for concentrations of BHB, fatty acids, BUN, total Ca, Mg, total protein, and albumin using a Kone 20 XT clinical chemistry analyzer (Thermo Fisher Scientific, Waltham, MA), with reagents supplied by Randox Laboratories (Crumlin, UK) for fatty acids, BUN, Ca, and Mg, and Regional Laboratory Services (Benalla, Victoria, Australia) for BHB, albumin, and total protein. Globulin concentrations were calculated as total protein concentration minus albumin concentration. Milk samples were collected as part of routine milk recording by the Herd Improvement Co-Operative Australia (Maffra, Victoria, Australia). Samples were preserved with SomaGlo (proprietary formulation, Bentley Instruments, Chaska, MN) and analyzed fresh using MIR spectroscopy (Bentley Instruments NexGen FTS Combi) by TasHerd Pty Ltd. (Hadspen, Tasmania, Australia).

Farm E Independent Validation Data Set. To test the robustness of MIR-prediction equations, a second independent validation data set was assembled in March 2018. This data set comprised a further 105 blood and milk samples taken from cows of different breeds, managed under a different production system, and calving in a different season (autumn as opposed to spring). The farm was located in the Gippsland region of southeastern Australia, and the herd consisted of Jersey, Australian Red, Holstein-Friesian, and crossbred cows. Cows were fed a diet consisting of grazed chicory, a ration of pasture silage, cottonseed and canola meal, and a wheat-barley grain mix fed in the parlor at milking time. Blood samples were collected immediately after the afternoon milking. Blood and milk samples were analyzed using the same protocols described for data set 1.

Statistical Analysis

Effect of Week of Lactation, Parity and Farm on Metabolite Concentrations. Fixed effects models were constructed to evaluate the effect of weeks in milk, parity, and farm, on the concentrations of each metabolite:

$$y_{ijkl} = \mu + WIM_i + P_j + F_k + e_{ijkl}, \quad [1]$$

where y is the metabolite concentration (BHB, fatty acids, Ca, Mg, urea, albumin, and globulin), μ is the mean, *WIM* is weeks in milk (from 1 to 8), P is parity (primiparous vs. multiparous), F is the effect of farm, and e is the random error term. Phenotypic correlations between metabolite concentrations were investigated by calculating the Pearson correlations between the residuals of each model.

Optimum metabolite concentration ranges were defined based on thresholds obtained from the literature and are shown in Table 1. Each metabolite concentration for every animal was classified as being either within or outside the defined optimum range, thus converting each continuous metabolite concentration variable into a binary trait. The prevalence of each metabolic disorder was then calculated as the percentage of animals that had a metabolite concentration outside the optimum range.

MIR Predictions. All MIR spectral data analysis was performed with Matlab R2017a (MathWorks, Natick, MA) utilizing the PLS Toolbox (Eigenvector Research, Manson, WA).

Preprocessing of Metabolite Concentrations. The distributions of serum metabolite concentrations were visually assessed for normality using frequency histograms. The fatty acid and BHB concentration distributions were both skewed, with lower values over-represented; this type of distribution leads to decreased accuracy in predicting high values in partial least squares (**PLS**) regression (Grelet et al., 2016), so a logarithmic (10) transformation was applied to BHB

Table 1. Upper and lower concentration thresholds for serum metabolites used for metabolic profile analyses

		Optimum concentration of serum metabolites	
Metabolite	Reference	Lower threshold	Upper threshold
BHB (mmol/L)	McArt et al., 2012; Compton et al., 2015	—	1.2
Fatty acids (mmol/L)	Ospina et al., 2010a	—	0.7
Ca (mmol/L)	DeGaris and Lean, 2008	2.0	—
Mg (mmol/L)	Anderson, 2009	0.62	—
Urea (mmol/L)	Butler et al., 1996; Macrae et al., 2006	1.7	6.78
Albumin (g/L)	Whitaker, 2004	30	—
Globulin (g/L)	Whitaker, 2004	—	50

concentrations and a square root transformation was applied to fatty acid concentrations (Figure 1).

Preprocessing of Spectra. The MIR spectra were expressed in absorbance, with 899 spectral points between 649 and 3,998 cm^{-1}. Preliminary analysis of the spectral data was conducted using principal component analysis. No outliers were identified in the data set. Spectral regions associated with the O–H bending and stretching regions of water were excluded (Afseth et al., 2010; Belay et al., 2017a). This left 538 spectral wavelengths between 928 and 1,596 as well as 1,693 and 3,025 cm^{-1} for the subsequent chemometric analysis. The MIR spectra were preprocessed with Savitzky–Golay second derivative transformation and smoothing, removal, of linear trend and autoscaling (Eigenvector, 2018).

Calibration and Validation. The relationships between blood metabolite concentrations and milk MIR spectra were investigated using PLS regression analysis. Partial least squares discriminant analysis (**PLS-DA**) was used to classify animals as being either affected or not affected with a metabolic disorder based on the aforementioned binary metabolic profile classifications.

Calibration equations were constructed using a randomly allocated subset of data set 1, which consisted of serum metabolic profile results and MIR spectral data from 579 animals (hereafter referred to as the calibration data set). These calibration equations were used for all subsequent analyses. The data from the remaining 194 animals from data set 1 were used for external validation (hereafter referred to as the random validation data set). The calibration and random validation data sets were designed to have a representative number of samples from each farm and parity category (primiparous or multiparous) and were balanced for DIM.

The number of latent variables (**LV**) included in each calibration model was based on maximizing the percentage of variance captured while minimizing the root mean square error of cross-validation ($\mathbf{RMSE_{CV}}$). The optimum number of LV was determined for each calibration model by examining a plot of $RMSE_{CV}$ as a function of number of LV.

Each calibration model was assessed for over-fitting using a permutation test with 50 iterations. Permutation testing of regression models involved randomly reordering the y block, and nominally assigning an incorrect y value to each vector of x values (Eigenvector, 2018). For example, with our data this involved randomly assigning an incorrect serum metabolite concentration

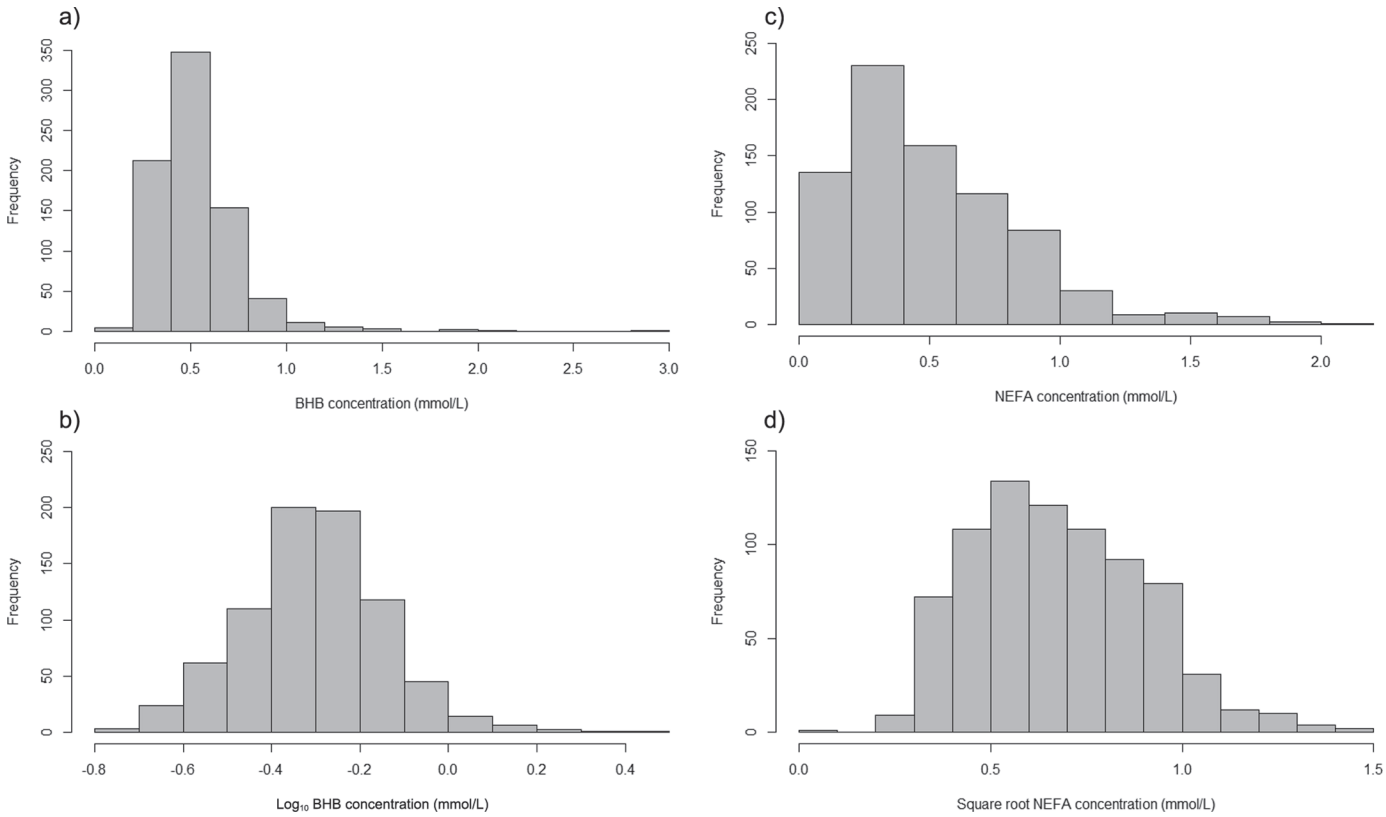


Figure 1. (a) Frequency distribution of untransformed serum BHB concentrations, (b) serum BHB concentrations following Log_{10} transformation, (c) untransformed serum fatty acids, and (d) serum fatty acid concentrations following square root transformation.

Table 2. Number of cows with metabolic profiles and milk mid-infrared spectral data by farm and by data set, including stage of lactation (DIM means and ranges) and percentage of animals in their first lactation

Data	No. of cows	% Primiparous	DIM		
			Mean	Minimum	Maximum
Data set 1					
Farm A	315	27	18.3	5	49
Farm B	132	27	20.6	5	39
Farm C	147	12	21.2	5	39
Farm D	179	18	29.4	6	52
Calibration and random validation data sets					
Calibration	579	22	21.6	5	49
Random validation	194	23	22.4	5	48
Total					
Data set 1 subtotal	773	22	21.8	5	49
Independent validation data set					
Farm E	105	33	29.5	11	46
All data					
Total	878	23	22.7	5	49

(the y value) to an MIR absorbance spectrum (a vector within the x block). The model was then rerun using the original parameters, but with the randomly aligned data. This process was repeated 50 times and the results obtained using the randomly assorted data were compared with the results of the correctly aligned data. A Wilcoxon signed-rank test was then used to assess the probability that the original model was significantly different from those built using the randomly assorted data (Eigenvector, 2018). A P-value of less than 0.05 indicated that the original model was significantly different to the random models and was therefore unlikely to be over-fitted.

Validation was performed in 3 ways. (1) Cross-validation was performed on the calibration data set (n = 579) using a venetian blinds method (which splits the data into 20 subsets and performs cross-validation on 2 samples per subset). (2) External validation was carried out using the random external validation data set (n = 194) and (3) external validation was done using the farm E independent validation data set (n = 105).

The accuracy of PLS models was assessed using the coefficient of determination ($\mathbf{R^2}$) and the root mean square error (**RMSE**). The accuracy of PLS-DA models was assessed by calculating the sensitivity, specificity, classification error (**CE**), and the area under the receiver operator curve (**AUC**).

RESULTS

Descriptive Statistics

Details of the animals included in the analysis are summarized in Table 2. Of the 878 animals included in the analysis, 36% (315 cows) were from farm A. The remaining 563 animals were evenly distributed between the remaining 4 farms. Of the animals sampled, 78% (682 cows) were in the first 30 d of lactation, which is the period of highest risk for development of metabolic disorders (LeBlanc et al., 2006). The overall percentage of primiparous animals in the data set was 23%, with a range of approximately 12 to 33% between farms.

The identity of the farm had a significant effect ($P < 0.05$) on the concentration of all metabolites. The number of weeks after a cow had calved had a significant effect on BHB, fatty acid, BUN, magnesium, and globulin concentrations. Parity had a significant effect on the concentration of all metabolites except those of fatty acids and albumin.

Descriptive statistics for the concentrations of each metabolite measured are summarized in Table 3. The distribution of metabolite concentrations in the calibration and random validation data sets were very similar; however, we found considerable differences in the distributions of fatty acid and urea concentrations between the calibration data set and the farm E independent validation data set.

Corrected mean metabolite concentrations for each 7-d period are shown in Figure 2. Both BHB and fatty acid concentrations were highest immediately postcalving and decreased over time. The concentrations of the remaining 5 metabolites exhibited an increasing trend over the 7-wk period. Calcium, urea, and albumin concentrations peaked at wk 7 postcalving, and globulin concentrations peaked at wk 5 postcalving. Magnesium concentrations peaked at wk 3 postcalving, then plateaued.

The number and percentages of animals with metabolite concentrations outside optimal ranges are shown in Table 4. A total of 56% (489 cows) had 1 or more metabolites outside optimal ranges. Aberrant protein concentrations were the most prevalent

disorder observed, with 39% (339 cows) having either elevated serum urea (31%) or globulin (3%) concentrations or low albumin (8%) concentrations. Less than 2% of animals had serum urea concentration below the optimal range (<1.7 mmol/L), and this disorder is not discussed beyond this point. A total of 23% of animals (205 cows) had 1 or more energy metabolites outside of optimal ranges. Less than 2% (15 cows) had BHB concentrations greater than 1.2 mmol/L, with a peak incidence of 20% (2/10) at 35 DIM. A total of 22% (199 cows) had fatty acids concentrations greater than 0.7 mmol/L, with a peak incidence of 67% (18/27) at 8 DIM. Of the 15 hyperketonemic cows, 6 did not have a concurrent elevation in fatty acids concentrations. The prevalence of hypocalcemia and hypomagnesemia were less than 2 (15 cows) and 1% (6 cows), respectively.

Phenotypic correlations between serum metabolite concentrations, corrected for fixed effects outlined in model 1, are shown in Table 5. Significant ($P < 0.01$) positive correlations were observed between BHB and fatty acids (0.32), Ca and albumin (0.39), Mg and albumin (0.34), urea and Mg (0.1), and urea and albumin (0.26). Significant negative correlations were noted between BHB and Ca (−0.10), fatty acids and Ca (−0.22), fatty acids and urea (−0.12), fatty acids and globulins (−0.09), Ca and globulins (−0.11), Mg and globulins (−0.21), urea and globulins (−0.16), and albumin and globulins (−0.41).

MIR Calibration and Validation

The R^2 and RMSE of PLS regression models investigating the relationships between blood metabolite concentrations and MIR spectra from milk samples are shown in Table 6. The R^2 of cross-validation $\left(\mathbf{R^2_{CV}}\right)$ and random validation $\left(\mathbf{R^2_{RV}}\right)$ for serum BHB predictions were 0.53 and 0.48, respectively. Predictions of serum fatty acids concentration were slightly more accurate, with an R^2_{CV} of 0.56 and an R^2_{RV} of 0.61. The $RMSE_{CV}$ and RMSE of random validation ($\mathbf{RMSE_{RV}}$) of BHB and fatty acids predictions were 0.11 and 0.12, and 0.15 and 0.14, respectively. The most promising results were for predictions of serum urea concentration, which had R^2_{CV} and R^2_{RV} of 0.90, $RMSE_{CV}$ of 0.75, and $RMSE_{RV}$ of 0.82. The accuracy of models predicting serum Ca, Mg, and globulin concentrations were poor, with R^2_{CV} and R^2_{RV} values less than 0.15. The model predicting serum albumin concentration performed slightly better, with R^2_{CV} of 0.23 and R^2_{RV} of 0.17.

The accuracies of prediction models when applied to the farm E independent validation data set (reported as $\mathbf{R^2_{IV}}$ and $\mathbf{RMSE_{IV}}$) are also reported in Table 6. The R^2_{IV} of BHB and fatty acids predictions were similar to the R^2_{CV}, at 0.60 and 0.45, respectively. The $RMSE_{IV}$ of BHB and fatty acids predictions were 0.11 and 0.14, respectively, both very close to the respective $RMSE_{RV}$ and $RMSE_{CV}$ values. The R^2_{IV} for prediction of serum urea concentration however was only 0.35, which was considerably lower than the R^2_{RV} (0.90). The $RMSE_{IV}$ of urea prediction was 1.53 mmol/L, almost double the $RMSE_{RV}$ (0.82 mmol/L). The models predicting serum albumin, globulin, Ca, and Mg concentrations all performed extremely poorly when applied to the independent validation data set, with R^2_{IV} values between 0.00 and 0.03.

Table 3. Mean and SD (in parentheses) of metabolite concentrations for each farm, the calibration data set, and the random validation and independent farm E validation data sets

		Metabolite						
Data	N[1]	BHB	Fatty acids	Ca	Mg	Urea	Albumin	Globulin
Data set 1								
Farm A	315	0.56 (0.22)	0.75 (0.33)	2.33 (0.15)	0.98 (0.11)	5.72 (1.37)	33.59 (2.23)	39.14 (5.28)
Farm B	132	0.49 (0.19)	0.31 (0.21)	2.32 (0.12)	0.96 (0.12)	4.94 (1.52)	35.13 (2.05)	37.07 (6.22)
Farm C	147	0.37 (0.16)	0.51 (0.29)	2.37 (0.16)	0.99 (0.10)	2.77 (0.74)	32.40 (2.13)	40.49 (5.63)
Farm D	179	0.66 (0.24)	0.26 (0.14)	2.33 (0.13)	1.02 (0.10)	9.00 (1.23)	32.69 (2.41)	40.00 (6.44)
Randomly assigned calibration and validation data sets								
Calibration	579	0.53 (0.21)	0.51 (0.34)	2.33 (0.14)	0.99 (0.11)	5.80 (2.40)	33.37 (2.42)	39.33 (6.03)
Validation	194	0.54 (0.29)	0.53 (0.34)	2.34 (0.14)	0.98 (0.11)	5.75 (2.50)	33.55 (2.34)	38.98 (5.48)
Total								
Data set 1 subtotal	773	0.53 (0.23)	0.51 (0.34)	2.34 (0.14)	0.98 (0.11)	5.79 (2.43)	33.42 (2.40)	39.24 (5.89)
Independent validation data set								
Farm E	105	0.56 (0.26)	0.29 (0.26)	2.43 (0.13)	1.10 (0.10)	3.75 (0.78)	34.60 (2.30)	37.18 (4.87)
All data								
Total	878	0.54 (0.23)	0.49 (0.34)	2.35 (0.14)	1.00 (0.12)	5.54 (2.38)	33.56 (2.42)	39.00 (5.81)

[1]Number of cows in the data set.

The results of PLS-DA models, where affected and unaffected groups were defined using previously described metabolic profile thresholds, are shown in Table 7. Models for the prediction of elevated BHB and globulin concentrations, as well as low Ca, Mg, and albumin concentrations, were deemed to be overfitted and therefore not significant ($P > 0.05$) based on pairwise Wilcoxon signed rank permutation testing. Models for the prediction of elevated fatty acids and urea concentrations, however, were highly significant ($P < 0.001$). The sensitivity and specificity for the prediction of elevated fatty acids concentrations when applied to the random validation data set were 73 and 81%, respectively, and the CE was 23% and the AUC was 0.87. The MIR predicted prevalence of elevated serum fatty acids concentrations was 35%. The sensitivity and specificity of the prediction of elevated urea concentrations in the random validation data set were 90 and 98%, respectively, and the CE was 6% and the AUC was 0.98. The MIR-predicted prevalence

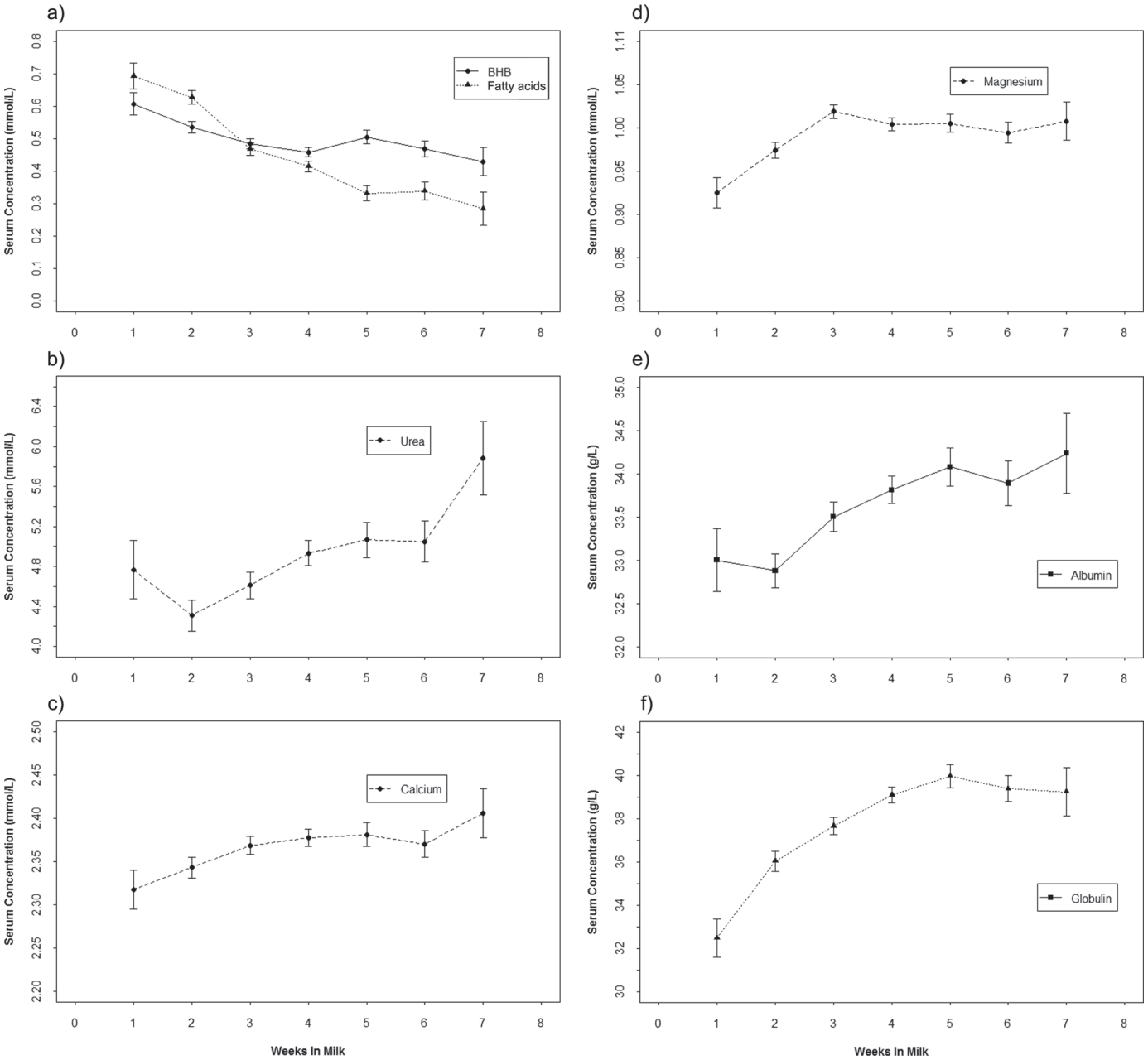


Figure 2. Estimated marginal means (±SEM) of serum metabolite concentrations for each 7-d period, corrected for farm identification and parity.

Table 4. The number and percentage (in parentheses) of animals with serum metabolite concentrations outside optimum ranges for each farm, the calibration data set, and the random validation and independent farm E validation data sets

		Metabolite						
Data	N[1]	BHB	Fatty acids	Ca	Mg	Urea	Albumin	Globulin
Data set 1								
Farm A	315	8 (3)	162 (51)	4 (1)	3 (1)	78 (25)	24 (8)	9 (3)
Farm B	132	1 (1)	5 (4)	4 (3)	2 (2)	18 (14)	2 (2)	4 (3)
Farm C	147	1 (1)	27 (18)	1 (1)	1 (1)	0	22 (15)	7 (5)
Farm D	179	3 (2)	1 (1)	4 (2)	0	175 (98)	21 (12)	8 (4)
Randomly assorted calibration and validation data sets								
Calibration	579	8 (1)	139 (24)	11 (2)	3 (1)	202 (35)	54 (9)	24 (4)
Validation	194	5 (3)	56 (29)	2 (1)	3 (2)	69 (36)	15 (8)	4 (2)
Data set 1 subtotal	773	13 (2)	195 (25)	13 (2)	6 (1)	271 (36)	69 (9)	28 (8)
Independent external validation data set								
Farm E	105	2 (2)	4 (4)	2 (2)	0	0	4 (4)	1 (1)
All data								
Total	878	15 (2)	199 (23)	15 (2)	6 (1)	271 (31)	73 (8)	29 (3)

[1]Number of cows in the data set.

of elevated serum urea concentrations in the random validation data set was 33%.

The accuracy of PLS-DA models, when validated using the farm E independent data set, are also shown in Table 7. The sensitivity and specificity for the prediction of elevated fatty acid concentrations were 25 and 90%, respectively, and the CE was 42% and the AUC was 0.82. The MIR-predicted prevalence of elevated fatty acids in this data set was 11%. The sensitivity and specificity of independent validation for the prediction of elevated urea concentrations were 100 and 89%, respectively, and the CE was 6%. The AUC could not be calculated, as no positive results (serum urea concentration >6.8 mmol/L) were recorded. The predicted prevalence of elevated urea concentrations was 11%.

DISCUSSION

Serum concentrations of BHB, fatty acids, Ca, Mg, urea, albumin, and globulin, as measured by colorimetric methods, are routinely used to assess the metabolic health of dairy cows. Although some studies have investigated the use of MIR spectroscopy of milk to predict serum BHB and fatty acids concentrations, to the best of our knowledge this is the first study to investigate the use of MIR spectral data to predict serum concentrations of all the above metabolites. We believe this is also the first reported use of PLS-DA models to classify animals as being either affected or not affected with metabolic disorders directly from MIR spectra. Our results indicate that MIR spectral data may be a useful predictor of serum BHB, fatty acids, and urea concentrations, but not Ca, Mg, albumin, or globulin concentrations. The performance of both PLS and PLS-DA models were affected by the distribution of the calibration and validation data sets, and larger and more diverse data sets are required to improve the accuracy of predictions.

Prevalence of Metabolic Disorders

No recent studies have investigated the epidemiology of metabolic disorders in the Australian dairy herd; thus, all concentration thresholds used in ou study are based on work undertaken in New Zealand, Europe, and the United States. It should be noted that our study was not intended as an epidemiological investigation, and prevalence data are presented principally to

Table 5. Pearson correlations between serum metabolite concentrations, corrected for weeks in milk, farm identification, and parity

Item	Fatty acids	Ca	Mg	Urea	Albumin	Globulin
BHB	0.32*	−0.10*	−0.03	0.07	−0.02	−0.07
Fatty acids		−0.22*	−0.06	−0.12*	0.01	−0.09*
Ca			0.08	0.08	0.39*	−0.11*
Mg				0.1*	0.34*	−0.21*
Urea					0.26*	−0.16*
Albumin						−0.41*

*$P < 0.01$.

Table 6. Results of partial least square regression models for the prediction of serum metabolite concentrations using milk mid-infrared spectra

Metabolite	*P*-value[1]	No. of LV[2]	Cross-validation[3] (n = 579)		Random validation[4] (n = 194)		Farm E validation[5] (n = 105)	
			R^2_{CV}	$RMSE_{CV}$	R^2_{RV}	$RMSE_{RV}$	R^2_{IV}	$RMSE_{IV}$
BHB	<0.001	6	0.53	0.11	0.48	0.12	0.60	0.11
Fatty acids	<0.001	8	0.56	0.15	0.61	0.14	0.45	0.14
Urea	<0.001	20	0.90	0.75	0.90	0.82	0.35	1.53
Calcium	<0.001	4	0.08	0.14	0.12	0.13	0.03	0.18
Magnesium	<0.001	5	0.06	0.11	0.08	0.10	0.01	0.17
Albumin	<0.001	5	0.23	2.14	0.17	2.18	0.02	2.40
Globulin	<0.001	3	0.12	5.67	0.06	5.42	0.00	5.57

[1]*P*-value for pairwise Wilcoxon signed rank test.

[2]Number of latent variables (LV) included in the model.

[3]R^2_{CV} = coefficient of determination of cross-validation; $RMSE_{CV}$ = root mean square error of cross-validation.

[4]R^2_{RV} = coefficient of determination of random external validation; $RMSE_{RV}$ = root mean square error of random external validation.

[5]R^2_{IV} = coefficient of determination of independent validation; $RMSE_{IV}$ = root mean square error of independent validation.

illustrate the data used to develop and validate PLS-DA models.

The prevalence of hyperketonemia in our data set was 2% (15 cows), which is considerably lower than values reported in studies undertaken in New Zealand, Europe, and North America (McArt et al., 2012; Compton et al., 2014; Suthar et al., 2013). The low prevalence may have been because only approximately 4% of animals (34 cows) in our data set were in the first week of lactation, which McArt et al. (2012), demonstrated to be the period of highest hyperketonemia incidence. This was an unavoidable consequence of our study design, which involved convenience sampling on the day of routine milk recording in commercial herds. Furthermore, farmers with seasonal calving herds are often reluctant to record milk in early lactation, as it is generally a busy time of year. The timing of sampling, during or immediately after concentrate feeding, may also have affected our results, as BHB concentrations are known to vary over time, when access to feed is not constant, and to peak 4 to 5 h after feeding (Oetzel, 2004). All 5 farms were well-managed and implemented good transition cow management practices, which are known to minimize the incidence of ketosis. The prevalence of elevated fatty acid concentrations was 23% (199 cows), with a peak incidence of 67% (18/27) at 6 d after calving. This was consistent with the results of Ospina et al. (2010b), who found that 65% of herds sampled had greater than 15% prevalence of elevated fatty acid concentration (>0.70 mmol/L) in cows between 3 and 14 DIM. Elevated fatty acid concentrations are arguably more significant than elevated BHB concentrations, as fatty acid concentrations have been demonstrated to be more stable over time than BHB concentrations

Table 7. Results of partial least square discriminant analysis models for the classification of serum metabolite concentrations based on metabolic profile thresholds using milk mid-infrared spectra

Metabolite	*P*-value[1]	LV[2]	Cross-validation (n = 579)				Random validation (n = 194)				Independent validation (n = 105)			
			Sens[3]	Spec[4]	CE[5]	AUC[6]	Sens	Spec	CE	AUC	Sens	Spec	CE	AUC
BHB	>0.05	4	1.00	0.94	0.03	0.97	0.40	0.93	0.33	0.92	1.00	0.83	0.03	0.99
Fatty acids	<0.001	5	0.82	0.75	0.22	0.85	0.73	0.81	0.23	0.87	0.25	0.90	0.42	0.82
Urea	<0.001	20	0.81	0.91	0.15	0.94	0.90	0.98	0.06	0.98	1.00	0.89	0.06	—
Calcium	>0.05	2	0.36	0.72	0.46	0.61	0.00	0.76	0.62	0.51	1.00	0.06	0.47	0.75
Magnesium	>0.05	2	0.33	0.93	0.37	0.61	0.00	0.91	0.55	0.54	1.00	0.95	0.02	—
Albumin	>0.05	3	0.67	0.67	0.33	0.73	0.53	0.64	0.41	0.59	0.25	0.99	0.38	0.85
Globulin	>0.05	5	0.46	0.73	0.41	0.63	0.75	0.76	0.24	0.87	1.00	0.42	0.29	0.99

[1]*P*-value for pairwise Wilcoxon signed rank test.

[2]Number of latent variables (LV) included in the model.

[3]Sensitivity.

[4]Specificity.

[5]Classification error.

[6]Area under the receiver operator characteristic curve.

(Eicher et al., 1999) and have a higher association with subsequent adverse health events (Ospina et al., 2010a; Sordillo and Raphael, 2013).

A considerable percentage of animals sampled (31%) had serum urea concentrations greater than 6.8 mmol/L, whereas less than 2% had a urea concentration less than 1.7 mmol/L. These results are consistent with animals grazing rapidly growing forage with high levels of RDP (Macrae et al., 2006). Urea concentrations in serum can be used to monitor RDP intake and the ratio of RDP to energy in the ration (Roseler et al., 1993; Macrae et al., 2006). Low blood urea concentrations can indicate insufficient RDP intake, whereas high serum urea concentrations can indicate excessive RDP intake, often in the form of high protein pasture. Both insufficient and excessive RDP intake are known to limit milk production, the former due to reduced rumen microbial protein synthesis and the latter due to the significant metabolic cost of removing and detoxifying excess nitrogenous by-products from the rumen (Waghorn and Wolff, 1984; Ulyatt, 1997). Elevated serum urea concentrations before AI have also been suggested to have a negative effect on reproductive performance (Raboisson et al., 2017). Several studies have used serum urea concentrations to predict nitrogen efficiency and urinary nitrogen excretion in cattle (Kohn et al., 2005; Kume et al., 2008), which is becoming an environmental concern for the global dairy industry.

Our results indicate that metabolic disorders are prevalent in the Australian dairy herd, but further studies are required to better understand the epidemiology of early-lactation metabolic disorders. Further work is also required to determine appropriate serum metabolic profile concentration thresholds and herd-level thresholds for Australia's diverse dairy production systems.

PLS Regression Models for Predicting Serum Metabolite Concentrations

The accuracy of our MIR prediction model for serum BHB was moderate ($0.48 \leq R^2 \leq 0.60$), which was better than those reported by Belay et al. (2017a) and similar to those reported by Smith et al. (2016) and Pralle et al. (2018). The moderate accuracy may have been in part due to the low prevalence of elevated serum BHB concentrations in our data set. The skewed distribution of the data likely resulted in lower prediction accuracy of higher BHB concentrations. Despite this, the R^2_{IV} was higher than the R^2_{CV} and the R^2_{RV}, suggesting that the model may be a useful indicator of ketosis risk when applied to independent data. This finding was supported by fact that the RMSE of all 3 validation methods were similar (0.11–0.12 mmol/L). The reported accuracies of MIR predictions of milk ketone bodies are considerably higher than the accuracies of MIR serum BHB predictions (de Roos et al., 2007; Grelet et al., 2016); however, serum BHB is considered to be a superior biomarker of ketosis (Duffield et al., 1997; Denis-Robichaud et al., 2014). Whether a less-accurate predictor of the gold standard biomarker is superior to a higher accuracy predictor of a less-valuable biomarker requires further investigation. This discussion is further complicated by the fact that, arguably, no true gold standard tests exist for many of the animal health traits being investigated (Krogh et al., 2011).

Serum fatty acid concentrations are routinely used to quantify the degree of fat mobilization, and therefore the magnitude of negative energy balance in early lactation (Ospina et al., 2010a). Few studies, however, have investigated the use of MIR of milk for predicting serum fatty acid concentrations. The R^2_{RV} of our fatty acids prediction was 0.61, which is similar to the result of Smith et al. (2016), who reported a correlation coefficient of 0.80 ($R^2 = 0.64$) between measured serum fatty acid concentration and MIR-predicted fatty acids concentration. McParland et al. (2011) were able to predict computed energy balance with reasonable accuracy ($R^2 = 0.56$), but they noted that their prediction equations were not robust when applied to data obtained from cows managed differently to the animals in the reference population (McParland et al., 2012). Similarly, the accuracy of our fatty acids prediction was lower when applied to the independent farm E validation data set ($R^2 = 0.45$), but may still be a useful indicator of energy balance.

Mid-infrared is routinely used to predict MUN concentrations (Gengler et al., 2016) with good accuracy. Serum and milk urea concentrations are linearly correlated (Moore and Varga, 1996), so it follows that the MIR prediction of serum urea concentration had the highest coefficient of determination of all the metabolite models we tested. The accuracy of the urea prediction model, when applied to data from farm E, was considerably lower than when applied to the random external validation data. This may have been due to differences in the distribution of urea concentrations between the farm E independent data set and the reference population data set (see Table 3), as the range of data is known to have a significant effect on the R^2 of PLS regression models (Davies and Fearn., 2006). Further validation with larger, more varied data sets is required to better understand these results.The MIR predictions of serum urea may be accurate enough to be a useful indicator of the protein nutrition of a herd. Large-scale predictions of serum urea concentrations could also be

used to identify variation in nitrogen efficiency and excretion between individuals. This could be exploited in breeding programs to lower the environmental impact of dairy cattle and could also be considered as a way to increase the accuracy of genomic predictions of feed efficiency due to the high energy cost of removing excess nitrogenous by-products from the rumen. The MIR predictions of serum urea concentration could also help manage the nitrogen output of the global dairy industry, which is becoming an increasingly important environmental issue.

Mid-infrared spectroscopy has been used successfully to quantify the concentration of several milk proteins, including casein, α_{S1}-CN, whey protein, and β-LG, with reasonable accuracy (De Marchi et al., 2009; Bonfatti et al., 2011; McDermott et al., 2016); however, MIR predictions of serum albumin and globulin concentrations in our study were extremely poor. Similarly, MIR spectral data has been used to estimate the Ca and P concentration of milk with reasonable accuracy (Toffanin et al., 2015); however, its ability to predict serum Ca and Mg concentrations was poor. Studies have demonstrated that animals suffering from subclinical hypocalcemia showed significant changes in their serum proteome (Wang et al., 2016; Fan et al., 2017). Given that changes in serum albumin and globulin concentrations could not be identified with milk MIR spectra, it follows that changes in serum proteome associated with hypocalcemia are also not detectable using milk MIR spectral analysis. This may suggest that changes in serum protein concentrations are not reflected in milk composition, or that a significant delay occurs between changes in serum metabolome and subsequent changes in milk composition. It should also be noted that we found considerably less variation in the concentrations of these metabolites (see Table 2) compared with the concentrations of BHB, fatty acids, and urea. As discussed previously, the distribution of data is known to have a significant effect on the accuracy of calibration models. Our results highlight the need for further work to investigate the relationships between the proteomes and metabolomes of serum and milk.

Several authors have observed that many milk MIR predictions of animal health traits are not sufficiently accurate to provide useful information on the health status of individual animals (de Roos et al., 2007; van der Drift et al., 2012; Grelet et al., 2016). These predictions may, however, be considered accurate enough to employ in genomic evaluations, as described by Bastin et al. (2016) and Bonfatti et al. (2017). The accuracy of our MIR prediction of serum BHB was better than that of Belay et al. (2017b), who used MIR predictions of serum BHB to investigate the genetic parameters of ketosis and the genetic relationships between serum BHB concentration and milk production traits. Similarly, our MIR prediction of serum fatty acids concentration had comparable accuracy to the MIR prediction of energy balance reported by McParland et al. (2015), who found favorable correlations between MIR-predicted and measured energy balance. As well as forming the basis of new traits, Pryce et al. (2016) suggested that MIR-predicted traits could be included in multitrait models to improve the accuracy of existing genomic predictions, but exactly how accurate such MIR predictions need to be to provide useful phenotypic information requires further investigation.

PLS-DA for Classifying Animals Based on Metabolic Profile Testing Thresholds

The aim of metabolic profile testing is to gain objective information on the nutritional status and metabolic health of a herd by estimating the prevalence of certain metabolic disorders. This requires a diagnostic test that can classify animals as being either affected or not affected with metabolic disorders with reasonable accuracy. The PLS-DA models are routinely used in chemometric studies to classify samples based on multivariate data. It follows, therefore, that these models may be useful for identifying cows with metabolic disorders based on their milk MIR spectra. Several authors have converted MIR predictions of continuous traits, such as BHB concentration, into binary traits based on metabolic profile thresholds (Gelé et al., 2015; Pralle et al., 2018). As far as we know, ours is the first report using PLS-DA models to classify animals directly using milk MIR spectral data. Given the poor accuracy of MIR predictions of serum Ca, Mg, albumin, and globulin concentrations, only BHB, fatty acids, and urea PLS-DA models will be discussed beyond this point.

Cows that experience a serum fatty acid concentration greater than 0.7 mmol/L in the immediate postpartum period are more likely to develop clinical ketosis, metritis, or a displaced abomasum, and are more likely to be culled early than cows with normal fatty acid concentrations (McArt et al., 2013). At herd level, several authors have demonstrated that an increase in the prevalence of cows with elevated postpartum fatty acid concentrations is associated with reduced milk production and poorer fertility (Ospina et al., 2010a; Chapinal et al., 2012; McArt et al., 2013). The same authors demonstrated similar results for elevated postpartum BHB concentrations. When applied to the random validation data set, our PLS-DA model was able to predict elevated serum fatty acid concentrations (>0.7 mmol/L) with a sensitivity of 73% and a specificity of 81%. Our findings are reasonably consistent with those of Gelé et al. (2015), who used a combination of BHB

and fatty acid concentrations to identify animals at risk of subclinical ketosis with a sensitivity of 81% and a specificity of 69%. The true prevalence of elevated fatty acid concentrations in our data set was 29% (95% CI = 22.5–35.2). The MIR-predicted prevalence of elevated serum fatty acid concentrations was 35%, within the 95% confidence interval of the true prevalence. When applied to the farm E validation set, the sensitivity of our fatty acids prediction decreased significantly to only 25%. The MIR-predicted prevalence of elevated fatty acid concentrations in this data set was 11%, which was significantly higher than the true prevalence of 4% (95% CI = 0.1–7.5). This highlights the need for larger and more diverse calibration data sets to improve the accuracy of predictions before they can be used for on-farm management purposes.

Butler et al. (1996) reported that animals with a plasma urea concentration of greater than 6.8 mmol/L had lower pregnancy rates than animals with normal plasma urea concentrations. Similarly, Raboisson et al. (2017) showed a 43% lower odds of pregnancy when serum urea concentrations were above 7 mmol/L, especially before AI. This may be highly relevant for the largely pasture-based Australian dairy industry, as most cows are mated in late spring at a time when they are grazing pasture high in RDP. When applied to the random validation data set, PLS-DA models could identify animals with elevated serum urea concentrations with good sensitivity (90%) and specificity (98%). The predicted prevalence of elevated urea concentrations was 33%, close to the true prevalence of 36% (95% CI = 28.8–42.3). When applied to the independent farm E validation data set, the sensitivity of the urea prediction increased to 100% and the specificity decreased to 89%. We believe these results are misleading and are artifacts of the independent validation data set being (1) relatively small, (2) having a different distribution and narrower range than the calibration data set, and (3) containing no positive results (urea concentrations >6.8 mmol/L). Given that the R^2_{IV} of the PLS prediction of serum urea concentration was considerably lower than the R^2_{RV} (0.35 and 0.90, respectively), we would expect the accuracy of PLS-DA predictions to be similarly lower. The MIR-predicted prevalence of elevated urea concentrations was 11%, significantly higher than the true prevalence of 0%. This is further evidence that a larger, more varied calibration data set is required to improve the accuracy of predictions.

The results of our PLS-DA model to predict elevated serum BHB concentrations were not significant ($P > 0.05$) based on pairwise Wilcoxon rank testing of permuted samples. This was likely due to the low number of hyperketonemic samples in the data set. Lowering the threshold of BHB concentration to 1.0 mmol/L increased the statistical significance of the model. Given that the accuracy of BHB and fatty acids PLS models were similar, we would expect that the addition of more hyperketonemic samples to our data set will yield a statistically significant PLS-DA model for estimating the prevalence of subclinical ketosis.

If their accuracy can be improved, PLS-DA predictions offer a potentially useful tool to monitor the prevalence of elevated serum fatty acids, BHB, and urea concentrations. This could provide dairy producers with a valuable early warning tool that would allow them to address dietary imbalances, and thereby optimize animal health, production, and fertility.

CONCLUSIONS

We assessed the accuracy of MIR spectroscopy, performed as part of routine milk recording, for predicting the metabolic health and nutritional status of early-lactation dairy cows. We found that MIR spectroscopy of milk provided a potentially useful prediction of energy balance by reasonable estimation of serum BHB and fatty acid concentrations. The accuracy of MIR prediction of serum urea concentration was good when the validation data set had a similar range and distribution to the calibration data set. However, when the model was applied to an independent data set taken from cows of differing breeds that were managed differently, the accuracy of the prediction dropped significantly. The accuracy of MIR predictions of serum Ca, Mg, albumin, and globulin concentrations were poor. Our results demonstrate that MIR PLS-DA models may be a useful tool for estimating the prevalence of metabolic disorders in early lactation, but more data are required to improve the accuracy of prediction equations. The MIR PLS models offer potential for large-scale phenotyping that can be employed in breeding programs to breed more resilient animals with smaller environmental footprints. We aim to improve the accuracy of our prediction models by sampling more animals, particularly in the first 2 wk of lactation, thereby increasing the size and variation of our data set.

ACKNOWLEDGMENTS

We thank DairyBio, jointly funded by Dairy Australia (Melbourne, Australia) and Agriculture Victoria (Melbourne, Australia), for funding this project and T. Luke's PhD project. This work was part of a project "MIRprofit: integrating very large genomic and milk midinfrared data to improve profitability of dairy cows," supported by funding from the Australian Gov-

ernment Department of Agriculture (Canberra, Australia) as part of the Rural R&D for Profit programme. The authors also thank Di Mapleson, Brigid Ribaux, and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance, Erika Oakes and Michelle Axford of Datagene (Bundoora, Australia) for their work coordinating this study, and the farmers who took part in this project.

REFERENCES

Anderson, D. E. 2009. Current Veterinary Therapy: Food Animal Practice. Saunders Elsevier, St. Louis, MO.

Afseth, N. K., H. Martens, A. Randby, L. Gidskehaug, B. Narum, K. Jorgensen, S. Lien, and A. Kohler. 2010. Predicting the fatty acid composition of milk: A comparison of two Fourier transform infrared sampling techniques. Appl. Spectrosc. 64:700–707. https://doi.org/10.1366/000370210791666200.

Bastin, C., L. Théron, A. Lainé, and N. Gengler. 2016. On the role of mid-infrared predicted phenotypes in fertility and health dairy breeding programs. J. Dairy Sci. 99:4080–4094. https://doi.org/10.3168/jds.2015-10087.

Belay, T. K., B. S. Dagnachew, Z. M. Kowalski, and T. Ådnøy. 2017a. An attempt at predicting blood β-hydroxybutyrate from Fourier-transform mid-infrared spectra of milk using multivariate mixed models in Polish dairy cattle. J. Dairy Sci. 100:6312–6326. https://doi.org/10.3168/jds.2016-12252.

Belay, T. K., M. Svendsen, Z. M. Kowalski, and T. Ådnøy. 2017b. 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. 100:6298–6311. https://doi.org/10.3168/jds.2016-12458.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011. 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. 94:5776–5785. https://doi.org/10.3168/jds.2011-4401.

Bonfatti, V., D. Vicario, A. Lugo, and P. Carnier. 2017. 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. 100:5526–5540. https://doi.org/10.3168/jds.2016-11667.

Butler, W. R., J. J. Calaman, and S. W. Beam. 1996. Plasma and milk urea nitrogen in relation to pregnancy rate in lactating dairy cattle. J. Anim. Sci. 74:858–865.

Chapinal, N., S. J. Leblanc, M. E. Carson, K. E. Leslie, S. Godden, M. Capel, J. E. Santos, M. W. Overton, and T. F. Duffield. 2012. Herd-level association of serum metabolites in the transition period with disease, milk production, and early lactation reproductive performance. J. Dairy Sci. 95:5676–5682. https://doi.org/10.3168/jds.2011-5132.

Compton, C. W., S. McDougall, L. Young, and M. Bryan. 2014. Prevalence of subclinical ketosis in mainly pasture-grazed dairy cows in New Zealand in early lactation. N. Z. Vet. J. 62:30–37. https://doi.org/10.1080/00480169.2013.823829.

Compton, C. W., L. Young, and S. McDougall. 2015. Subclinical ketosis in post-partum dairy cows fed a predominantly pasture-based diet: defining cut-points for diagnosis using concentrations of beta-hydroxybutyrate in blood and determining prevalence. N. Z. Vet. J. 63:241–248. https://doi.org/10.1080/00480169.2014.999841.

Davies, A., and T. Fearn. 2006. Back to basics: Calibration statistics. Spectroscopy Europe 18:31–32.

De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2009. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Ital. J. Anim. Sci. 8:399–401. https://doi.org/10.4081/ijas.2009.s2.399.

de Roos, A. P. W., H. J. C. M. van Den Bijgaart, J. Hørlyk, and G. de Jong. 2007. Screening for subclinical ketosis in dairy cattle by fourier transform infrared spectrometry. J. Dairy Sci. 90:1761–1766. https://doi.org/10.3168/jds.2006-203.

DeGaris, P. J., and I. J. Lean. 2008. Milk fever in dairy cows: A review of pathophysiology and control principles. Vet. J. 176:58–69. https://doi.org/10.1016/j.tvjl.2007.12.029.

Denis-Robichaud, J., J. Dubuc, D. Lefebvre, and L. Descôteaux. 2014. Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. J. Dairy Sci. 97:3364–3370. https://doi.org/10.3168/jds.2013-6744.

Duffield, T. F., D. F. Kelton, K. E. Leslie, K. D. Lissemore, and J. H. Lumsden. 1997. Use of test day milk fat and milk protein to detect subclinical ketosis in dairy cattle in Ontario. Can. Vet. J. 38:713.

Eicher, R., A. Liesegang, E. Bouchard, and A. Tremblay. 1999. Effect of cow-specific factors and feeding frequency of concentrate on diurnal variations of blood metabolites in dairy cows. Am. J. Vet. Res. 60:1493–1499.

Eigenvector. 2018. PLS Toolbox User Guide. Vol. 2018. Eignevector Research Inc., Manson, WA.

Fan, Z., S. Shu, C. Xu, X. Xiao, G. Wang, Y. Bai, C. Xia, L. Wu, H. Zhang, C. Xu, and W. Yang. 2017. Protein profiling of plasma proteins in dairy cows with subclinical hypocalcaemia. Ir. Vet. J. 70:3. https://doi.org/10.1186/s13620-017-0082-0.

Gelé, M., M. Ferrand-Calmels, L. Brun-Lafleur, A. Werner, and F. Gollé-Leidreiter. 2015. Predicting the risk of ketosis using mid infrared spectrometry. Pages 19–24 in Proceedings of ICAR Technical Series. International Committee for Animal Recording (ICAR), Rome, Italy.

Gengler, N., H. Soyeurt, F. Dehareng, C. Bastin, F. Colinet, H. Hammami, M. L. Vanrobays, A. Lainé, S. Vanderick, C. Grelet, A. Vanlierde, E. Froidmont, and P. Dardenne. 2016. Capitalizing on fine milk composition for breeding and management of dairy cows. J. Dairy Sci. 99:4071–4079. https://doi.org/10.3168/jds.2015-10140.

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477.

Hamann, J., and V. Krömker. 1997. Potential of specific milk composition variables for cow health management. Livest. Prod. Sci. 48:201–208. https://doi.org/10.1016/S0301-6226(97)00027-4.

Jonker, J. S., R. A. Kohn, and J. High. 2002. Use of milk urea nitrogen to improve dairy cow diets. J. Dairy Sci. 85:939–946. https://doi.org/10.3168/jds.S0022-0302(02)74152-0.

Kohn, R. A., M. M. Dinneen, and E. Russek-Cohen. 2005. Using blood urea nitrogen to predict nitrogen excretion and efficiency of nitrogen utilization in cattle, sheep, goats, horses, pigs, and rats. J. Anim. Sci. 83:879. https://doi.org/10.2527/2005.834879x.

Krogh, M. A., N. Toft, and C. Enevoldsen. 2011. Latent class evaluation of a milk test, a urine test, and the fat-to-protein percentage ratio in milk to diagnose ketosis in dairy cows. J. Dairy Sci. 94:2360–2367. https://doi.org/10.3168/jds.2010-3816.

Kume, S., K. Numata, Y. Takeya, Y. Miyagawa, S. Ikeda, M. Kitagawa, K. Nonaka, T. Oshita, and T. Kozakai. 2008. Evaluation of urinary nitrogen excretion from plasma urea nitrogen in dry and lactating cows. (Report). Asian-Australas. J. Anim. Sci. 21:1159. https://doi.org/10.5713/ajas.2008.70734.

LeBlanc, S. J., K. D. Lissemore, D. F. Kelton, T. F. Duffield, and K. E. Leslie. 2006. Major advances in disease prevention in dairy cattle. J. Dairy Sci. 89:1267–1279. https://doi.org/10.3168/jds.S0022-0302(06)72195-6.

Macrae, A. I., D. A. Whitaker, E. Burrough, A. Dowell, and J. M. Kelly. 2006. Use of metabolic profiles for the assessment of dietary adequacy in UK dairy herds. Vet. Rec. 159:655. https://doi.org/10.1136/vr.159.20.655.

McArt, J. A. A., D. V. Nydam, and G. R. Oetzel. 2012. Epidemiology of subclinical ketosis in early lactation dairy cattle. J. Dairy Sci. 95:5056–5066. https://doi.org/10.3168/jds.2012-5443.

McArt, J. A. A., D. V. Nydam, G. R. Oetzel, T. R. Overton, and P. A. Ospina. 2013. Elevated non-esterified fatty acids and

β-hydroxybutyrate and their association with transition dairy cow performance. Vet. J. 198:560. https://doi.org/10.1016/j.tvjl.2013.08.011.

McArt, J. A. A., D. V. Nydam, and M. W. Overton. 2015. Hyperketonemia in early lactation dairy cattle: A deterministic estimate of component and total cost per case. J. Dairy Sci. 98:2043–2054. https://doi.org/10.3168/jds.2014-8740.

McDermott, A., G. Visentin, M. De Marchi, D. P. Berry, M. A. Fenelon, P. M. O'connor, O. A. Kenny, and S. McParland. 2016. 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. 99:3171–3182. https://doi.org/10.3168/jds.2015-9747.

McParland, S., G. Banos, B. McCarthy, E. Lewis, M. P. Coffey, B. O'neill, M. O'donovan, E. Wall, and D. P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. J. Dairy Sci. 95:7225–7235. https://doi.org/10.3168/jds.2012-5406.

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651–3661. https://doi.org/10.3168/jds.2010-3965.

McParland, S., E. Kennedy, E. Lewis, S. G. Moore, B. McCarthy, M. O'donovan, and D. P. Berry. 2015. Genetic parameters of dairy cow energy intake and body energy status predicted using mid-infrared spectrometry of milk. J. Dairy Sci. 98:1310–1320. https://doi.org/10.3168/jds.2014-8892.

Moore, D. A., and G. Varga. 1996. BUN and MUN: Urea nitrogen testing in dairy cattle. Compend. Contin. Educ. Pract. Vet. 18:712–720.

National Health and Medical Research Council. 2013. Australian Code for the Care and Use of Animals for Scientific Purposes. 8th ed. National Health and Medical Research Council, Canberra, Australia.

Nousiainen, J., K. J. Shingfield, and P. Huhtanen. 2004. Evaluation of milk urea nitrogen as a diagnostic of protein feeding. J. Dairy Sci. 87:386–398. https://doi.org/10.3168/jds.S0022-0302(04)73178-1.

Oetzel, G. R. 2004. Monitoring and testing dairy herds for metabolic disease. Vet. Clin. North Am. Food Anim. Pract. 20:651–674. https://doi.org/10.1016/j.cvfa.2004.06.006.

Ospina, P. A., D. Nydam, T. Stokol, and T. Overton. 2010a. Associations of elevated nonesterified fatty acids and [beta]-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. J. Dairy Sci. 93:1596–1603.

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010b. Association between the proportion of sampled transition cows with increased nonesterified fatty acids and beta-hydroxybutyrate and disease incidence, pregnancy rate, and milk production at the herd level. J. Dairy Sci. 93:3595–3601. https://doi.org/10.3168/jds.2010-3074.

Paudyal, S., F. P. Maunsell, C. A. Risco, A. Donovan, A. D. Vries, D. Manriquez, and P. J. Pinedo. 2016. Evaluating milk fat to protein ratio and milk fat to lactose ratio as indicators for early lactation disease. J. Anim. Sci. 94:69. (Abstr.)

Payne, J. M., S. M. Dew, R. Manston, and M. Faulks. 1970. The use of a metabolic profile test in dairy herds. Vet. Rec. 87:150. https://doi.org/10.1136/vr.87.6.150.

Pralle, R. S., K. W. Weigel, and H. M. White. 2018. Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. J. Dairy Sci. 101:4378–4387. https://doi.org/10.3168/jds.2017-14076.

Pryce, J. E., K. L. Parker Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B. Heringstad, C. Egger-Danner, K. F. Stock, A. J. Bradley, and J. B. Cole. 2016. Invited review: Opportunities for genetic improvement of metabolic diseases. J. Dairy Sci. 99:6855–6873. https://doi.org/10.3168/jds.2016-10854.

Raboisson, D., A. Albaaj, G. Nonne, and G. Foucras. 2017. High urea and pregnancy or conception in dairy cows: A meta-analysis to define the appropriate urea threshold. J. Dairy Sci. 100:7581. (Abstr.) https://doi.org/10.3168/jds.2016-12009.

Roseler, D. K., J. D. Ferguson, C. J. Sniffen, and J. Herrema. 1993. Dietary protein degradability effects on plasma and milk urea nitrogen and milk nonprotein nitrogen in Holstein cows. J. Dairy Sci. 76:525–534. https://doi.org/10.3168/jds.S0022-0302(93)77372-5.

Schcolnik, T. 2016. Using milk fat-to-protein ratio to evaluate dairy cows energy balance status. J. Anim. Sci. 94:54. (Abstr.)

Smith, S., V. Hicks, R. Cooper, J. Forrest, A. Macrae, and E. Wall. 2016. Developing a bovine ketosis risk indicator using spectral and phenotypic data. Page 301 in Proc. 67th Annu. Meeting of EAAP, Belfast, UK. Wageningen Academic Publishers, Wageningen, the Netherlands.

Sordillo, L. M., and W. Raphael. 2013. Significance of metabolic stress, lipid mobilization, and inflammation on transition cow disorders. Vet. Clin. North Am. Food Anim. Pract. 29:267–278. https://doi.org/10.1016/j.cvfa.2013.03.002.

Suthar, V. S., J. Canelas-Raposo, A. Deniz, and W. Heuwieser. 2013. Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. J. Dairy Sci. 96:2925–2938. https://doi.org/10.3168/jds.2012-6035.

Toffanin, V., M. De Marchi, N. Lopez-Villalobos, and M. Cassandro. 2015. 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. 41:68–73. https://doi.org/10.1016/j.idairyj.2014.10.002.

Toni, F., L. Vincenti, L. Grigoletto, A. Ricci, and Y. H. Schukken. 2011. Early lactation ratio of fat and protein percentage in milk is associated with health, milk production, and survival. J. Dairy Sci. 94:1772–1783. https://doi.org/10.3168/jds.2010-3389.

Ulyatt, M. J. 1997. Can protein utilisation from pasture be improved? Pages 4–8 in Proc. NZ Soc. of Anim. Prod. Annu. Meeting, Palmerston North, New Zealand. N. Z. Soc. Anim. Prod., Wellington, New Zealand.

van der Drift, S. G. A., R. Jorritsma, J. T. Schonewille, H. M. Knijn, and J. A. Stegeman. 2012. Routine detection of hyperketonemia in dairy cows using Fourier transform infrared spectroscopy analysis of β-hydroxybutyrate and acetone in milk in combination with test-day information. J. Dairy Sci. 95:4886–4898. https://doi.org/10.3168/jds.2011-4417.

van Knegsel, A. T. M., S. G. A. van Der Drift, M. Horneman, A. P. W. de Roos, B. Kemp, and E. A. M. Graat. 2010. Short communication: Ketone body concentration in milk determined by Fourier transform infrared spectroscopy: Value for the detection of hyperketonemia in dairy cows. J. Dairy Sci. 93:3065–3069. https://doi.org/10.3168/jds.2009-2847.

Waghorn, G. C., and J. E. Wolff. 1984. Theoretical considerations for partitioning nutrients between muscle and adipose tissue. Pages 193–200 in Proc. NZ Soc. of Anim. Prod. Annu. Meeting. N. Z. Soc. Anim. Prod., Wellington, New Zealand.

Wang, P. X., S. Shu, C. Xia, Z. Wang, L. Wu, B. Wang, C. C. Xu, and J. Liu. 2016. Protein expression in dairy cows with and without subclinical hypocalcaemia. N. Z. Vet. J. 64:101–106. https://doi.org/10.1080/00480169.2015.1100970.

Whitaker, D. A. 2004. Metabolic profiles. Pages 804–817 in Bovine Medicine: Diseases and Husbandry of Cattle. 2nd ed. Blackwell Science, Oxford, UK.

Chapter 4:

A Tale of Two Biomarkers: Untargeted ^{1}H NMR Metabolomic Fingerprinting of BHBA and NEFA in Early Lactation Dairy Cows

Metabolites. 2020, 10, 247. doi:10.3390/metabo10060247

T. D. W. Luke, J. E. Pryce, W. J. Wales and S. J. Rochfort

Published: 15th of June 2020 (open access)

metabolites

MDPI

Article

A Tale of Two Biomarkers: Untargeted ^{1}H NMR Metabolomic Fingerprinting of BHBA and NEFA in Early Lactation Dairy Cows

Timothy D. W. Luke [1,2], Jennie E. Pryce [1,2], William J. Wales [3,4] and Simone J. Rochfort [1,2,*]

1 Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia; tim.luke@agriculture.vic.gov.au (T.D.W.L.); jennie.pryce@agriculture.vic.gov.au (J.E.P.)
2 School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia
3 Agriculture Victoria Research, Ellinbank Centre, Ellinbank, VIC 3821, Australia; bill.wales@agriculture.vic.gov.au
4 Centre for Agricultural Innovation, School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC 3010, Australia
* Correspondence: simone.rochfort@agriculture.vic.gov.au

Received: 21 May 2020; Accepted: 12 June 2020; Published: 15 June 2020

Abstract: Disorders of energy metabolism, which can result from a failure to adapt to the period of negative energy balance immediately after calving, have significant negative effects on the health, welfare and profitability of dairy cows. The most common biomarkers of energy balance in dairy cows are β-hydroxybutyrate (BHBA) and non-esterified fatty acids (NEFA). While elevated concentrations of these biomarkers are associated with similar negative health and production outcomes, the phenotypic and genetic correlations between them are weak. In this study, we used an untargeted ^{1}H NMR metabolomics approach to investigate the serum metabolomic fingerprints of BHBA and NEFA. Serum samples were collected from 298 cows in early lactation (calibration dataset $N = 248$, validation $N = 50$). Metabolomic fingerprinting was done by regressing ^{1}H NMR spectra against BHBA and NEFA concentrations (determined using colorimetric assays) using orthogonal partial least squares regression. Prediction accuracies were high for BHBA models, and moderately high for NEFA models (R^2 of external validation of 0.88 and 0.75, respectively). We identified 16 metabolites that were significantly (variable importance of projection score > 1) correlated with the concentration of one or both biomarkers. These metabolites were primarily intermediates of energy, phospholipid, and/or methyl donor metabolism. Of the significant metabolites identified; (1) two (acetate and creatine) were positively correlated with BHBA but negatively correlated with NEFA, (2) nine had similar associations with both BHBA and NEFA, (3) two were correlated with only BHBA concentration, and (4) three were only correlated with NEFA concentration. Overall, our results suggest that BHBA and NEFA are indicative of similar metabolic states in clinically healthy animals, but that several significant metabolic differences exist that help to explain the weak correlations between them. We also identified several metabolites that may be useful intermediate phenotypes in genomic selection for improved metabolic health.

Keywords: metabolic profile; ketosis; transition period; livestock; methyl donor; one-carbon metabolism; negative energy balance

1. Introduction

Most dairy cows experience a period of negative energy balance immediately after calving due to both a reduction in feed intake preceding calving [1], and an increase in energy requirements for milk production [2]. A successful transition from pregnancy to lactation requires a series of complex and

coordinated changes in metabolism and nutrient partitioning, known as homeorhesis [3]. Failure of these homeorhetic controls can lead to the development of metabolic disorders such as ketosis and fatty liver [4]. These disorders can have significant negative effects on the health, welfare and profitability of early-lactation dairy cows due to their (1) relatively high incidence [5,6], (2) demonstrated association with other diseases [4,7] and (3) their significant economic costs [8,9].

Serum β-hydroxybutyrate (BHBA) and non-esterified fatty acids (NEFA) are biomarkers that are commonly used to evaluate the energy balance of dairy cows in the transition period [6,10,11]. One of the main physiological responses to reduced energy intake is the mobilization of stored energy from adipose tissue as NEFA. Serum NEFA concentration is a measure of the degree of lipolysis, and therefore an indicator of the magnitude of negative energy balance [12]. Once released, NEFA are transported via the bloodstream to the mammary gland for milk fat synthesis, or to the liver where they undergo either (1) complete oxidation via the TCA cycle, (2) partial oxidation to ketone bodies (BHBA, acetone and acetoacetate), or (3) re-esterification to form triglycerides which can either be stored or exported as very low density lipoprotein (VLDL). BHBA is the most stable of the three ketone bodies [13], and is commonly used as a biomarker of energy balance [14].

Mild elevations in serum BHBA and/or NEFA concentration during the transition period are considered normal [15], but marked elevations are indicative of excessive negative energy balance and/or perturbed metabolism [16]. Elevated concentrations of both BHBA and NEFA can be observed in clinically healthy animals (i.e., showing no visible signs of illness), and are associated with (1) reduced reproductive performance [11,17], (2) an increased incidence of clinical diseases such as displaced abomasa and metritis [15,17,18], (3) decreased milk production [6,11,19] and (4) an increased risk of culling [6,15,20]. However, despite these similarities, both the phenotypic [21,22] and genetic [23] correlations between these two biomarkers are low. This is not necessarily important if biomarkers are being used for management purposes (such as the identification of sick animals or the assessment of nutritional status) but may be significant if the biomarkers are used as phenotypes for genetic selection for improved animal health and resilience. There is therefore a need to better understand the metabolic states represented by BHBA and NEFA.

Untargeted metabolomics combines high throughput molecular analytical techniques such as proton nuclear magnetic resonance (^{1}H NMR) spectroscopy with multivariate statistical modelling, to characterize the metabolic response of a biological system to pathophysiological stimuli [24]. Examples in dairy cattle include studies of ketosis [25,26], fatty liver [27], hypocalcaemia [28] and displaced abomasa [29]. The collective metabolic features of a given state or condition can be described as its "metabolomic fingerprint". As well improving our understanding of the biological processes, metabolomic studies can uncover intermediate molecular phenotypes (metabotypes) associated with complex animal health traits such as metabolic resilience. These metabotypes can then be integrated with genomic data to (1) elucidate the genetic architecture of these traits, and (2) improve genomic prediction accuracies [30,31].

The aim of this study was therefore to use an untargeted ^{1}H NMR metabolomic approach to investigate the metabolomic fingerprints of serum BHBA and NEFA concentrations in clinical healthy dairy cows in early lactation, and in so doing (1) identify common and differential metabolic pathways, and (2) identify novel metabotypes for application to genetic selection for improved metabolic health.

2. Results

2.1. Analysis of Experimental Metadata

Descriptive statistics of the datasets used in this experiment are shown in Table 1. BHBA concentrations were significantly higher in Dataset 1 than in Dataset 2 ($p < 0.001$). The differences in all other parameters were not statistically significant ($p > 0.05$). The correlation between BHBA and NEFA concentrations was 0.45 in Dataset 1 and 0.40 in Dataset 2.

Table 1. Descriptive statistics of the datasets used in this experiment, including number of animals (N), stage of lactation defined as days in milk (DIM), age in years, and β-hydroxybutyrate (BHBA) and non-esterified fatty acid (NEFA) concentrations (mmol/L) in the serum obtained from clinically healthy dairy cows.

Variable	Dataset 1 (N = 248)			Dataset 2 (N = 50)			p [1]
	Min	Max	Mean (SD)	Min	Max	Mean (SD)	
DIM (days)	4	30	16.7 (6.0)	4	30	18.6 (7.3)	0.09
Age (years)	2	12	3.7 (2.0)	2	9	3.9 (1.8)	0.22
BHBA (mmol/L)	0.22	1.86	0.55 (0.21)	0.23	0.94	0.42 (0.17)	<0.001
NEFA (mmol/L)	0.11	2.18	0.75 (0.32)	0.14	1.91	0.67 (0.36)	0.07

[1] Statistical significance of the differences between Datasets 1 and 2 were determined using paired *t*-test for DIM, and a paired Wilcoxon signed-rank test for age, BHBA and NEFA.

2.2. ^{1}H NMR Spectra

Twenty-four metabolites could be clearly identified from the ^{1}H NMR spectra. Two metabolites, cholate and 3-phenyllactate, were tentatively identified. Figure 1 shows representative spectra from animals in Dataset 1 with (a) elevated BHBA concentration, (b) elevated NEFA concentration and (c) normal BHBA and NEFA concentrations. Upfield regions of spectra were dominated by branched-chain amino acids (leucine, isoleucine and valine), organics acids (BHBA, lactate, acetate) and the methyl and methylene groups of low density (LDL) and very low density lipoproteins (VLDL) at δ 0.86 ppm and δ 1.25 ppm, respectively [32]. We also observed a prominent peak at δ 2.03 ppm which was consistent with the N-acetyl groups of glycoproteins [33]. The singlet at δ 3.14 ppm was identified as dimethyl sulfone ($DMSO_2$) [34,35]. The middle of the spectrum was complex and dominated by glucose. Signal overlap and weak 2D signal strength meant that hippurate was the only compound that could be clearly identified in the downfield region. Relative chemical shifts and the multiplicity of identified peaks are available in the supplementary material (Table S1).

Unsupervised analysis of the data using PCA showed no obvious clustering of samples by dataset. Results of ANOVA-simultaneous component analysis showed that fixed effects (cow age, herd of origin and days in milk (DIM)) explained only 13.94% of the spectral variation (Table S2). Only the effect of age was statistically significant ($p < 0.05$). This suggests that most spectral variation is due to differences between individual animals.

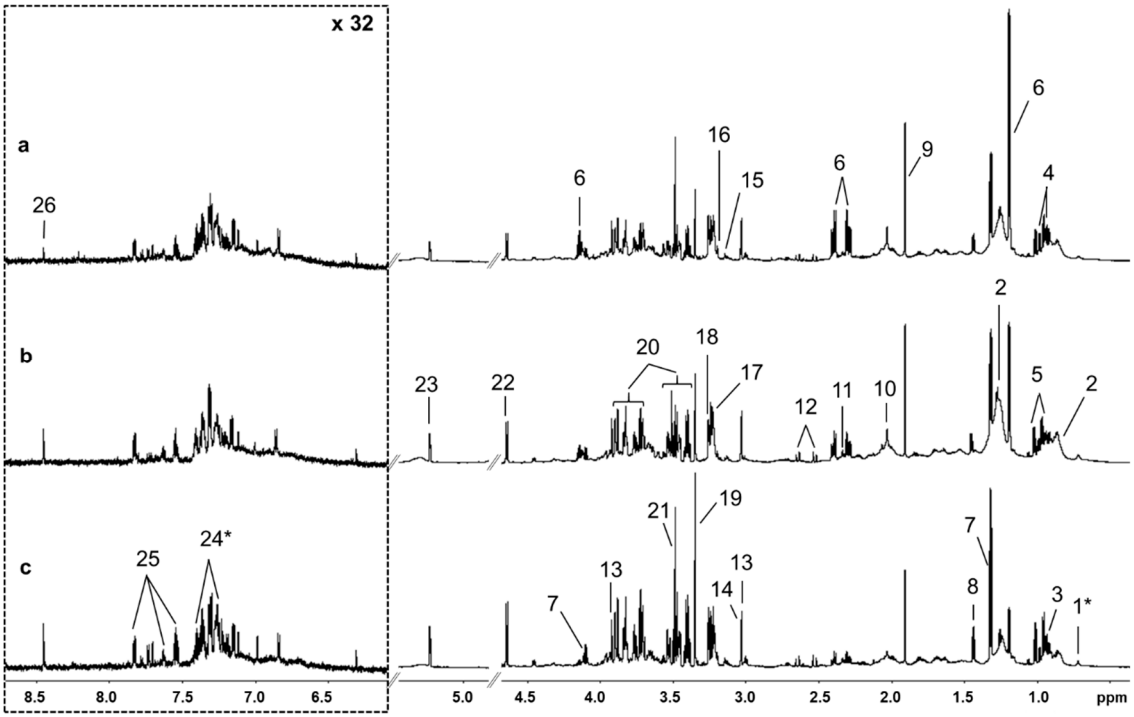
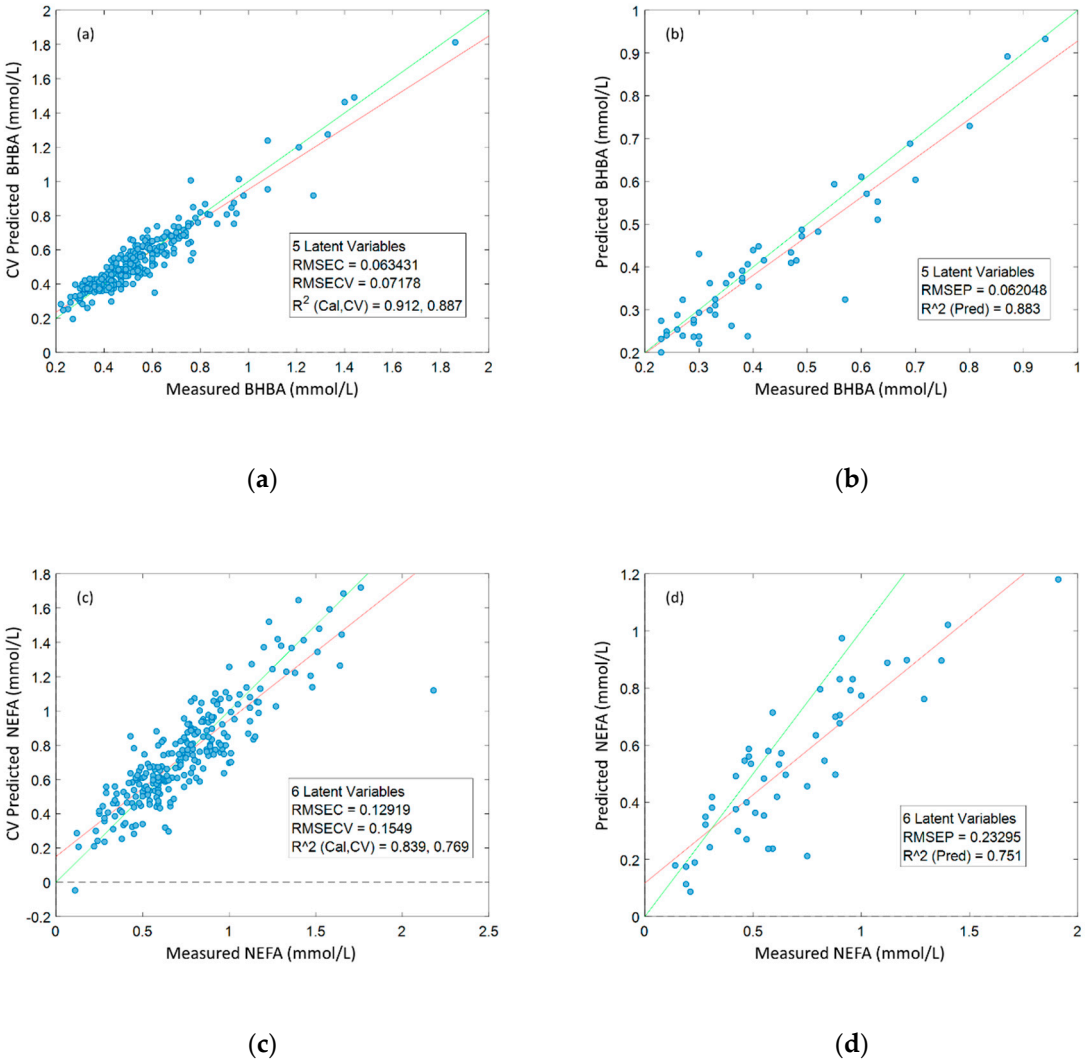


Figure 1. Representative 700 MHz ^{1}H nuclear magnetic resonance spectra of serum samples from early lactation dairy cows with (**a**) elevated β-hydroxybutyrate (BHBA), (**b**) elevated non-esterified fatty acid (NEFA), and (**c**) normal BHBA and NEFA concentrations. Downfield regions were vertically expanded 32 times for clarity. Legend: 1, cholate; 2, very low density lipoprotein/low density lipoprotein; 3, leucine; 4, isoleucine; 5, valine; 6, β-hydroxybutyrate; 7, lactate; 8, alanine; 9, acetate; 10, N-acetyl glycoprotein; 11, pyruvate; 12, citrate; 13, creatine; 14, creatine phosphate; 15, dimethyl sulfone ($DMSO_2$); 16, choline; 17, phosphocholine; 18, betaine; 19, methanol; 20, glucose; 21, glycine; 22, β-Glu; 23, α-Glu; 24, 3-phenyllactate; 25, hippurate; 26; formate. * = tentative identification.

2.3. Accuracy and Robustness of Prediction Models

The robustness of the orthogonal partial least squares (OPLS) regression models built using data from Dataset 1 was assessed using (1) 10-fold cross-validation (Figure 2a,c) and (2) external validation with data from Dataset 2 (Figure 2b,d). Prediction accuracies derived from external validation were high for BHBA (R^2 = 0.88), and moderately high for NEFA (R^2 = 0.75). BHBA models were remarkably robust, with external validation R^2 and RMSE results almost identical to cross-validation results. Models predicting serum NEFA concentration were less accurate than those predicting BHBA (NRMSE 0.32 and 0.50, respectively), but external validation results indicated that these models were still quite robust. *p*-values derived from permutation testing were < 0.001 for all models, indicating that models were not over-fitted.



(**a**) (**b**)

(**c**) (**d**)

Figure 2. Accuracy of orthogonal partial least squares (OPLS) regression models predicting serum β-hydroxybutyrate (BHBA) and non-esterified fatty acid (NEFA) concentrations from ^{1}H NMR spectra, built using data from Dataset 1 (N = 248); (**a**) 10-fold cross-validation (CV)-predicted BHBA vs. measured BHBA; (**b**) external validation (N = 50)-predicted BHBA vs. actual BHBA; (**c**) CV-predicted NEFA vs. measured NEFA; (**d**) external validation-predicted NEFA vs measured NEFA.

2.4. Metabolomic Fingerprints of BHBA and NEFA

The metabolomic fingerprints associated with BHBA and NEFA were investigated using OPLS regression. Larger scores on the first latent variable (LV1) correspond to higher concentrations of both BHBA and NEFA (Figure 3a,b). LV1 loadings plots were used to identify which spectral features contributed most to the variation in the reference biomarker concentrations [36] (Figure 3c,d). Spectral features with positive loadings correspond to metabolites that are positively correlated with reference biomarker concentrations, and vice-versa. Peaks with a variable importance of projection (VIP) score greater than one were considered statistically significant [37] (Figure S2).

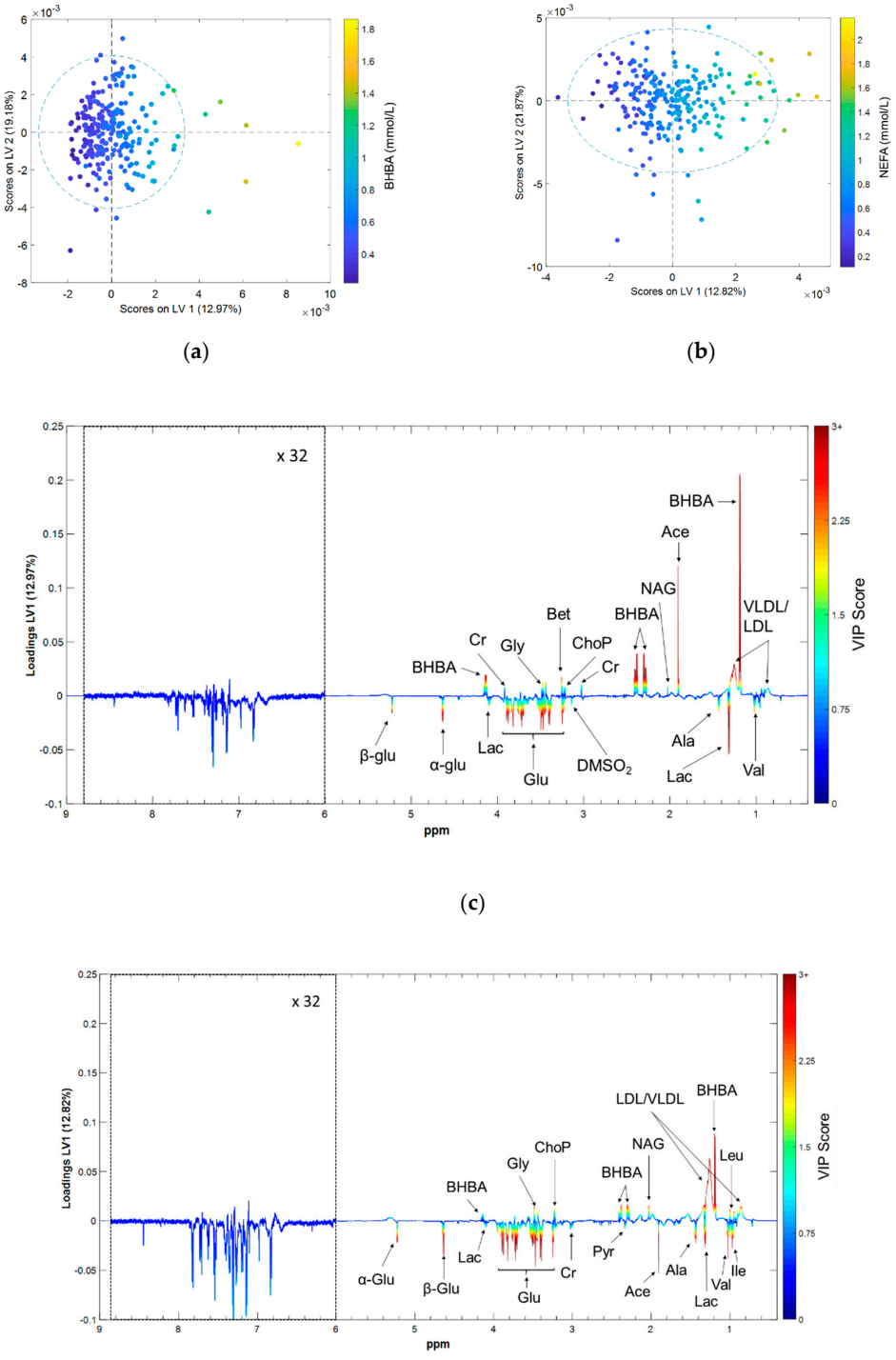


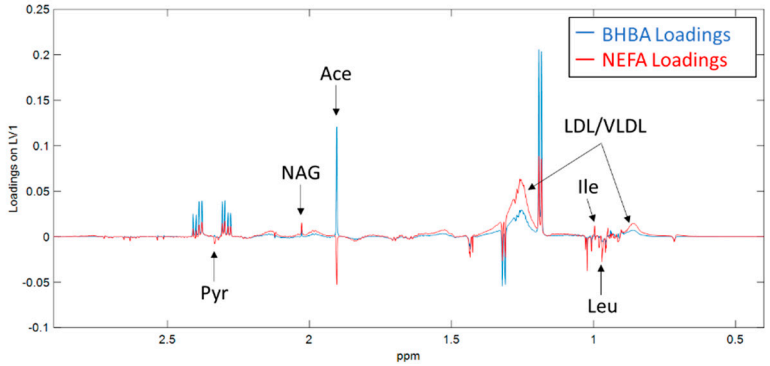
Figure 3. Results of the orthogonal partial least squares (OPLS) regression models predicting serum BHBA and NEFA concentrations from ^{1}H NMR spectra; (**a**) First latent variable (LV1) vs. second latent variable (LV2) scores for the BHBA prediction model; (**b**) LV1 vs. LV2 scores for the NEFA prediction model; (**c**) LV1 loadings for the BHBA prediction model; (**d**) LV1 loadings for the NEFA prediction model. Scores plots color-coded by reference biomarker concentration, loadings plots by VIP score. α-Glu = α glucose, β-Glu = β glucose, Ace = acetate, Ala = alanine, Bet = betaine, BHBA = β hydroxybutyrate, Cr = creatine, $DMSO_2$ = dimethyl sulfone, Glu = glucose, Gly = glycine, Ile = isoleucine, Lac = lactate, Leu = leucine, NAG = N-acetyl glycoprotein, ChoP = phosphocholine, Pyr = pyruvate, Val = valine, LDL = low density lipoprotein; VLDL = very low density lipoprotein.

2.4.1. Commonalities in the Metabolomic Fingerprints of BHBA and NEFA

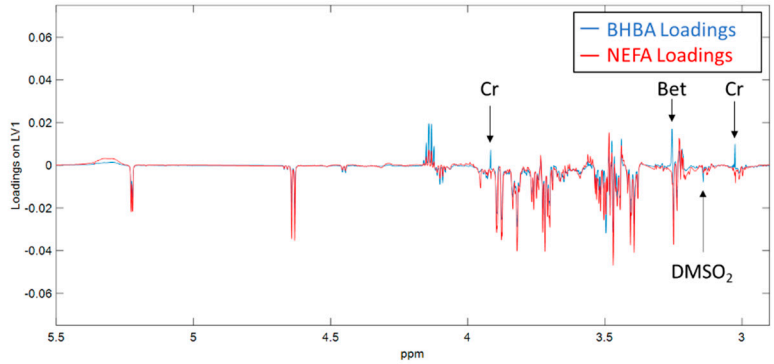
The results of this study show that several metabolites showed similar co-variances with both BHBA and NEFA concentrations. The largest effect we observed was from peaks assigned to glucose, which were negatively correlated with both biomarkers. Other metabolites with common co-variances included lactate, valine and alanine (negatively correlated), and glycine and phosphocholine (positively correlated). Spectral regions attributed to lipoproteins (LDL and VLDL) and glycoproteins were positively correlated with both BHBA and NEFA concentrations.

2.4.2. Differences between the Metabolomic Fingerprints of BHBA and NEFA

Figure 4 highlights the differences we observed between the metabolomic fingerprints of BHBA and NEFA. Acetate and creatine were positively correlated with BHBA, and negatively correlated with NEFA. A small number of metabolites showed significant co-variance with only one of the biomarkers. BHBA concentration was positively correlated with betaine, and negatively correlated with dimethyl sulfone ($DMSO_2$), while NEFA concentration was positively correlated with isoleucine and negatively correlated with leucine.



(**a**)



(**b**)

Figure 4. Loadings on the first latent variable (LV1) derived from orthogonal partial least squares (OPLS) regression of ^{1}H NMR spectra against serum BHBA (blue) and NEFA (red) concentrations in early lactation dairy cows. Spectral regions between (**a**) δ 0.2 ppm to 2.9 ppm and (b) δ 2.9 ppm to 5.5 ppm are shown. Figure (**b**) has been for clarity purposes. Ace = acetate, Bet = betaine, ChoP = Phosphocholine, Cr = creatine, $DMSO_2$ = dimethyl sulfone, Ile = isoleucine, Leu = leucine, LDL/VLDL = low/very low-density lipoprotein, NAG = N-acetyl glycoprotein, Pyr = pyruvate.

3. Discussion

3.1. Similarities between BHBA and NEFA

Not surprisingly, many of the metabolites identified as having common co-variance with both BHBA and NEFA concentrations are involved in hepatic energy metabolism. These relationships are summarized in Figure 5. Most obvious was the negative relationship between both biomarkers and glucose. Hypoglycaemia has been widely reported in early lactation dairy cows due to the massive demand for glucose for lactogenesis [3,38]. More recently, NMR metabolomics studies have identified serum glucose concentration as being (1) directly correlated to energy balance (r = 0.84) [39], and (2) lower in cows with clinical and subclinical ketosis [25] and fatty liver [27] when compared to healthy controls. Our results offer further evidence of the pivotal role glucose plays in the early lactation metabolic health in dairy cows.

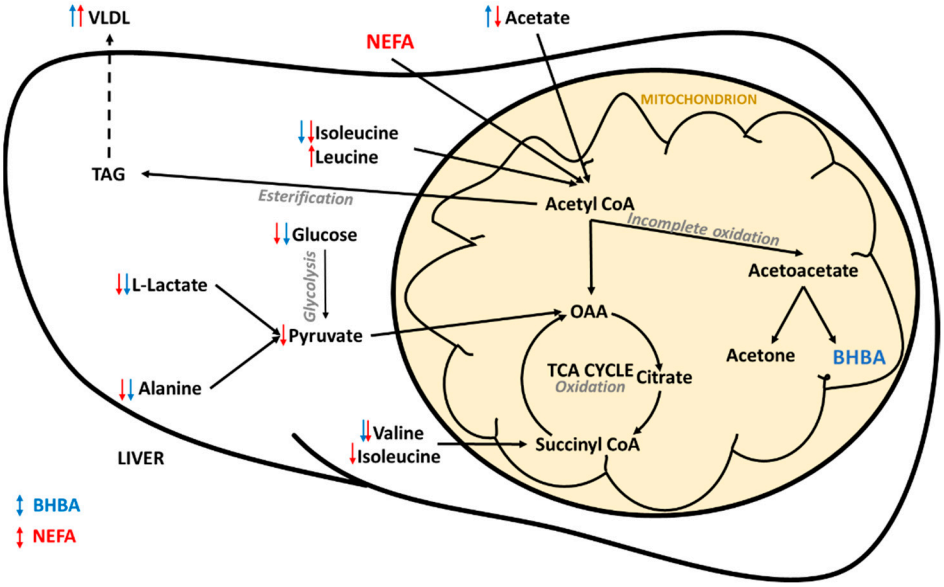


Figure 5. Summary of hepatic energy metabolism in early lactation dairy cows. Arrows indicate the direction of the relationship between the metabolites and the reference BHBA (blue) and non-esterified fatty acid (NEFA) (red) concentrations. BHBA = β-hydroxybutyrate; OAA = oxaloacetate; TAG = triglyceride, TCA = tricarboxylic acid, VLDL = very low density lipoprotein.

Lactate and alanine, important gluconeogenic substrates in ruminants [40,41], were also negatively associated with both BHBA and NEFA, as was valine (another gluconeogenic amino acid). Interestingly, Xu et al. [39] found no correlation between calculated energy balance in early lactation dairy cows and the concentrations of any of the branched-chain amino acids or lactate. Conversely, when compared to healthy controls, cows with fatty liver and displaced abomasa have been shown to have lower serum alanine concentrations [27,29], and cows with ketosis have lower lactate and alanine concentrations [25,42]. This suggests that alterations in glucogenic precursors, in particular lactate and alanine, are indicative of a perturbed metabolism, not simply negative energy balance. We previously showed that lactate concentration in pasture-fed dairy cows is heavily influenced by herd-specific management factors [43], and as such may not be heavily influenced by genetic factors. Alanine has been shown to be the most important glucogenic amino acid, and the most important gluconeogenic precursor after lactate and propionate, in dairy cows [41]. Therefore, genetic selection for cows with higher serum concentrations of alanine in early lactation may help to increase endogenous glucose supply.

Spectral features attributed to VLDL and LDL were positively correlated with the concentrations of both BHBA and NEFA. These results need to be interpreted with caution as the methanol extraction

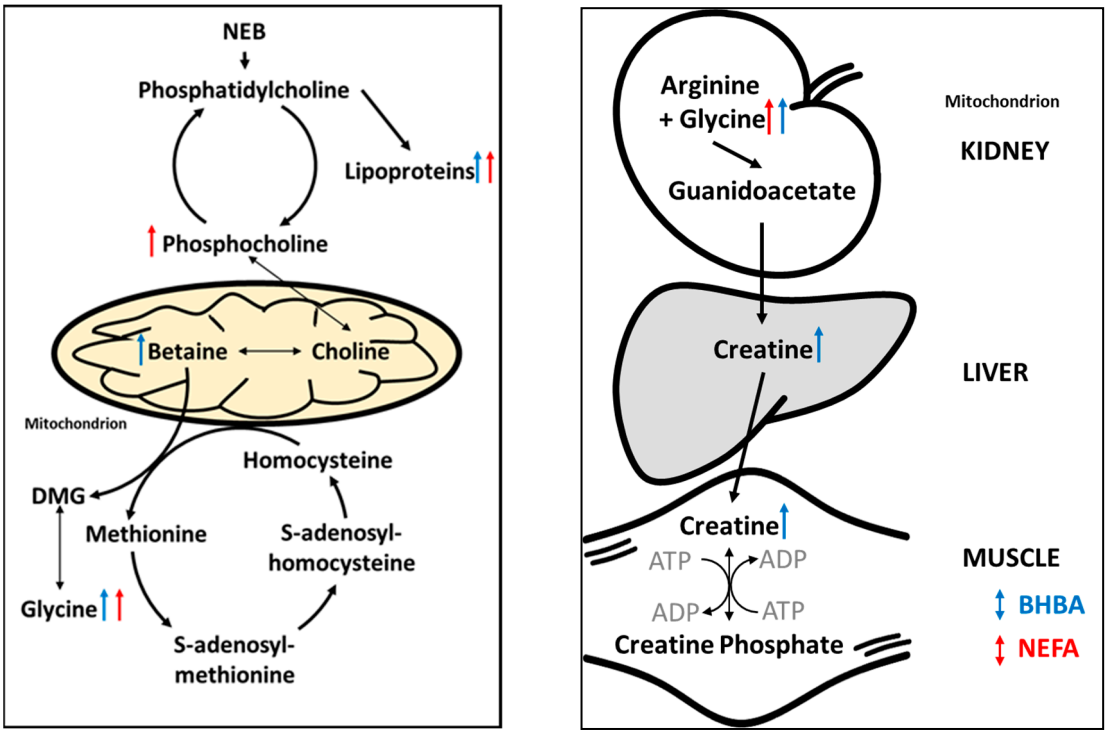
used in this study removed much of the protein from the samples and may have introduced experimental artefacts. Interestingly, ^{1}H NMR spectroscopy has recently been shown capable of providing high-throughput and accurate quantification of lipoprotein subclasses in human serum and plasma samples [32,44]. It is important to note that these protocols used different pulse sequences and involved the dilution of plasma/serum in a deuterated water/phosphate buffer solution without any metabolite extraction, such as the one used in our study. The findings of these studies cannot, therefore, be applied directly to our results. However, lipoprotein metabolism is central to early lactation health in dairy cows, and impaired VLDL production in the liver can result in hepatic triglyceride (TAG) accumulation (Figure 4) and the development of fatty liver [45]. Dyslipoproteinaemia is also an important feature of metabolic syndrome in humans, and the quantification of lipoprotein subclasses is considered critical to the better understanding of this disease [44]. We believe that the investigation of serum lipoproteins using ^{1}H NMR spectroscopy holds great promise in the research of early lactation metabolic health in dairy cows, and we plan to validate the aforementioned protocols on bovine serum and plasma samples.

The region of the spectrum associated with glycoproteins was also significantly positively correlated with both NEFA and BHBA concentrations. Glycoproteins are acute phase proteins which can be used as indicators of inflammation in cattle [46]. In dairy cattle, increased serum NEFA concentrations in early lactation are associated with uncontrolled inflammation, and this inflammatory dysfunction is hypothesized to be a central link between metabolic and infectious disorders [14,47]. ^{1}H NMR spectroscopy is showing promise for the quantification of glycoprotein A (GlcA) in human research into metabolic diseases such obesity, diabetes mellitus and the metabolic syndrome [33]. Given that these syndromes have much in common with early lactation metabolic disease in dairy cows (e.g., insulin resistance), we believe that further research into GlcA as a biomarker for early lactation health is warranted. Overall, our results offer further evidence that inflammation plays an important role in early lactation metabolic health of dairy cows.

Glycine was positively correlated with the concentrations of both BHBA and NEFA. Metabolomics studies comparing healthy and ketotic dairy cows have reported (1) no change in glycine concentrations [25], (2) increased glycine concentrations in cows with sub-clinical ketosis [26], (3) increased glycine concentrations in cows with clinical ketosis [48] and (4) decreased glycine concentrations in cows with clinical ketosis [26] and fatty liver [49]. Glycine concentration has also been shown to increase in response to lipolysis [50]. These differing results suggest that changes in glycine concentration may be dependent on the severity of the metabolic disorder (i.e., increased in mild cases, and decreased in more severe cases). Most interesting are the findings of a recent metabolomics study that showed that glycine concentrations in plasma and milk were strongly negatively correlated with energy balance in early lactation dairy cows ($r = -0.80$ and $r = -0.79$, respectively) [39]. The authors of this study hypothesized that this relationship was due to an increase in one-carbon or methyl donor metabolism, specifically an increase in the conversion of choline to glycine. Given that all cows in our study were clinically healthy, our results are consistent with glycine being an indicator of negative energy balance, lipolysis, and/or sub-clinical ketosis. Further work is required to better understand the role of glycine metabolism in clinical metabolic disease.

The positive correlations between phosphocholine and both BHBA and NEFA concentrations, and between betaine and BHBA concentration, are consistent with an increase in methyl donor metabolism in cows experiencing negative energy balance. Methyl donor metabolism and nutrition are receiving a great deal of attention in dairy science due to links with early-lactation cow health (including fatty liver), milk production and immune function [51]. Betaine, phosphocholine and glycine are intermediates in several important one-carbon metabolic pathways including the folate and methionine cycles, and the cytidine diphosphate (CDP)–choline pathway [51] (Figure 6a). The positive correlation between NEFA and phosphocholine may be due to increased breakdown of phosphatidylcholine (Figure 6a). This is consistent with the findings of Imhasly et al. [52] who showed that serum concentrations of lyso-phosphatidylcholines and phosphatidylcholines increase in response to negative energy balance

in post-partum dairy cows. The positive association observed between betaine and BHBA could be due to increased oxidation of choline. A detailed description of these pathways is beyond the scope of this study, however our results suggest that methyl donor metabolism has an important influence on both BHBA and NEFA concentrations in early-lactation dairy cows.



(**a**) (**b**)

Figure 6. Summary of (**a**) phospholipid and one-carbon/methyl donor metabolism [53,54], and (**b**) creatine metabolism in early lactation dairy cows. Arrows indicate the direction of the relationship between the metabolites identified using untargeted ^{1}H NMR metabolomics, and reference BHBA (blue) and non-esterified fatty acids (NEFA) (red) concentrations. ADP = adenosine diphosphate; ATP = adenosine triphosphate; DMG = dimethylglycine; NEB = negative energy balance.

3.2. Differences between BHBA and NEFA

Despite many similarities, we observed some significant differences between the metabolomic fingerprints of BHBA and NEFA. Most obvious was the difference in the direction of correlation between acetate and the two biomarkers. Acetate is a volatile fatty acid produced by microbial fermentation of feedstuffs in the rumen, and is an important energy source [55] (via oxidation or the partial oxidation of acetyl-CoA in the liver) and substrate for de novo milk fat synthesis [56] in cows. The negative relationship we observed between acetate and NEFA is consistent with the findings of Bielak et al. [57], who reported a negative correlation (r = 0.44) between plasma NEFA and acetate concentrations in early lactation dairy cows, possibly due to the down-regulation of the active transport of acetate across the rumen wall. The positive association between acetate and BHBA is consistent with previously discussed metabolomic studies of ketosis and fatty liver [25,27]. These results suggest that differences in acetate metabolism may help to explain the weak correlation between serum BHBA and NEFA concentrations in early lactation dairy cows.

The positive correlation between creatine and BHBA concentration is consistent with previous reports that creatine is a potentially useful biomarker of ketosis and severe energy deficiency in dairy cows [25,26,39]. Creatine is an important intermediate in energy metabolism, and this result may represent increased breakdown of creatine phosphate in skeletal muscle and the release of high-energy

phosphate for the conversion of adenosine diphosphate (ADP) to adenosine triphosphate (ATP) (Figure 6b). Interestingly, creatine concentration was negatively correlated with NEFA concentration (albeit weakly and non-significantly (VIP < 1)). That mobilization of energy from skeletal muscle is a feature of the BHBA metabolomic fingerprint, but not that of NEFA, suggests that elevated BHBA concentrations are indicative of a more severe energy deficiency than are elevated NEFA concentrations. However, the ability to rapidly mobilize energy from skeletal muscle may be advantageous to early-lactation dairy cows, and we believe the role of creatine metabolism in transition cow health warrants further investigation. We therefore plan to undertake genome-wide association studies to better understand the genetic relationships between hepatic and skeletal muscle energy metabolism.

The significant negative correlation between $DMSO_2$ and BHBA concentration was an interesting finding of this study. $DMSO_2$ concentration in the milk and rumen fluid of dairy cows has been shown to vary according to feeding system; higher in pasture-fed cows managed outdoors than in cows fed a total mixed ration indoors [58]. Maher et al. [59] showed that the concentrations of $DMSO_2$ in milk and plasma are highly correlated (r = 0.69), so serum $DMSO_2$ may also be an indicator of pasture intake. Given that all animals in this experiment were fed pasture, the negative association we observed between $DMSO_2$ and BHBA concentration may indicate that hyperketonemic cows are consuming less feed.

4. Materials and Methods

All procedures undertaken in this study were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (National Health and Medical Research Council, 2013). Approval to proceed was granted by the Agricultural Research and Extension Animal Ethics Committee of the Department of Jobs, Precincts and Regions Animal Ethics Committee (DJPR, 475 Mickleham Road, Attwood, Victoria 3049, Australia), and the Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIPWE Animal Biosecurity and Welfare Branch, 13 St Johns Avenue, New Town, Tasmania 7008, Australia). AEC project approval code 2017-05.

4.1. Animals and Datasets

A total of 298 Holstein-Friesian cows were used in this experiment. The calibration dataset (Dataset 1) was collected between August and September 2017 from 248 animals located at the Ellinbank Dairy Research Centre, Ellinbank, Victoria, Australia. An independent validation dataset (Dataset 2) was collected in September 2018, from 50 cows located on a commercial dairy farm in Smithton, Tasmania, Australia. All cows were clinically healthy, and had been calved for between 4 and 30 days at the time of sampling. Feeding systems on Australian dairy farms are diverse but can be classified into 5 main feeding systems [60]. Both farms operated under feeding system 2; grazed pasture plus moderate to high level concentrate feeding (>1.0 tonne of concentrate fed in the milking parlour per cow per year).

4.2. Blood Sample Collection and Reference Biomarker Measurements

A single serum sample was taken from each cow immediately after morning milking (approximately 07:00) according to the protocol described in Luke et al. [43]. Cows were fed their concentrate ration as soon as they entered the milking parlour, meaning that samples were collected approximately 10 min after grain feeding.

An aliquot of each serum sample was shipped on ice to Regional Laboratory Services (Benalla, Victoria, Australia) for BHBA and NEFA analyses. Colorimetric assays were performed using a Kone 20 XT clinical chemistry analyser (Thermo Fisher Scientific, Waltham, MA, USA); an enzymatic kinetic assay for BHBA (McMurray et al., 1984) and enzymatic end point assay for NEFA (Randox Laboratories, Crumlin, UK). The uncertainty of measurement (at a 95% confidence level) was ±0.060 mmol/L at 0.85 mmol/L for BHBA, and ±0.031 mM at 1.45 mM for NEFA. A second aliquot was stored at −20 °C until processing for NMR spectroscopy.

4.3. Sample Preparation for NMR Spectroscopy

Details of the sample preparation and metabolite extraction protocols used in this study can be found in Luke et al. [43]. Briefly, 300 µL of serum was (1) mixed with 600 µL of methanol, (2) vortexed, (3) incubated at −20 °C for 20 min, and (4) centrifuged at 11,360× *g* at 21 °C for 30 min to pellet proteins. A 600 µL aliquot of the supernatant was then transferred to a clean 2 mL microcentrifuge tube, dried under vacuum at 21 °C overnight using a SpeedVac Savant SPD 2010 Concentrator (Thermo Fisher Scientific, Waltham, MA, USA) then reconstituted in a D_2O phosphate buffer solution (100 mM K_2HPO_4) containing 0.25 mM DSS-d6 as an internal standard. A 550 µL aliquot of reconstituted extract was transferred to a 5 mm NMR tube for analysis.

4.4. ^{1}H NMR Data Acquisition and Pre-Processing

One-dimensional proton spectra were acquired using a Bruker Ascend 700 MHz spectrometer equipped with cryoprobe and SampleJet automatic sample changer (Bruker Biospin, Rheinstetten, Germany). A Bruker noesypr1d pulse sequence was used over a −0.76–10.32 ppm spectral range with the following acquisition parameters; (1) a temperature of 298 K, (2) 256 scans after eight dummy scans (3) acquisition time per increment of 2.11 s, and (4) relaxation delay (D1) of 2.00 s. This resulted in 32,768 data points. A line broadening of 0.3 Hz was applied to all spectra prior to Fourier transformation. Spectra were manually phased, baseline corrected and referenced to the internal standard (DSS-d6) at δ 0.00 ppm using the Topspin v.3.6.1 software (Bruker Biospin, Rheinstetten, Germany).

Data pre-processing was performed in MatLab v.R20017b (Mathworks, Natick, MA, USA). Spectra were imported as a matrix of signal intensities using the ProMetab v.1.1 script [61]. Spectral pre-processing involved (1) deletion of the residual water peak region (δ 4.68–5.00 ppm), (2) spectral alignment using the correlation optimized warping algorithm [62], (3) normalization to total signal area (area = 1), (4) deletion of methanol (δ 3.32–3.36 ppm) and DSS-d6 (δ 0.4–0.60 ppm) peak regions, and the non-informative region beyond 9.00 ppm, (5) baseline correction using automatic weighted least squares and (6) mean centering.

4.5. Statistical Analysis

Statistical analysis of experimental metadata was performed in R v3.6.2 [63]. Differences between the 2 datasets were analysed using a paired *t*-test or a Wilcoxon signed-rank test depending on the normality of the data.

Multivariate statistical analyses were performed using the PLS Toolbox v. 8.5.2 (Eigenvector Research Inc., Manson, WA, USA). Preliminary data analysis and outlier detection was performed using an unsupervised PCA. Examination of PC1 vs. PC2 scores plot showed 14 samples from Dataset 1 outside the 95% confidence level ellipse (Figure S1). These samples were individually examined, and a single spectrum with poor water suppression and baseline correction was removed from subsequent analyses. The influences of fixed effects (DIM, age and herd) on spectra were investigated using ANOVA simultaneous component analysis with 1000 permutations [64]. Untargeted metabolomic fingerprinting was done by regressing reference NEFA and BHBA concentrations against ^{1}H NMR spectra using supervised OPLS regression. Variable importance of projection (VIP) scores for the first latent variable were used to identify the most statistically significant peaks in each model. Peaks of interest were identified using the Chenomx NMR suite software v.8.4 (Chenomx Inc., Edmonton, AB, Canada), comparison to the literature, 2D NMR analysis (COSY, gHMBC and gHSQC), and statistical total correlation spectroscopy [65].

OPLS models were constructed using data from Dataset 1. The robustness of models was assessed using (1) cross-validation using a venetian blind technique (10 sample splits with 1 sample per blind) and (2) external validation using data from Dataset 2. The prediction accuracy of OPLS models was assessed using the coefficient of determination (R^2) and root mean square error (RMSE). Normalized

RMSE (NRMSE) values, calculated as external validation RMSE divided by the interquartile interval of the observed data, were used to compare RMSE estimates for NEFA and BHBA predictions. Permutation testing (50 iterations and statistical significance determined using a Wilcoxon signed-rank test) was performed to ensure that models were not over-fitted.

5. Conclusions

In this study we used an untargeted ^{1}H NMR metabolomics approach to investigate the serum metabolic fingerprints of the two most common biomarkers of energy balance in dairy cows, BHBA and NEFA. Our results suggest that while BHBA and NEFA are indicative of similar metabolic states in early-lactation dairy cows, there are significant differences between the two biomarkers. Metabolites with common co-variances were intermediates of energy, phospholipid, and methyl donor metabolism. The most significant differences in the metabolomic fingerprints were related to acetate and creatine metabolism. We also identified several intermediate metabotypes which, when combined with genomic data, will enable further the investigation of the genetic architecture of metabolic health in early lactation dairy cows.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-1989/10/6/247/s1, Table S1: ^{1}H NMR chemical shifts (δ) and multiplicity of metabolites in bovine serum run in deuterated water (D_2O). Clearly observed resonances are indicated in bold text. s, singlet; d, doublet; dd, doublet of a doublet; m, multiplet; t, triplet. The right two columns show the direction of the relationship with serum β-hydroxybutyrate (BHBA) and non-esterified fatty acid (NEFA) concentrations determined by colorimetric assays, Table S2: Results of ANOVA-simultaneous component analysis (ASCA) of ^{1}H NMR spectra of bovine serum (N= 298). Effect describes the relative influence of each variable (herd, age and days in milk (DIM)) on spectra. p-value is derived from permutation testing (1000 iterations), Figure S1: Results of PCA of ^{1}H NMR spectra of serum obtained from 298 dairy cows in early lactation from the Ellinbank research farm (Dataset 1, N = 248) and a commercial dairy farm in Tasmania (Dataset 2, N = 50), Figure S2: VIP scores from OPLS regressions of ^{1}H NMR spectra of serum obtained from 298 dairy cows in early lactation against (a) BHBA concentration and (b) NEFA concentration.

Author Contributions: Conceptualization, T.D.W.L., J.E.P., W.J.W., and S.J.R.; formal analysis, T.D.W.L. and S.J.R.; funding acquisition, J.E.P.; investigation, T.D.W.L. and S.J.R.; methodology, T.D.W.L. and S.J.R.; project administration, J.E.P., W.J.W. and S.J.R.; resources, S.J.R.; supervision, J.E.P., W.J.W., and S.J.R.; visualization, T.D.W.L.; writing—original draft, T.D.W.L.; writing—review and editing, T.D.W.L, J.E.P., W.J.W., and S.J.R. All authors have read and agreed to the published version of the manuscript.

Funding: DairyBio: jointly funded by Dairy Australia (Melbourne, Australia), the Gardiner Foundation (Melbourne, Australia), and Agriculture Victoria (Melbourne, Australia) funded this study and T.D.W.L.'s PhD project.

Acknowledgments: The authors thank Aaron Elkins and Simone Vassiliadis for assistance with ^{1}H NMR protocol and metabolite identification, Di Mapleson, Brigid Ribaux and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance, Erika Oakes and the staff at Datagene (Bundoora, Australia) for their work coordinating this study, and the farmers who took part in this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grant, R.J.; Albright, J.L. Feeding behavior and management factors during the transition period in dairy cattle. *J. Anim. Sci.* **1995**, *73*, 2791–2803. [CrossRef] [PubMed]
2. Bell, A.W. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. *J. Anim. Sci.* **1995**, *73*, 2804–2819. [CrossRef] [PubMed]
3. Bauman, D.E.; Bruce Currie, W. Partitioning of Nutrients During Pregnancy and Lactation: A Review of Mechanisms Involving Homeostasis and Homeorhesis. *J. Dairy Sci.* **1980**, *63*, 1514–1529. [CrossRef]
4. Herdt, T.H. Ruminant Adaptation to Negative Energy Balance: Influences on the Etiology of Ketosis and Fatty Liver. *Vet. Clin. N. Am. Food Anim. Pract.* **2000**, *16*, 215–230. [CrossRef]
5. Compton, C.; McDougall, S.; Young, L.; Bryan, M. Prevalence of subclinical ketosis in mainly pasture-grazed dairy cows in New Zealand in early lactation. *N. Z. Vet. J.* **2013**, *62*, 30–37. [CrossRef]
6. McArt, J.A.A.; Nydam, D.V.; Oetzel, G.R. Epidemiology of subclinical ketosis in early lactation dairy cattle. *J. Dairy Sci.* **2012**, *95*, 5056–5066. [CrossRef]

7. Cameron, R.E.B.; Dyk, P.B.; Herdt, T.H.; Kaneene, J.B.; Miller, R.; Bucholtz, H.F.; Liesman, J.S.; Vandehaar, M.J.; Emery, R.S. Dry Cow Diet, Management, and Energy Balance as Risk Factors for Displaced Abomasum in High Producing Dairy Herds. *J. Dairy Sci.* **1998**, *81*, 132–139. [CrossRef]
8. Drackley, J.K. Biology of Dairy Cows During the Transition Period: The Final Frontier? *J. Dairy Sci.* **1999**, *82*, 2259–2273. [CrossRef]
9. McArt, J.A.A.; Nydam, D.V.; Overton, M.W. Hyperketonemia in early lactation dairy cattle: A deterministic estimate of component and total cost per case. *J. Dairy Sci.* **2015**, *98*, 2043–2054. [CrossRef]
10. Chapinal, N.; Leblanc, S.J.; Carson, M.E.; Leslie, K.E.; Godden, S.; Capel, M.; Santos, J.E.; Overton, M.W.; Duffield, T.F. Herd-level association of serum metabolites in the transition period with disease, milk production, and early lactation reproductive performance. *J. Dairy Sci.* **2012**, *95*, 5676–5682. [CrossRef]
11. Ospina, P.; Nydam, D.; Stokol, T.; Overton, T. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. *J. Dairy Sci.* **2010**, *93*, 1596–1601. [CrossRef] [PubMed]
12. Adewuyi, A.A.; Gruys, E.; Van Eerdenburg, F.J.C.M. Non esterified fatty acids (NEFA) in dairy cattle. A review. *Vet. Q.* **2005**, *27*, 117–126. [CrossRef] [PubMed]
13. Laffel, L. Ketone bodies: A review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes Metab. Res. Rev.* **1999**, *15*, 412–426. [CrossRef]
14. Sordillo, L.M.; Raphael, W. Significance of Metabolic Stress, Lipid Mobilization, and Inflammation on Transition Cow Disorders. *Vet. Clin. N. Am. Food Anim. Pract.* **2013**, *29*, 267–278. [CrossRef] [PubMed]
15. Ospina, P.A.; Nydam, D.V.; Stokol, T.; Overton, T.R. Evaluation of nonesterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. *J. Dairy Sci.* **2010**, *93*, 546–554. [CrossRef]
16. McArt, J.A.A.; Nydam, D.V.; Oetzel, G.R.; Overton, T.R.; Ospina, P.A. Elevated non-esterified fatty acids and β-hydroxybutyrate and their association with transition dairy cow performance. *Vet. J.* **2013**, *198*, 560–570. [CrossRef] [PubMed]
17. Compton, C.; Young, L.; McDougall, S. Subclinical ketosis in post-partum dairy cows fed a predominantly pasture-based diet: Defining cut-points for diagnosis using concentrations of beta-hydroxybutyrate in blood and determining prevalence. *N. Z. Vet. J.* **2014**, *63*, 241–248. [CrossRef]
18. LeBlanc, S.J.; Leslie, K.E.; Duffield, T.F. Metabolic predictors of displaced abomasum in dairy cattle. *J. Dairy Sci.* **2005**, *88*, 159–170. [CrossRef]
19. Duffield, T.F.; Lissemore, K.D.; McBride, B.W.; Leslie, K.E. Impact of hyperketonemia in early lactation dairy cows on health and production. *J. Dairy Sci.* **2009**, *92*, 571–580. [CrossRef]
20. Seifi, H.A.; Leblanc, S.J.; Leslie, K.E.; Duffield, T.F. Metabolic predictors of post-partum disease and culling risk in dairy cattle. *Vet. J.* **2011**, *188*, 216–220. [CrossRef]
21. Luke, T.D.W.; Rochfort, S.; Wales, W.J.; Bonfatti, V.; Marett, L.; Pryce, J.E. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. *J. Dairy Sci.* **2019**, *102*, 1747–1760. [CrossRef] [PubMed]
22. McCarthy, M.M.; Mann, S.; Nydam, D.V.; Overton, T.R.; McArt, J.A.A. Short communication: Concentrations of nonesterified fatty acids and β-hydroxybutyrate in dairy cows are not well correlated during the transition period. *J. Dairy Sci.* **2015**, *98*, 6284–6290. [CrossRef]
23. Luke, T.D.W.; Nguyen, T.T.T.; Rochfort, S.; Wales, W.J.; Richardson, C.M.; Abdelsayed, M.; Pryce, J.E. Genomic prediction of serum biomarkers of health in early lactation. *J. Dairy Sci.* **2019**, *102*, 11142–11152. [CrossRef]
24. Nicholson, J.K.; Lindon, J.C.; Holmes, E. 'Metabonomics': Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29*, 1181–1189. [CrossRef] [PubMed]
25. Sun, L.W.; Zhang, H.Y.; Wu, L.; Shu, S.; Xia, C.; Xu, C.; Zheng, J.S. 1H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. *J. Dairy Sci.* **2014**, *97*, 1552–1562. [CrossRef] [PubMed]
26. Wang, Y.; Gao, Y.; Xia, C.; Zhang, H.; Qian, W.; Cao, Y. Pathway analysis of plasma different metabolites for dairy cow ketosis. *Ital. J. Anim. Sci.* **2016**, *15*, 545–551. [CrossRef]
27. Xu, C.; Sun, L.-W.; Xia, C.; Zhang, H.-Y.; Zheng, J.-S.; Wang, J.-S. (1)H-Nuclear Magnetic Resonance-Based Plasma Metabolic Profiling of Dairy Cows with Fatty Liver. *Asian-Australas. J. Anim. Sci.* **2016**, *29*, 219–229. [CrossRef]

28. Sun, Y.; Xu, C.; Li, C.; Xia, C.; Xu, C.; Wu, L.; Zhang, H. Characterization of the serum metabolic profile of dairy cows with milk fever using 1H-NMR spectroscopy. *Vet. Q.* **2014**, *34*, 159–163. [CrossRef]
29. Basoglu, A.; Baspinar, N.; Tenori, L.; Licari, C.; Gulersoy, E. Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum. *Vet. Q.* **2020**, *40*, 1–15. [CrossRef]
30. Daetwyler, H.D.; Xiang, R.; Yuan, Z.; Bolormaa, S.; Vander Jagt, C.J.; Hayes, B.J.; van der Werf, J.H.J.; Pryce, J.E.; Chamberlain, A.J.; Macleod, I.M.; et al. Integration of functional genomics and phenomics into genomic prediction raises its accuracy in sheep and dairy cattle. In Proceedings of the Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW, Australia, 27 October–1 November 2019; pp. 11–14.
31. Xiang, R.; Berg, I.v.D.; Macleod, I.M.; Hayes, B.J.; Prowse-Wilkins, C.P.; Wang, M.; Bolormaa, S.; Liu, Z.; Rochfort, S.J.; Reich, C.M.; et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19398–19408. [CrossRef]
32. Aru, V.; Lam, C.; Khakimov, B.; Hoefsloot, H.C.J.; Zwanenburg, G.; Lind, M.V.; Schäfer, H.; van Duynhoven, J.; Jacobs, D.M.; Smilde, A.K.; et al. Quantification of lipoprotein profiles by nuclear magnetic resonance spectroscopy and multivariate data analysis. *Trends Anal. Chem* **2017**, *94*, 210–219. [CrossRef]
33. Fuertes-Martin, R.; Correig, X.; Vallve, J.C.; Amigo, N. Human Serum/Plasma Glycoprotein Analysis by (1)H-NMR, an Emerging Method of Inflammatory Assessment. *J. Clin. Med.* **2020**, *9*, 354. [CrossRef] [PubMed]
34. Engelke, U.F.; Tangerman, A.; Willemsen, M.A.; Moskau, D.; Loss, S.; Mudd, S.H.; Wevers, R.A. Dimethyl sulfone in human cerebrospinal fluid and blood plasma confirmed by one-dimensional (1)H and two-dimensional (1)H-(13)C NMR. *NMR Biomed.* **2005**, *18*, 331–336. [CrossRef]
35. Maher, A.D.; Coles, C.; White, J.; Bateman, J.F.; Fuller, E.S.; Burkhardt, D.; Little, C.B.; Cake, M.; Read, R.; McDonagh, M.B.; et al. 1H NMR spectroscopy of serum reveals unique metabolic fingerprints associated with subtypes of surgically induced osteoarthritis in sheep. *J. Proteome Res.* **2012**, *11*, 4261–4268. [CrossRef] [PubMed]
36. Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2013**, *1*, 92–107. [CrossRef]
37. Chong, I.-G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112. [CrossRef]
38. Annison, E.F.; Bickerstaffe, R.; Linzell, J.L. Glucose and fatty acid metabolism in cows producing milk of low fat content. *J. Agric. Sci* **1974**, *82*, 87–95. [CrossRef]
39. Xu, W.; Vervoort, J.; Saccenti, E.; Kemp, B.; van Hoeij, R.J.; van Knegsel, A.T.M. Relationship between energy balance and metabolic profiles in plasma and milk of dairy cows in early lactation. *J. Dairy Sci.* **2020**, *103*, 4795–4805. [CrossRef]
40. Aschenbach, J.R.; Kristensen, N.B.; Donkin, S.S.; Hammon, H.M.; Penner, G.B. Gluconeogenesis in dairy cows: The secret of making sweet milk from sour dough. *IUBMB Life* **2010**, *62*, 869–877. [CrossRef]
41. Drackley, J.K.; Overton, T.R.; Douglas, G.N. Adaptations of Glucose and Long-Chain Fatty Acid Metabolism in Liver of Dairy Cows during the Periparturient Period. *J. Dairy Sci.* **2001**, *84*, E100–E112. [CrossRef]
42. Zhang, H.; Wu, L.; Xu, C.; Xia, C.; Sun, L.; Shu, S. Plasma metabolomic profiling of dairy cows affected with ketosis using gas chromatography/mass spectrometry. *BMC Vet. Res.* **2013**, *9*, 186. [CrossRef] [PubMed]
43. Luke, T.D.W.; Pryce, J.E.; Elkins, A.C.; Wales, W.J.; Rochfort, S.J. Use of Large and Diverse Datasets for ^{1}H NMR Serum Metabolic Profiling of Early Lactation Dairy Cows. *Metabolites* **2020**, *10*, 180. [CrossRef] [PubMed]
44. Jimenez, B.; Holmes, E.; Heude, C.; Tolson, R.F.; Harvey, N.; Lodge, S.L.; Chetwynd, A.J.; Cannet, C.; Fang, F.; Pearce, J.T.M.; et al. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by (1)H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* **2018**, *90*, 11962–11971. [CrossRef]
45. Grummer, R.R. Etiology of Lipid-Related Metabolic Disorders in Periparturient Dairy Cows. *J. Dairy Sci.* **1993**, *76*, 3882–3896. [CrossRef]
46. Horadagoda, N.U.; Knox, K.M.G.; Gibbs, H.A.; Reid, S.W.J.; Horadagoda, A.; Edwards, S.E.R.; Eckersall, P.D. Acute phase proteins in cattle: Discrimination between acute and chronic inflammation. *Vet. Rec.* **1999**, *144*, 437–441. [CrossRef]

47. Sordillo, L.M. Nutritional strategies to optimize dairy cattle immunity. *J. Dairy Sci.* **2016**, *99*, 4967–4982. [CrossRef]
48. Li, Y.; Xu, C.; Xia, C.; Zhang, H.; Sun, L.; Gao, Y. Plasma metabolic profiling of dairy cows affected with clinical ketosis using LC/MS technology. *Vet. Q.* **2014**, *34*, 152–158. [CrossRef]
49. Imhasly, S.; Naegeli, H.; Baumann, S.; Von Bergen, M.; Luch, A.; Jungnickel, H.; Potratz, S.; Gerspach, C. Metabolomic biomarkers correlating with hepatic lipidosis in dairy cows. *BMC Vet. Res.* **2014**, *10*, 122. [CrossRef]
50. Humer, E.; Khol-Parisini, A.; Metzler-Zebeli, B.U.; Gruber, L.; Zebeli, Q. Alterations of the Lipid Metabolome in Dairy Cows Experiencing Excessive Lipolysis Early Postpartum. *PLoS ONE* **2016**, *11*, e0158633. [CrossRef]
51. McFadden, J.W.; Girard, C.L.; Tao, S.; Zhou, Z.; Bernard, J.K.; Duplessis, M.; White, H.M. Symposium review: One-carbon metabolism and methyl donor nutrition in the dairy cow. *J. Dairy Sci.* **2020**. [CrossRef]
52. Imhasly, S.; Bieli, C.; Naegeli, H.; Nyström, L.; Ruetten, M.; Gerspach, C. Blood plasma lipidome profile of dairy cows during the transition period. *BMC Vet. Res.* **2015**, *11*, 252. [CrossRef] [PubMed]
53. Artegoitia, V.M.; Middleton, J.L.; Harte, F.M.; Campagna, S.R.; de Veth, M.J. Choline and Choline Metabolite Patterns and Associations in Blood and Milk during Lactation in Dairy Cows. *PLoS ONE* **2014**, *9*, e103412. [CrossRef] [PubMed]
54. Zhao, G.; He, F.; Wu, C.; Li, P.; Li, N.; Deng, J.; Zhu, G.; Ren, W.; Peng, Y. Betaine in Inflammation: Mechanistic Aspects and Applications. *Front. Immunol.* **2018**, *9*, 1070. [CrossRef] [PubMed]
55. Bergman, E. Energy Contributions of Volatile Fatty Acids from the Gastrointestinal Tract in Various Species. *Physiol. Rev.* **1990**, *70*, 567. [CrossRef] [PubMed]
56. Urrutia, N.L.; Harvatine, K.J. Acetate Dose-Dependently Stimulates Milk Fat Synthesis in Lactating Dairy Cows. *J. Nutr* **2017**, *147*, 763–769. [CrossRef]
57. Bielak, A.; Derno, M.; Tuchscherer, A.; Hammon, H.M.; Susenbeth, A.; Kuhla, B. Body fat mobilization in early lactation influences methane production of dairy cows. *Sci. Rep.* **2016**, *6*, 28135. [CrossRef]
58. O'Callaghan, T.F.; Vázquez-Fresno, R.; Serra-Cayuela, A.; Dong, E.; Mandal, R.; Hennessy, D.; McAuliffe, S.; Dillon, P.; Wishart, D.S.; Stanton, C.; et al. Pasture Feeding Changes the Bovine Rumen and Milk Metabolome. *Metabolites* **2018**, *8*, 27. [CrossRef]
59. Maher, A.D.; Hayes, B.; Cocks, B.; Marett, L.; Wales, W.J.; Rochfort, S.J. Latent biochemical relationships in the blood-milk metabolic axis of dairy cows revealed by statistical integration of 1H NMR spectroscopic data. *J. Proteome Res.* **2013**, *12*, 1428–1435. [CrossRef]
60. Little, S. *Feeding Systems Used by Australian Dairy Farmers*; Dairy Australia: Melbourne, Australia, 2010.
61. Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943–948. [CrossRef]
62. Nielsen, N.-P.V.; Carstensen, J.M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr.* **1998**, *805*, 17–35. [CrossRef]
63. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
64. Smilde, A.; Jansen, J.; Hoefsloot, H.; Lamers, R.-J.; van Der Greef, J.; Timmerman, M. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics* **2005**, *21*, 3043–3048. [CrossRef] [PubMed]
65. Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R.H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J.C.; Holmes, E.; et al. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Anal. Chem.* **2005**, *77*, 1282–1289. [CrossRef] [PubMed]

Chapter 5:

Use of Large and Diverse Datasets for ^{1}H NMR Serum Metabolic Profiling of Early Lactation Dairy Cows

Metabolites. 2020, 10, 247. doi:10.3390/metabo10060247

T. D. W. Luke, J. E. Pryce, A. C. Elkins, W. J. Wales and S. J. Rochfort

Published: 30th of April 2020 (open access)

metabolites

MDPI

Article

Use of Large and Diverse Datasets for ^{1}H NMR Serum Metabolic Profiling of Early Lactation Dairy Cows

Timothy D. W. Luke [1,2], Jennie E. Pryce [1,2], Aaron C. Elkins [1], William J. Wales [3] and Simone J. Rochfort [1,2,*]

1 Agriculture Victoria Research, AgriBio, Centre for AgriBioscience, Bundoora 3083, Australia; tim.luke@agriculture.vic.gov.au (T.D.W.L.); jennie.pryce@agriculture.vic.gov.au (J.E.P.); aaron.elkins@agriculture.vic.gov.au (A.C.E.)
2 School of Applied Systems Biology, La Trobe University, Bundoora 3083, Australia
3 Agriculture Victoria Research, Ellinbank Centre, Ellinbank 3821, Australia; bill.wales@agriculture.vic.gov.au
* Correspondence: simone.rochfort@agriculture.vic.gov.au

Received: 11 April 2020; Accepted: 27 April 2020; Published: 30 April 2020

Abstract: Most livestock metabolomic studies involve relatively small, homogenous populations of animals. However, livestock farming systems are non-homogenous, and large and more diverse datasets are required to ensure that biomarkers are robust. The aims of this study were therefore to (1) investigate the feasibility of using a large and diverse dataset for untargeted proton nuclear magnetic resonance (^{1}H NMR) serum metabolomic profiling, and (2) investigate the impact of fixed effects (farm of origin, parity and stage of lactation) on the serum metabolome of early-lactation dairy cows. First, we used multiple linear regression to correct a large spectral dataset (707 cows from 13 farms) for fixed effects prior to multivariate statistical analysis with principal component analysis (PCA). Results showed that farm of origin accounted for up to 57% of overall spectral variation, and nearly 80% of variation for some individual metabolite concentrations. Parity and week of lactation had much smaller effects on both the spectra as a whole and individual metabolites (<3% and <20%, respectively). In order to assess the effect of fixed effects on prediction accuracy and biomarker discovery, we used orthogonal partial least squares (OPLS) regression to quantify the relationship between NMR spectra and concentrations of the current gold standard serum biomarker of energy balance, β-hydroxybutyrate (BHBA). Models constructed using data from multiple farms provided reasonably robust predictions of serum BHBA concentration ($0.05 \leq \text{RMSE} \leq 0.18$). Fixed effects influenced the results biomarker discovery; however, these impacts could be controlled using the proposed method of linear regression spectral correction.

Keywords: NMR; metabotype; metabolomics; transition; ketosis; cattle; chemometrics; spectral correction

1. Introduction

Modern metabolomic techniques such as proton nuclear magnetic resonance (^{1}H NMR) spectroscopy allow high-throughput, synchronous characterization of the small metabolites present in biological matrices [1]. In dairy cows, the metabolome gives a snapshot of the complex interactions between host genetics, the rumen microbiome, and the environment at a given time point. ^{1}H NMR-based metabolomics therefore offers exciting opportunities to better understand and characterize the complex physiological and biochemical challenges facing cows in the transition period (defined as the three weeks before and after calving [2,3]) which is the period of greatest disease risk [4]. This in turn can facilitate identification of new molecular phenotypes (metabotypes) for genetic selection for improved animal health. These "intermediate phenotypes," so-termed because they sit

between the genome and external phenotype [5], can then be integrated with genomic data to improve genomic prediction accuracies of complex traits [6,7]. The aim of metabotype identification is therefore to identify biomarkers that represent inter-animal variation free of confounding environmental factors.

Another aim of dairy cattle metabolomic studies is to identify biomarkers which enable early identification of health disorders in the transition period such as ketosis [8,9], hypocalcemia [10] and displaced abomasa [11]. Of particular interest are studies that have identified biomarkers that are predictive of transition period disorders, such as that by Hailemariam et al. [12], who identified a panel of three metabolites that could predict the occurrence of peri-parturient disease up to four weeks before calving. If robust, such predictive biomarkers would enable producers and veterinarians to implement preventive nutritional, management and/or veterinary interventions before the onset of disease.Unlike metabotype biomarkers used for genetic selection, the aim of biomarkers used for management purposes is to predict the external phenotype, and these must therefore capture all sources of phenotypic variation (i.e., host genetics, rumen microbiome, and the environment).

To date, most serum ^{1}H NMR-based metabolomic studies of livestock have involved relatively small numbers of animals, often of a single breed, and often located on a single farm. In their review, Goldansaz et al. [13] identified limited sample size and diversity as limitations of many livestock metabolomics studies and highlighted the need for larger and more diverse datasets to ensure models and biomarkers are robust. However this needs to be balanced against the need for careful experimental design to account for potential confounding from systematic environmental effects such as diet/nutritional management, parity and stage of lactation, which are known to affect the metabolic status of cows [13]. However, in order to achieve large datasets, it may be necessary to obtain samples from multiple different farms, especially when the prevalence of the condition being investigated is low (e.g., displaced abomasa). Previous studies have reported differences in the milk metabolome of animals from different geographical regions [14], farms [15], and of different breeds [16]. However, given that there is not a strong relationship between blood and milk metabolomes [17,18], these findings cannot be extrapolated to the blood serum/plasma metabolome. More information is therefore needed on the impact of systematic environmental effects on the serum metabolome of livestock.

Linear models are routinely used by quantitative geneticists to account for the influence of systematic environmental effects (also known as fixed effects) known to have significant effects on phenotypic variation [19], and thus disentangle genetic from non-genetic effects. Frequently used fixed effects include stage of lactation, parity, and herd-year-season. Similar approaches have recently been applied to metabolomic data, for example Wanichthanarak et al. [20], who used linear mixed-effects models and patient metadata to account for biological variation in metabolomics data, and Laine et al. [21], who used linear models to study the effect of pregnancy on mid-infrared spectral data derived from cows' milk.

The aim of this study was therefore to investigate the feasibility of using of large and diverse datasets in livestock metabolomics studies by examining the effects of fixed environmental and physiological effects on the ^{1}H NMR serum metabolome of clinically healthy dairy cows in early lactation. We propose a method that uses linear models to correct spectra for fixed effects and demonstrate its potential utility by quantifying the relationship between ^{1}H NMR spectra and the current gold-standard serum biomarker of energy balance, β-hydroxybutyrate (BHBA) [22,23].

2. Results

2.1. Dataset

Serum samples were collected from 707 early lactation cows (<30 d in milk) from 13 farms located in southeastern Australia. Descriptive statistics of the animals included in the experiment, including herd of origin, stage of lactation (reported as days in milk, or the number of days post-calving), parity and serum BHBA concentrations and are summarized in Table 1. Of particular interest were the BHBA

results obtained from Farm 1, which had a greater mean and standard deviation than observed in other farms.

Table 1. Descriptive statistics of dataset used in this experiment, including farm details, number of cows (N), mean and standard deviation (shown in parentheses) of parity, days in milk (DIM), and serum β hydroxybutyrate (BHBA) concentrations obtained from dairy cows in the first 30 days of lactation from 13 farms in south eastern Australia.

Farm	N	Location	Parity	DIM	BHBA
1	129	Sth Gipp [1]	2.9 (1.1)	19.4 (7.2)	1.25 (0.69)
2	11	Sth Gipp	2.6 (1.2)	20.4 (8.1)	0.34 (0.12)
3	12	W Gipp [2]	2.6 (1.4)	22.8 (5.7)	0.33 (0.10)
4	11	W Gipp	3.1 (1.2)	17.9 (10.2)	0.54 (0.15)
5	18	MID [3]	2.9 (1.1)	22.6 (5.1)	0.61 (0.25)
6	248	W Gipp	2.1 (1.0)	16.7 (6.0)	0.55 (0.21)
7	9	GV [4]	2.6 (1.0)	13.9 (6.7)	0.53 (0.27)
8	24	MID	2.4 (1.2)	17.7 (8.2)	0.38 (0.09)
9	33	Sth Gipp	2.5 (1.2)	18.3 (7.2)	0.55 (0.33)
10	27	Sth Gipp	1.8 (1.1)	13.1 (7.7)	0.50 (0.14)
11	50	Tas [5]	2.6 (1.3)	18.6 (7.3)	0.42 (0.17)
12	123	MID	2.8 (1.2)	15.8 (8.6)	0.38 (0.15)
13	12	Tas	2.7 (0.8)	16.0 (7.6)	0.58 (0.22)
ALL	707	-	2.5 (1.2)	17.4 (7.3)	0.63 (0.46)

[1] South Gippsland Region, [2] West Gippsland Region, [3] Macalister Irrigation District, [4] Goulburn Valley Region, [5] Tasmania.

2.2. ^{1}H NMR Spectroscopy of Serum Samples

^{1}H NMR spectra were complex; however, more than 20 metabolites could be identified. Spectra were dominated by organic acids, amino acids, glucose and phospholipid intermediates (Figure S1 and Table S1).

2.3. Preliminary Data Analysis Using Principal Component Analysis

Preliminary data analysis and outlier identification was performed using principal component analysis (PCA). Plots of the first 2 principal components (PCs) identified several samples located outside the 95% confidence level. These spectra were manually inspected, and a single outlier with erroneous phasing was identified and removed from subsequent analyses.

PCA was repeated after outlier removal. The first 13 PCs explained greater than 90% of the variation in the spectra. Scores plots of the first three PCs, which explain 47.64%, 15.59%, and 7.45% of variation, respectively, are shown in Figure 1a–c. There was obvious clustering of samples by herd of origin. Samples from Farm 1 showed greater variation than those from the other farms. The separation between farm clusters was most obvious along PC1 and PC2. Visual comparisons based on stage of lactation (defined as weeks in milk (WIM)) and parity were also performed, but no obvious clustering or separation was observed. Loadings plots of the first three PCs show that energy metabolites BHBA, lactate, acetate and glucose, have the largest influences on spectral differences (Figure 1d–f), with smaller influences from the branched chain amino acids, lipoproteins, glycine, creatine, and betaine.

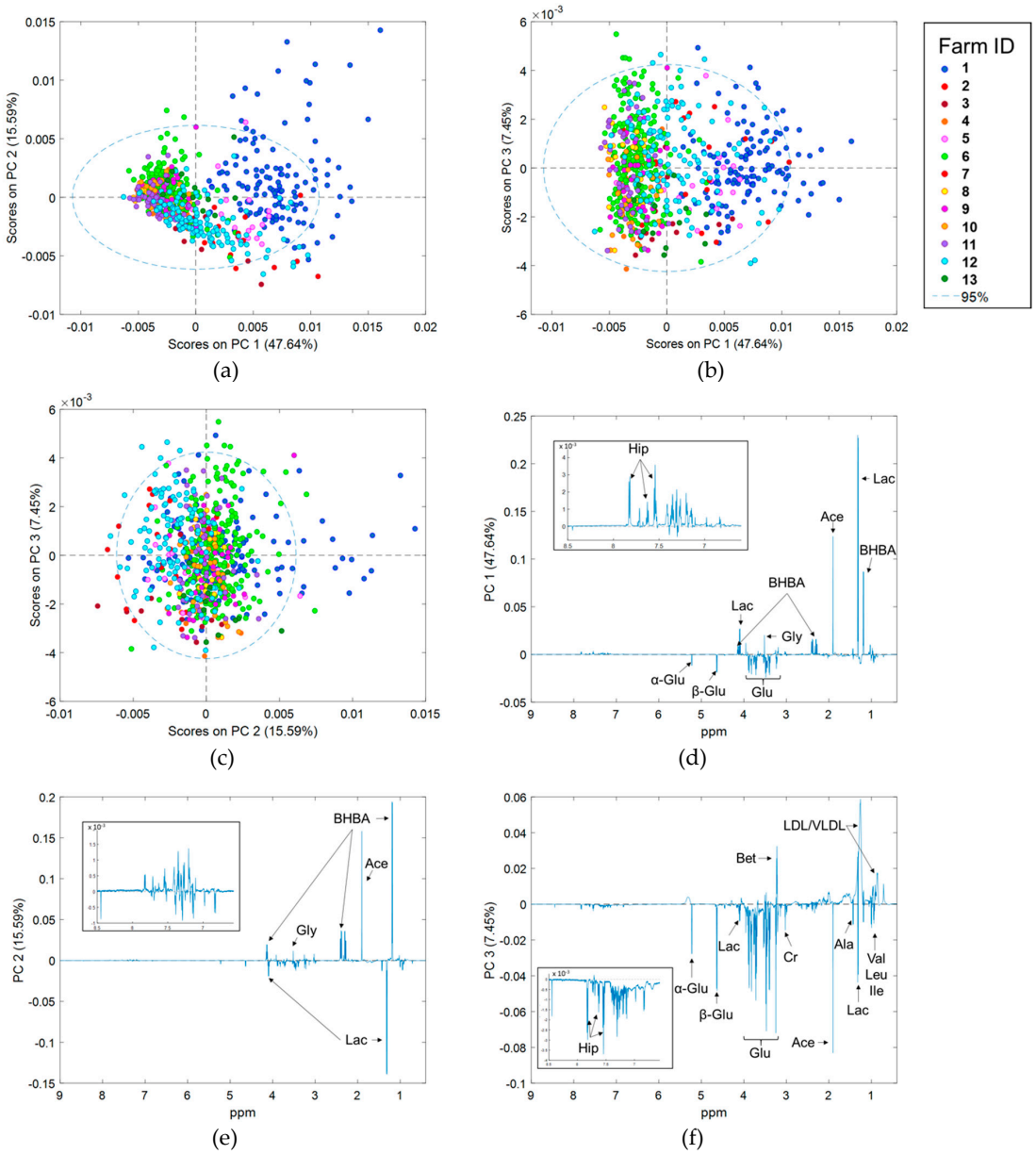


Figure 1. Results of principal component analysis (PCA) of 707 proton nuclear magnetic resonance (^{1}H NMR) spectra of serum obtained from dairy cows in early lactation; (**a**) principal component (PC) 1 vs. PC 2 scores, (**b**) PC 1 vs. PC 3 scores, (**c**) PC 2 vs. PC 3 scores, (**d**) PC 1 loadings, (**e**) PC 2 loadings, and (**f**) PC 3 loadings plots. Scores plots are colored by farm of origin. The δ 6.5 to 8.5 region of loadings plots have been magnified for clarity purposes. α-Glu = α glucose, Ace = acetate, Ala = alanine, β-Glu = β glucose, Bet = betaine, BHBA = β hydroxybutyrate, Cr = creatine, Glu = glucose, Gly = glycine, Hip = hippurate, Ile = isoleucine, Lac = lactate, Leu = leucine, Val = valine, VLDL/LDL = Very low density lipoprotein and low density lipoprotein.

2.4. Principal Component Analysis of Spectra Corrected for Fixed Effects

Principal component analysis (PCA) was repeated on spectra that had been corrected for (1) WIM, (2) Parity, (3) Herd, and (4) WIM, Parity and Herd simultaneously (hereafter referred to as all fixed effects) (Models 1 to 4). Results derived from spectra corrected separately for WIM and Parity are nearly identical to uncorrected spectra (Figures S2 and S3). By contrast, scores plots derived from PCA of spectra corrected for Herd (Figure S4), and spectra corrected for all fixed effects (Figure 2a–c), show no obvious clustering of samples by Herd, WIM or Parity. There is, however, still considerable

separation of samples along all three PC axes, suggesting that significant inter-animal variation in the serum metabolome exists after accounting for fixed effects. Compared to the uncorrected data; (1) more PCs were required to explain >90% of spectral variation (24 vs. 13), (2) the percentage of variation captured by PC1 was lower (25.70% vs. 47.64%), and (3) the percentage of variation captured by PC2 and PC3 was higher (16.92% vs. 15.59% and 11.79% vs. 7.45%, respectively). Loadings plots are shown in Figure 2d–f. Interestingly, separation of samples along PC1 (25.70%) is due almost entirely to lactate. Loadings on PC2 and PC3 are similar to uncorrected data.

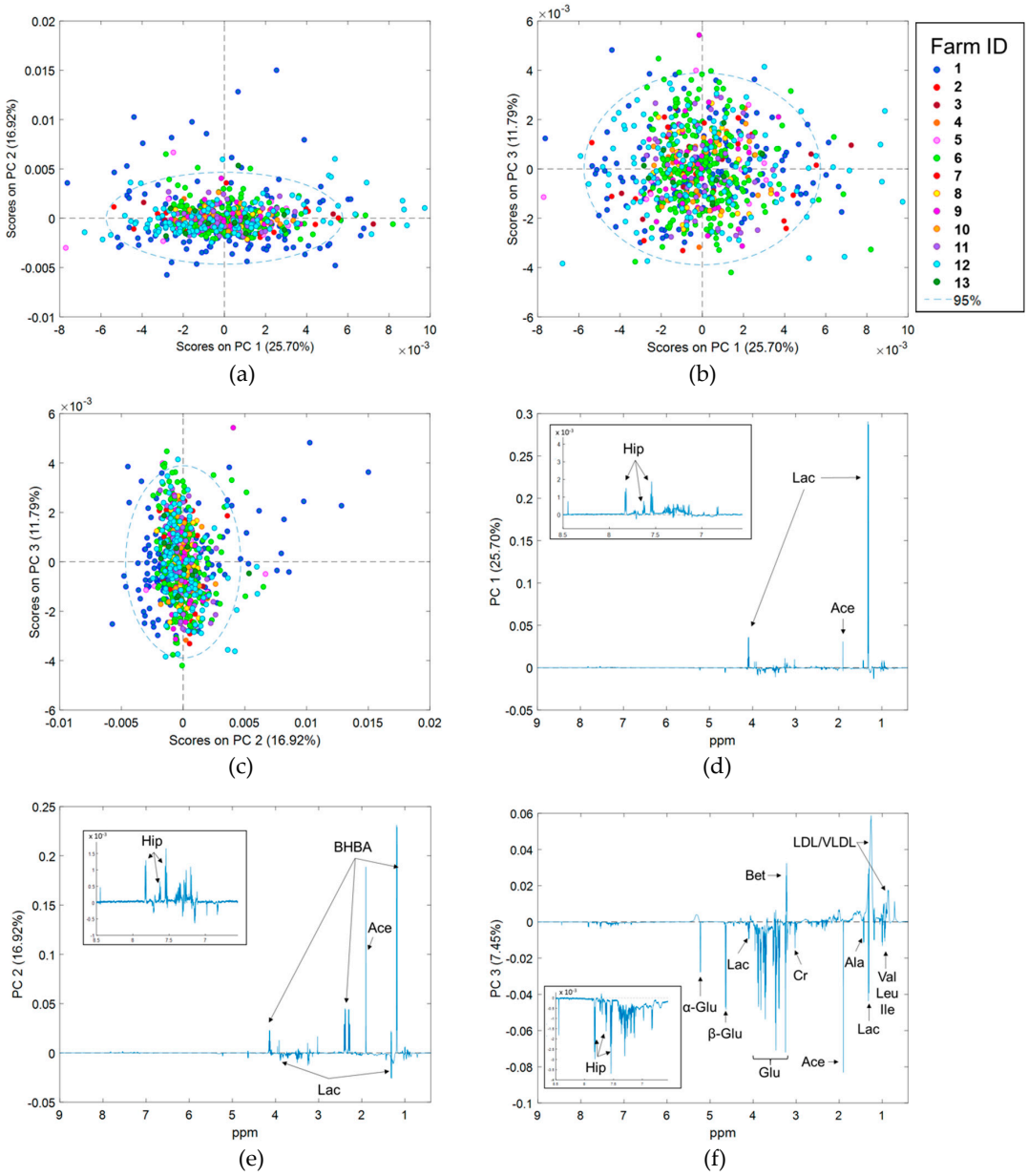


Figure 2. Results of PCA of 707 ^{1}H NMR spectra of serum, corrected for herd of origin, week of lactation, and parity obtained from dairy cows in early lactation; (**a**) PC 1 vs. PC 2 scores, (**b**) PC 1 vs. PC 3 scores, (**c**) PC 2 vs. PC 3 scores, (**d**) PC 1 loadings, (**e**) PC 2 loadings, and (**f**) PC 3 loadings plots. Scores plots are colored by farm of origin. The δ 6.5 to 8.5 region of loadings plots have been magnified for clarity purposes. α-Glu = α glucose, Ace = acetate, Ala = alanine, β-Glu = β glucose, Bet = betaine, BHBA = β hydroxybutyrate, Cr = creatine, Glu = glucose, Gly = glycine, Ile = isoleucine, Lac = lactate, Leu = leucine, Val = valine, VLDL/LDL = Very low density lipoprotein and low density lipoprotein.

2.5. Effect of Stage of Lactation, Parity, and Herd Effects on ^{1}H NMR Spectra

In order to quantify the effect of each fixed effect on NMR spectra, we calculated Pearson's correlations between scores for the first three PCs from the previously described PCAs (Figure 3). The largest differences (i.e., lowest correlations) were seen between uncorrected spectra, and spectra corrected for Herd (r = 0.43). This suggests that there are significant differences between those 2 spectral datasets, and that Herd, therefore, has a significant effect on the serum NMR metabolome. This is consistent with the clustering of samples by farm in the original PCA (Figure 1a–c). By comparison, the correlations between PC scores derived from uncorrected spectra, and spectra corrected for WIM and Parity, were high (0.99 and 0.97, respectively). This suggests that these spectra are nearly identical, and that these fixed effects have minimal influence on the serum metabolome. Correlations between PC2 scores were consistent with those observed between PC1 scores, and correlations between PC3 scores were all high (≥0.89).

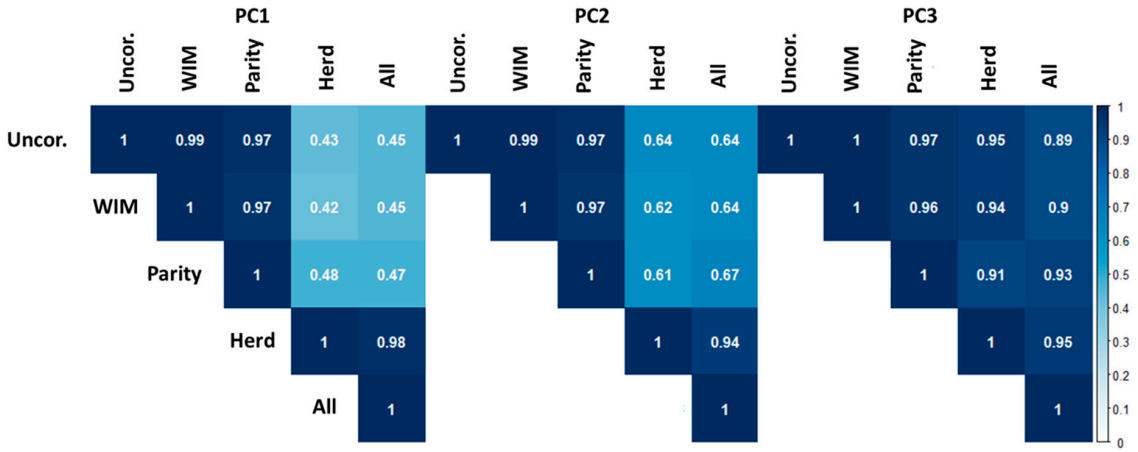


Figure 3. Pearson's correlations between scores derived from PCA of uncorrected ^{1}H NMR spectra of bovine serum, and the same spectra corrected using linear regression for week of lactation (WIM), parity, herd of origin, and WIM, parity, and herd simultaneously (All). Color map shows strength of Pearson's correlation.

To test the statistical significance of fixed effects on ^{1}H NMR spectra, we used conditional Wald F statistics derived from multiple linear regression models on the first three PC scores (Model 5). The higher the F statistic, the greater the effect of that variable on the PC score, and the lower the P value, the greater the statistical significance. Results derived from these models are summarized in Table 2. PC1 results were consistent with the results of Pearson's correlations, showing that Herd had the greatest effect. Interestingly, results for PC2 and PC3 differed slightly from Pearson's correlations. While Herd had a relatively large and significant (P < 0.001) impact on both PC2 and PC3, the effect of Parity was nearly as great on PC2 scores and greater on PC3 scores.

Table 2. Results of multiple linear regression models of principal component (PC) scores derived from PCA of ^{1}H NMR spectra, against weeks in milk (WIM), parity, and herd of origin. Conditional Wald F statistics (F-con) and corresponding *P* values describe the magnitude and statistical significance of each fixed effect, respectively.

	PC1 (47.64%)		PC2 (15.59%)		PC3 (7.45%)	
Fixed Effect	**F-con**	***P* Value**	**F-con**	***P* Value**	**F-con**	***P* Value**
WIM	2.66	0.047	5.42	0.001	2.14	0.094
Parity	2.78	0.041	20.39	<0.001	15.19	<0.001
Herd	158.29	<0.001	26.78	<0.001	6.66	<0.001

R^2 values obtained from models 1–4 were used to investigate which regions of the NMR spectra were most strongly influenced by the fixed effects. As the signal intensity at each chemical shift was treated as a separate response variable, the R^2 values from Models 1, 2, and 3 describe the effect of WIM, parity, and herd on each of the 24,349 chemical shifts, respectively. These R^2 values were color-coded, and overlaid on an average NMR spectrum. Plots showing the effects of WIM and Parity were unremarkable (all $R^2 < 0.2$, Figure S5), however R^2 values obtained from Model 2 showed that approximately 10–20% of the variation in glucose and acetate concentration could be explained by parity. The plot showing the effect of herd is shown in Figure 4. The strongest effect was seen in the downfield region of the spectrum, with close to 80% of variation in the concentration of some phenolic compounds being explained by Herd. Of these, hippurate could be clearly identified. Peaks at δ 7.31 and 7.39 were tentatively assigned to 3-phenyllactate, but the peak at δ 7.22 could not be identified. Lactate, acetate, BHBA, betaine, pyruvate, glycine, and glucose concentrations were also strongly influenced by herd effect.

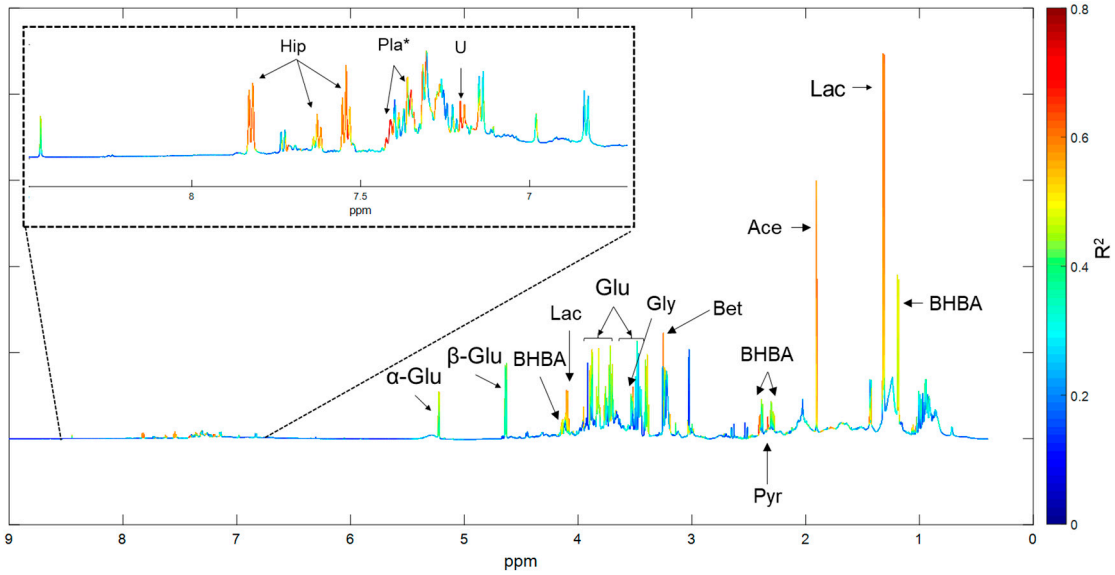


Figure 4. Average ^{1}H NMR spectrum of bovine serum. Color coding represents the percentage of variation in the signal at each chemical shift intensity that can be explained by herd of origin. The δ 6.5 to 8.5 region has been magnified for clarity purposes. Ace = acetate, Bet = betaine, BHBA = β hydroxybutyrate, Gly = glycine, Hip = hippurate, Lac = lactate, Pla = 3-phenyllactate, Pyr = pyruvate, U = unidentified peak. * indicates tentative identification.

The results of ANOVA-simultaneous component analysis (ASCA) were consistent with results of linear regression spectral correction and are shown in Table S2. Herd had the greatest effect (43.99, P = 0.02), followed by parity (4.10, P = 0.02) and WIM (1.37, P = 0.02). When ASCA was performed on corrected spectra, the effect of the fixed effect(s) was reduced to zero. For example, when ASCA was performed on spectra corrected for Herd, the effect of herd was zero (P = 1.00), but the effects of WIM (1.68, P = 0.02) and parity (3.30, P = 0.02) were retained.

2.6. Robustness of ^{1}H NMR Predictions of Serum BHBA

Our results show that ^{1}H NMR spectra can be used to predict serum BHBA concentration with good accuracy. This result is expected, as BHBA is directly quantifiable from NMR spectra. The overall robustness of our approach was assessed using a "leave-one-farm-out" external validation of OPLS models built using uncorrected data. This involved iteratively setting aside data from one farm, training models using data from the remaining 12 farms, then using the withheld data for external validation. R^2 results were variable ($0.30 \leq R^2 \leq 0.99$), however RMSE values remained relatively

low (≤ 0.18) (Table 3). Interestingly, RMSE values were considerably higher when Farm 1 data were withheld for validation.

Table 3. Results of leave-one-farm out external validation of orthogonal partial least squares (OPLS) regression models predicting serum BHBA concentration from uncorrected ^{1}H NMR spectra. Validation farm specifies the identity of the data used for validation, N the number of animals used in calibration and validation datasets. The coefficient of determination (R^2) and root mean square error (RMSE) are reported for each calibration/validation subset.

			Calibration			Cross Validation		External Validation		
Validation Farm	**P**	**LV**	**N**	**R^2**	**RMSE**	**R^2**	**RMSE**	**N**	**R^2**	**RMSE**
-	<0.05	3	707	0.95	0.10	0.95	0.10	-	-	-
1	<0.05	5	578	0.87	0.08	0.85	0.08	129	0.96	0.18
2	<0.05	3	696	0.95	0.10	0.95	0.10	11	0.59	0.10
3	<0.05	4	695	0.96	0.09	0.96	0.10	12	0.78	0.06
4	<0.05	3	696	0.95	0.10	0.95	0.10	11	0.93	0.09
5	<0.05	3	689	0.96	0.10	0.95	0.10	18	0.99	0.09
6	<0.05	3	459	0.96	0.11	0.96	0.11	248	0.87	0.10
7	<0.05	3	698	0.95	0.10	0.95	0.10	9	0.98	0.05
8	<0.05	3	683	0.95	0.10	0.95	0.10	24	0.30	0.07
9	<0.05	3	674	0.95	0.10	0.95	0.10	33	0.95	0.11
10	<0.05	3	680	0.95	0.10	0.95	0.10	27	0.85	0.09
11	<0.05	3	657	0.95	0.10	0.95	0.10	50	0.82	0.08
12	<0.05	3	584	0.97	0.09	0.96	0.09	123	0.52	0.12
13	<0.05	3	695	0.95	0.10	0.95	0.10	12	0.98	0.05

2.7. Influence of Fixed Effects on Interpretation of ^{1}H NMR Metabolomic Data

The impact of fixed effects on the interpretation of ^{1}H NMR metabolomic data was determined by comparing the results of OPLS models built using (1) data from Farm 1 only (used as a control), (2) uncorrected data from all farms, and (3) data from all farms corrected for all fixed effects. Fixed effects appeared to have minimal effect on the predictive ability of models. We observed similar 10-fold cross validation prediction accuracies for all 3 datasets (Table 4). Interestingly, RMSE results were quite close to the results of the leave-one-farm out external validation (0.05 ≤ RMSE ≤ 0.18).

Table 4. Results obtained from 10-fold cross validation of OPLS regression models predicting serum BHBA concentration from ^{1}H NMR spectra using data from Farm 1 only, uncorrected data from all farms, and data from all farms corrected for the effect of herd. Number of cows (N), number of latent variables included in each mode (LV), coefficient of determination (R^2) and root mean square error (RMSE) of calibration (C) and 10-fold cross validation (CV) are shown.

Dataset	**N**	**LVs**	**P Value** [1]	**R^2_C**	**$RMSE_C$**	**R^2_{CV}**	**$RMSE_{CV}$**
Farm 1 Uncorrected	129	4	<0.001	0.98	0.10	0.97	0.12
All Data Uncorrected	707	4	<0.001	0.96	0.09	0.96	0.10
All Data Corrected for Herd	707	4	<0.001	0.93	0.09	0.93	0.09

[1] P-value derived from permutation testing (50 iterations) and pairwise Wilcoxon signed rank test.

The influences of fixed effects on biomarker discovery were investigated by comparing loadings on LV1. Results obtained using only Farm 1 data were used as a reference and show a strong positive correlation between BHBA concentration and acetate, and strong negative correlations with lactate and glucose (Figure 5a,b). Loadings from the complete dataset corrected for all fixed effects were very similar (Figure 5e,f). Results from uncorrected data, however, were quite different (Figure 5c,d), with BHBA being positively correlated with lactate and glycine. Examination of scores plots shows obvious clustering and separation by herd (especially Farm 1) when uncorrected data are used (Figure S6a), but not when corrected data are used (Figure S6b). Results from the original PCA showed that samples from Farm 1 clustered at the positive end of PC1, and that lactate and glycine both had

strong positive influences on PC1 loadings. Therefore, it is possible that OPLS results are confounded by a strong herd effect when uncorrected data are used.

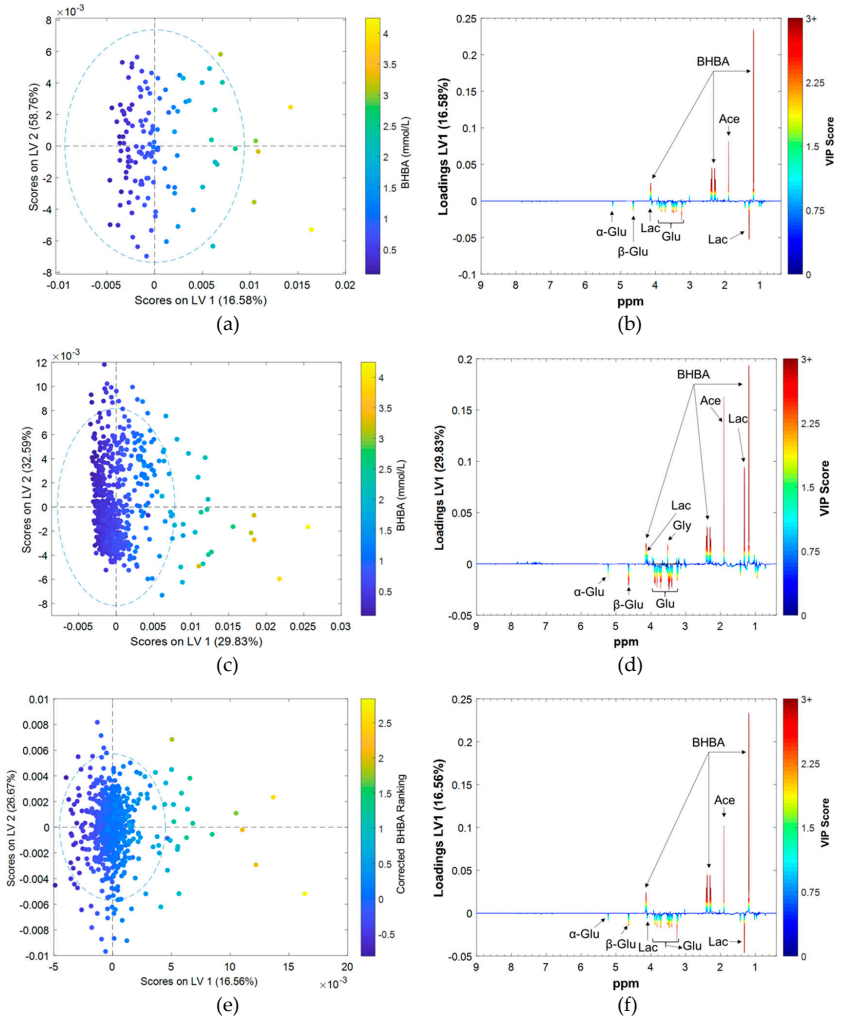


Figure 5. Results of OPLS regressions of serum BHBA concentration against ^{1}H NMR spectrum of bovine serum: (**a**) Farm 1 (N = 179) LV1 vs. LV2 scores and (**b**) LV1 loadings, (**c**) all farms (N = 707) uncorrected data LV1 vs. LV2 scores and (**d**) LV1 loadings, and (**e**) all farms data LV1 vs. LV2 scores and (**d**) LV1 loadings.

3. Discussion

To the best of the authors' knowledge, this is the first large-scale serum metabolomics study to investigate the impact of systematic environmental and physiological fixed effects on the ^{1}H NMR serum metabolome of clinically healthy dairy cattle. Our results indicate that herd-specific environmental factors have much greater effects on the serum metabolome of early lactation dairy cows than physiological factors such as WIM and parity. We demonstrate that, while confounding from herd effects can significantly influence the results of biomarker discovery, models built using data collected from multiple farms can give robust predictions of external phenotypes such as BHBA. In order to overcome the potential confounding of fixed effects on biomarker discovery, we propose a method to correct ^{1}H NMR spectra prior to multivariate analysis using multiple linear regression.

3.1. Differences in ^{1}H NMR Spectra Between Herds

Our results clearly demonstrate that there are significant differences in the serum metabolomes of animals from different herds. The fact that energy metabolites BHBA, lactate, acetate and glucose

dominated PCA loadings (Figure 2d–f), and that herd effect accounted for a large percentage of the variation seen in lactate, acetate, pyruvate, glucose, and BHBA concentrations, (Figure 4) suggests that metabolic state, in particular energy balance, varied significantly between farms.

The importance of lactate was particularly interesting. Lactate was one of the most abundant metabolites identified in this experiment. This is very different to the findings of Sun et al. [8], who reported that lactate was one of the weakest signals in serum ^{1}H NMR spectra obtained from early-lactation cows fed a total mixed ration. One possible explanation for the very high concentrations of lactate seen in our dataset could be ruminal lactate production. During spring, dairy cows in pastoral farming systems of southeastern Australia are typically fed rations high in fermentable carbohydrate, and low in neutral detergent fiber. As a consequence, ruminal acidosis is common [24]. Serum concentrations of lactate, and in particular D-lactate from microbial fermentation, have been shown to increase following experimental induction of ruminal acidosis [25,26]. Without the use of a shift reagent and specialized experiments it is not possible to differentiate between the different lactate isomers by ^{1}H NMR [27]. We therefore plan to quantify the relative contributions of L- and D- lactate to better understand the cause of high lactate concentrations in our dataset.

The strong influence of Herd on the concentration of phenolic compounds could also be consistent with ruminal acidosis. Signal intensities in the downfield region of 2D spectra were weak, meaning clear identification of some of the phenolic peaks in our dataset was not possible. Our tentative identification of 3-phenyllactate is consistent with the findings of Yang et al. [26], who demonstrated that beef steers fed high starch (corn) diets had higher plasma concentrations of phenyllactate compared to those fed low starch diets. This study also identified L-phenyllalanine biosynthesis and metabolism as important metabolic pathways in high starch feeding. We plan to (1) enrich samples and repeat 2D analyses and (2) perform LCMS-based metabolomics on a subset of samples to identify these compounds.

Nearly 80% of the variation seen in hippurate concentration could be explained by herd effect (Figure 4). Hippurate is formed by the conjugation of glycine and benzoic acid, and has been associated with microbial degradation of dietary compounds [28]. Concentrations of hippurate increase with increased consumption of phenolic compounds [13], which are present in relatively high concentrations in pasture species. Milk hippurate concentration has been proposed as a biomarker of pasture/forage intake in goats [29], and it is possible that our results represent differences in feeding regimens between farms. Hippurate has also been proposed as a biomarker for gut microbiome diversity in humans [30], and our results may indicate differences in the gastrointestinal health of animals from different farms (i.e., ruminal acidosis). Detailed information of ration formulations is very difficult to define in grazing systems as pasture quality and intake vary considerably within and between herds. This information was therefore not available for the herds in our dataset and more data are required to further investigate this finding.

Results of the initial PCA showed that data from Farm 1 were significantly different to, and showed more variation than, data from the other farms. The reasons for these differences are hard to determine from our dataset, as Farm 1 differed in environment/management, breed and reference BHBA concentrations (and therefore it is assumed animal metabolic status). Given that we also observed clustering and separation of the 12 Holstein-Friesian herds in the initial PCA (Figure 2), it appears that herd-specific environmental factors have a larger effect on the serum metabolome than breed. However, Liao et al. [31] recently reported clear differences in the serum metabolomes (GC-MS) of three different breeds of beef steers, all the same age, fed the same ration, and managed under the same conditions. Further data are therefore required to investigate if there are differences between the serum metabolomes of different dairy breeds.

Pre-analytical sample handling and processing have been shown to have significant effects on human metabolomic data [32], and considerable efforts are made to streamline and standardize sample collection and processing protocols [33–35]. Standardizing protocols in livestock studies provides its own challenges, when relatively large number of samples are being taken at once, often in diverse, challenging and remote locations. While all attempts were made to ensure consistency, there were

some unavoidable differences in the way samples from different herds were handled (for example time, between blood sample collection and centrifugation varied from approximately 2–4 h). It is therefore possible that some of the variation between farms seen in our data could be due to pre-analytical sample handling. However, overall our results suggest that metabolomic differences between animals from different farms are due largely to differences in diet/nutritional management. We plan to collect more samples from animals receiving different diets to investigate this further.

3.2. Effect of Lactation Stage and Parity on Serum Metabolome

Our results suggest that stage of lactation appeared to have a minimal effect on the NMR spectra. This is consistent with the findings of Ilves et al. [17] who found that the mass spectrometry (MS) based plasma metabolome of dairy cows was more heavily influenced by animal individuality than by lactation stage. By contrast, several authors report that both the NMR and MS-derived milk metabolome changes across lactation [17,36]. This suggests that blood-based metabolomics may be more suitable for identification of individual animal-specific differences within a population, and therefore provide more robust metabotypes for genetic selection.

Parity appeared to have a small but significant ($P < 0.05$) effect on the overall ^{1}H NMR serum metabolome. We could find no other reports in the literature describing the effect of parity on the entire serum metabolome. However, our results are consistent with other studies that showed parity has a significant effect on the concentration of several metabolites in serum including glucose, creatinine, urea and BHBA [37–39]. This suggests that parity should be taken into consideration when undertaking metabolomic studies in dairy cows.

3.3. Accuracy of OPLS Models for Predicting Serum BHBA Concentration

Despite the significant influence of fixed effects on the serum metabolome, results obtained from the leave-one-farm out external validation suggest that prediction models constructed with data from multiple farms are quite robust. R^2 values varied significantly depending on which farm was used for validation ($0.30 \leq R^2 \leq 0.99$); however, the R^2 is known to be affected by the range of the dataset, and RMSE is often considered to be a better predictor of model performance [40]. Promisingly, external validation RMSE results ($0.05 \leq R^2 \leq 0.18$) were close to those obtained from 10-fold cross validation of models built using only Farm 1 data (RMSE = 0.12) and all data (RMSE = 0.10). The fact that prediction errors were highest when Farm 1 data were withheld for validation suggests that the increased variation observed in Farm 1 data represents valuable biological variation rather than confounding/noise.

Correcting data for fixed effects had very little impact on the predictive ability of OPLS models. Furthermore, when corrected spectra were used, y-values also had to be corrected, making interpretation of phenotypic values difficult. Interestingly, Wanichthanarak, et al. [20] found that "readjusting" mass spectroscopy metabolite signals using patient metadata and linear mixed models improved the sensitivity and specificity of classification of human tissue samples with and without colorectal cancer. Conversely, Posma et al. [41] found that adjusting NMR data for confounding factors lead to a loss of predictive power for cardiovascular risk in a large-scale human NMR metabolomic dataset. Whether using NMR spectra corrected using linear regression will improve the performance of classification models (as opposed to regression against a continuous variable as used in this study) requires further investigation. Overall, our results suggest that models constructed using uncorrected data collated from multiple farms may be appropriate for prediction of external phenotypes which are influenced by both genetic and environmental factors.

3.4. Impact of Fixed Effects on the Interpretation of Metabolomic Data for Biomarker and Metabotype Discovery

Loadings from OPLS models built using uncorrected spectra from Farm 1, and spectra from all farms corrected for fixed effects, were consistent with the literature. BHBA and glucose concentrations have been shown to be negatively correlated in the serum of cows in early lactation dairy cows [42]. L-lactate is an important gluconeogenic substrate in dairy cows [43,44], so it follows that lactate

concentration is also negatively correlated with BHBA concentration. Our results are also consistent with the findings of Sun et al. [8] who showed that cows with subclinical (1.2 < BHBA < 2.9 mmol/L) and clinical ketosis (BHBA > 2.9 mmol/L) had lower lactate and glucose concentrations and higher BHBA and acetate concentrations than the healthy controls.

The fact that loadings were different when uncorrected spectra from all farms were used demonstrates that herd-specific environmental effects can influence the results of biomarker discovery. How significant this is ultimately depends on the research question being asked. If the study aim is to identify biomarkers of external phenotypes (i.e., biomarkers that represent both genetic and environmental factors which are used for management purposes such as disease prediction), then the impact of environmental effects is important and must be captured. However, if the aim is to identify biomarkers indicative of inter-animal differences free of environmental confounding, or to understand biological processes, our results suggest that the influence of environmental effects could lead to erroneous results. This is consistent with the findings of Posma et al. [41] who showed that differences in fixed effects between subjects from the north and south of China explained some metabolite associations, which had previously been attributed to cardiovascular disease risk. This study also reported that adjusting metabolomic data for confounding using an algorithm called Covariate-Adjusted Projection to Latent Structures (CA-PLS) improved model interpretability and led to the identification of more robust biomarkers. Our results are also consistent with other studies that have explored the impacts of data pretreatments on the interpretation of metabolomics data. For example, van den Berg et al. [45] showed that pretreatment methods such as scaling, centering and transformations can greatly affect the outcome of metabolomic analyses (including the biological ranking of important metabolites) and have the potential to enhance biological interpretability. Similarly, Emwas et al. [46] concluded that the choice of spectral processing and post-processing depended on many factors including the aim of the experiment and the quality of data.

We believe that our approach has particular application in animal breeding, where the aim is to understand the biological processes that underpin economically important traits [47] and to identify metabotypes that represent inter-animal variation independent of confounding from systematic environment effects. Even with the advent of genomic selection, livestock genetic studies require relatively large numbers of animals to ensure there is adequate genetic variation in the study population [48,49]. The same is likely to be true for metabotype discovery studies. Such large datasets can be hard to compile, especially when the trait of interest is difficult and/or expensive to measure. As well as collecting data from multiple farms, another potential solution is data sharing through international collaboration. This is routinely done by geneticists; for example, de Haas et al. [50] used data from Holstein cattle in Europe, North America and Australasia to improve genomic prediction accuracies for feed intake. The ability to correct metabolomic data for factors such as experimental batch, diet, herd, year and season should allow similar collaborations in metabotype studies.

4. Materials and Methods

All procedures undertaken in this study were conducted in accordance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes (National Health and Medical Research Council, 2013). Approval to proceed was granted by the Agricultural Research and Extension Animal Ethics Committee of the Department of Jobs, Precincts and Resources Animal Ethics Committee (DJPR, 475 Mickleham Road, Attwood, Victoria 3049, Australia), and the Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIPWE Animal Biosecurity and Welfare Branch, 13 St Johns Avenue, New Town, Tasmania 7008, Australia). AEC project approval codes 2017-05 and 2018-07.

4.1. Sample Collection

A single 10 mL blood sample was taken from 708 clinical healthy cows, located on 13 farms in south-eastern Australia between September 2017 and July 2019. All cows had been calved 30 days

or less at the time of sampling. Cows on all farms except Farm 1 were Australian Holstein-Friesians, while cows on Farm 1 were crossbred animals (including Holstein–Friesian, Jersey, and Australian Red breeds). All farms operated a feeding system reliant on grazed pasture plus other forages, and concentrates fed in the bail at milking time.

Blood samples were collected from the coccygeal vein into 10 mL serum clot activator vacutainer tubes (Becton Dickinson, Franklin Lakes, NJ, USA). Samples were allowed to clot at room temperature, before being centrifuged at 1000 g for 20 min at 20 °C. Sera were divided into two aliquots. The first aliquot was refrigerated at 4 °C then shipped on ice to a commercial laboratory for BHBA analysis. The second aliquot was stored at –20 °C until processing for NMR spectroscopy.

4.2. Reference BHBA Measurements

Serum BHBA concentrations were determined using a colorimetric enzymatic kinetic assay [51]. All assays were performed by Regional Laboratory Services (Benalla, Victoria, Australia) using a Kone 20 XT clinical chemistry analyzer (Thermo Fisher Scientific, Waltham, MA, USA). The uncertainty of measurement (at a 95% confidence level) was ± 0.060 mmol/L at 0.85 mmol/L.

4.3. Chemicals

Methanol (>99.9% pure) and dipotassium hydrogen phosphate (anhydrous) were purchased from Fisher Chemical (Fair Lawn, NJ, USA). Sodium 2,2-dimethyl- 2-silapentane-5-sulfonate (DSS-d6, 98%) and deuterium oxide (D_2O, 98%) were purchased from Cambridge Isotope Laboratories, Inc. (Tewksbury, MA, USA).

4.4. Sample Preparation for NMR Spectroscopy

Serum samples were thawed at room temperature for one hour and were prepared for NMR spectroscopy using a methanol protein precipitation method described by Nagana Gowda and Raftery [52]. Briefly, 300 µL of serum was mixed with 600 µL of methanol, vortexed (Ratek multi tube vortex mixer, MTV1), incubated at –20 °C for 20 min, then centrifuged to pellet proteins (11,360 g, 21 °C, 30 min). A 600 µL aliquot of supernatant was then transferred to a clean 2 mL microcentrifuge tube and dried under vacuum at 21 °C overnight using a SpeedVac Savant SPD 2010 Concentrator (Thermo Fisher Scientific, Waltham, MA, USA). Dried extracts were then reconstituted in a D_2O phosphate buffer solution (100 mM K_2HPO_4) containing 0.25 mM DSS-d6 as an internal standard. A 550 µL aliquot was transferred to 5 mm NMR tube for analysis.

4.5. ^{1}H NMR Data Acquisition

Routine 1D proton spectra were obtained on a Bruker Ascend 700 MHz spectrometer equipped with cryoprobe and SampleJet automatic sample changer (Bruker Biospin, Rheinstetten, Germany). A Bruker noesypr1d pulse sequence was used over –0.76 ppm to 10.32 ppm spectral range with 256 scans collected after eight dummy scans at 298K, with a total acquisition time of 2.11 seconds per increment and a relaxation delay (D1) of 2.00 seconds. The overall number of data points was 32,768. A line broadening of 0.3 Hz was applied to all spectra prior to Fourier transformation. Spectra were manually phased then baseline corrected in Topspin v.3.6.1 (Bruker Biospin, Rheinstetten, Germany). Samples were referenced to the internal standard (DSS-d6) at δ 0.00.

4.6. ^{1}H NMR Spectral Processing & Multivariate Statistical Analysis

NMR spectra were imported into MatLab v.R2017b (Mathworks, Natick, WA, USA) using the ProMetab v.1.1 script [53]. Each raw spectrum consisted of 31,313 data points between −0.60 and 10.00 ppm.

Statistical analyses were performed in MatLab utilizing the PLS Toolbox v. 8.5.2 (Eigenvector Research Inc., Manson, WA, USA). The spectral region containing the residual water peak (δ 4.68–5.00)

was removed. Spectra were aligned using the correlation optimized warping algorithm [54] to account for chemical shift drift, then normalized to total signal area to account for inherent concentration differences between samples. After normalization, spectral regions containing methanol (δ 3.32–3.36) and DSS-d6 (δ 0.4–−0.60) peaks, and the non-informative region beyond 9.00 ppm were removed. Finally, spectra were baseline corrected using automatic weighted least squares, and scaled by mean centering. After editing, a total of 24,349 chemical shift datapoints were included in subsequent statistical analyses.

For multivariate analyses, unsupervised principal component analysis (PCA) was used. Peaks of interest were identified using the Chenomx NMR suite software v.8.4 (Chenomx Inc., Edmonton, AB, Canada), comparison to the literature, and 2D NMR analysis.

4.7. Correction of ^{1}H NMR Spectra for the Effects of Systematic Environemtal and Physiological Effects

In order to investigate the effects on spectra of systematic environmental effects (also known as fixed effects) spectra were "corrected" using linear regression models. When correcting for a single categorical fixed effect, this is equivalent to scaling data using the "class centering" pre-processing step. Rather than mean centering, which involves subtracting the global mean from each variable, class centering subtracts the mean of each class. This allows investigation of intra-class variation by removing the effects of inter-class variation [55]. The advantage of using linear models rather than class centering is that the effect of multiple fixed effects or classes can be modelled simultaneously.

The approach we took was based on the principals of quantitative genetic models, where

$$\text{Phenotypic observation} = \text{environmental effects} + \text{genetic effects} + \text{residual effects} \quad (1)$$

In this study, we only want to remove the effect of environmental factors (as it is the variation in NMR spectra under genetic influence that we are interested in), so the equation can be further simplified to

$$\text{Phenotypic observation} = \text{genetic effects} + \text{residual effects} \quad (2)$$

The "corrected phenotype" (i.e., the phenotypic observation with the effects of the environmental effects removed) is defined as the residuals from the above model. For the purposes of this study each chemical shift was treated as a separate phenotype, with the signal intensity at each chemical shift being an individual phenotypic observation. The "corrected spectra" was a matrix of the residuals of each model.

A 707 × 24,349 matrix of signal intensities of pre-processed spectra was imported into the R statistical software package v 3.6.2 [56]. Each row in the matrix represented a single sample, and each column represented 1 of the 24,349 chemical shifts between δ 0.40 and δ 8.99 that made up an individual spectrum. The following 4 linear models were applied to each of the 24,349 columns in the matrix (i.e., the signal intensity at each chemical shift was treated as the response variable in a separate regression model):

$$y_{il} = \mu + WIM_i + e_{il} \text{ (Model 1)} \quad (3)$$

$$y_{jl} = \mu + P_j + ej_{jl} \text{ (Model 2)} \quad (4)$$

$$y_{kl} = \mu + H_k + e_{kl} \text{ (Model 3)} \quad (5)$$

$$y_{ijkl} = \mu + WIM_i + P_j + H_k + e_{ijkl} \text{ (Model 4)} \quad (6)$$

where y is the signal intensity at a given chemical shift, μ is the mean, WIM is weeks in milk (4 levels, defined as 1, 2, 3, or 4), P is parity (4 levels, defined as 1, 2, 3, or ≥ 4), H is the effect of herd (13 levels, with a range of 9 to 248 cows per herd), and e is the random error term. This resulted in four separate 707 × 24,349 matrices containing spectra corrected for the effects of WIM, parity, herd, and all fixed effects, respectively.

The R^2 values from each regression model were stored in a separate vector. This resulted in four vectors each containing 24,349 R^2 values; each value representing the percentage of variation in signal intensity explained by the fixed effect(s) at a given chemical shift.

4.8. Quantifying the Effect of Stage of Lactation, Parity and Herd on ^{1}H NMR Spectra

A separate PCA was performed on each of the 4 corrected spectral datasets (as described in 4.7). Scores of the first three PCs were extracted for each model, and for the PCA model constructed using uncorrected data. We then calculated Pearson's correlations between scores derived from the 5 PCAs using the corrplot package [57] in R v 3.6.2 [56]. This resulted in three correlation matrices (one for each PC). The lower the Pearson's correlation coefficient, the greater the differences between PC scores, the greater the differences between the two spectral datasets and therefore the greater the significance of the fixed effect(s).

An alternative approach to investigating the influence of fixed effects is to use multiple linear regression on PC scores from uncorrected spectra. The advantage of this approach is that all fixed effects can be fitted simultaneously, and the statistical significance of each fixed effect can be calculated. The model used was

$$y_{ijkl} = \mu + WIM_i + P_j + H_k + e_{ijkl} \text{ (Model 5)} \quad (7)$$

where y is the PC score (on either PC1, PC2, or PC3) and μ, WIM, P, H, and e are the mean, fixed effect, and error terms described previously. The statistical significance of each fixed effect was determined using conditional Wald F statistics in ASReml v 4.2 (VSN International Ltd., Hemel Hempstead, UK). Conditional F statistics are used in multiple linear regression to infer the significance of a given fixed effect assuming that the effect of remaining predictor variables have been accounted for [58].

Finally, we validated our results using the analysis of variance (ANOVA) simultaneous component analysis (ASCA) method in the PLS Toolbox [55]. ASCA is a generalization of ANOVA used to quantify the variation induced by fixed experimental design factors on complex multivariate datasets [59]. ASCA was performed on all spectral datasets (corrected and uncorrected). Statistical significance was determined using permutation testing (50 iterations).

4.9. The Relationships between ^{1}H NMR Spectra and Existing Energy Balance Biomarker Concentrations

In order to assess the utility of large and diverse datasets in livestock metabolomics studies, we used orthogonal partial least squares (OPLS) regression to compare ^{1}H NMR spectra to serum BHBA concentrations determined by colorimetric assay. The aims of this analysis were (1) to assess the robustness of OPLS models built using uncorrected data and (2) investigate the influence of systematic environmental effects on the interpretation of ^{1}H NMR spectra when used for untargeted metabolomic analyses.

4.9.1. Robustness of OPLS Models to Predict External Phenotypes Using Uncorrected Data

The robustness of OPLS models constructed using large and diverse datasets was assessed using a leave-one-farm-out external validation. This involved setting aside data from one farm, training OPLS models using data from the remaining 12 farms, then using the withheld data for external validation. This process was repeated until data from each farm was used as an external validation set once. Model performance was assessed using the R^2 and RMSE of calibration, cross validation (venetian blind CV with 10 data splits, and one sample per split), and external validation. The statistical significance of OPLS models was determined using permutation testing (cross validated, Wilcoxon test). Only uncorrected data were used for this part of the analysis.

4.9.2. Influence of Fixed Effects on Interpretation of ^{1}H NMR Metabolomic Data

To assess the impact of fixed effects on the results of untargeted metabolomic analyses we compared the results of OPLS models constructed from (1) uncorrected data from Farm 1 only

(N = 129), (2) uncorrected data from all farms, and (3) data from all farms corrected for all fixed effects (Model 4). Farm 1 data was used to simulate a more "typical" metabolomics experiment in which confounding from environmental effects is controlled through experimental design.

When corrected spectra were used, reference BHBA concentrations were corrected for the same fixed effects (Model 4). The residuals of this model represent the "corrected BHBA" concentration, which is the expected BHBA concentration of an individual accounting for differences in WIM, Parity and Herd. This poses some challenges in terms of interpretation, as negative residual values (i.e., negative BHBA concentrations) are possible. However, for the purposes of genetic evaluations, the ranking of an animal, or the relative phenotypic value, is of more interest than an absolute value. The corrected value can therefore be considered a "corrected phenotypic ranking."

The impact of fixed effects on the ability of NMR spectra to predict external phenotypes (i.e., to classify animals or predict biomarker concentrations for management purposes) was assessed by comparing the predictive ability of OPLS models. The influences of fixed effects on biomarker discovery were investigated using scores and loadings on LV1 which show the magnitude and direction of relationships between BHBA concentration and ^{1}H NMR spectral features. Variable importance of projection (VIP) scores were used to identify the most statistically significant spectral features in each model. Variables with VIP scores greater than one were considered significant [60].

5. Conclusions

In this study we investigated the feasibility of using large and diverse datasets for untargeted ^{1}H NMR serum metabolomic profiling of clinically healthy dairy cows in early lactation. In particular, we investigated the effects of systematic environmental factors on the serum metabolome. We used linear regression to correct spectra for (1) herd of origin; (2) parity; (3) WIM; and (4) herd, parity, and WIM simultaneously. Corrected and uncorrected spectra were then analyzed using PCA. Comparison of PCA results showed that herd of origin had a much greater impact on the serum metabolome than either parity or WIM. In order to simulate the impact of these effects in untargeted metabolomics, we used OPLS regression to quantify the relationship between both corrected and uncorrected NMR spectra, and the current gold-standard biomarker of energy balance in dairy cows, BHBA. Our results showed that (1) models constructed using uncorrected data from multiple farms provided reasonably robust predictions of serum BHBA concentration, (2) environmental effects can alter the results of biomarker discovery, and (3) that correcting spectra for environmental effects using linear regression may be useful when the aim of analysis is to investigate phenotypic variation free of confounding from environmental effects (e.g., identification of metabotypes for genetic selection).

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-1989/10/5/180/s1, Table S1. ^{1}H NMR chemical shifts (δ) and multiplicity of metabolites in bovine serum run in deuterated water (D2O).; Table S2. Results of ANOVA-simultaneous component analysis (ASCA) of uncorrected ^{1}H NMR spectra of bovine serum.; Figure S1. Representative 700MHz ^{1}H NMR spectrum (δ 0.4 to 9.0) of serum obtained from a Holstein–Friesian cow in early lactation.; Figure S2. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for weeks in milk using linear regression; (a) PC 1 vs. PC 2 scores, (b) PC 1 vs. PC 3 scores, (c) PC 2 vs. PC 3 scores, (d) PC 1 loadings, (e) PC 2 loadings, and (f) PC 3 loadings plots.; Figure S3. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for Parity using linear regression; (a) PC 1 vs. PC 2 scores, (b) PC 1 vs. PC 3 scores, (c) PC 2 vs. PC 3 scores, (d) PC 1 loadings, (e) PC 2 loadings, and (f) PC 3 loadings plots.; Figure S4. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for Herd using linear regression; (a) PC 1 vs. PC 2 scores, (b) PC 1 vs. PC 3 scores, (c) PC 2 vs. PC 3 scores, (d) PC 1 loadings, (e) PC 2 loadings, and (f) PC 3 loadings plots.; Figure S5. Average ^{1}H NMR spectrum of bovine serum. Color-coding represents the percentage of variation in the signal at each chemical shift intensity that can be explained by (a) WIM and (b) Parity: Figure S6: Results of OPLS regressions of serum BHBA concentration against ^{1}H NMR spectrum of bovine serum (n = 707): (a) LV1 vs. LV2 scores for uncorrected data (b) CV predicted vs. measured BHBA (c) LV1 vs. LV2 scores for corrected data (d) CV predicted vs. measured corrected BHBA ranking.

Author Contributions: Conceptualization, T.D.W.L., J.E.P., W.J.W., and S.J.R.; Formal analysis, T.D.W.L. and S.J.R.; Funding acquisition, J.E.P.; Investigation, T.D.W.L., A.C.E., and S.J.R.; Methodology, T.D.W.L., A.C.E., and S.J.R.; Project administration, J.E.P., W.J.W., and S.J.R.; Resources, S.J.R.; Supervision, J.E.P., W.J.W., and S.J.R.; Visualization, T.D.W.L.; Writing—original draft, T.D.W.L.; Writing—review and editing, J.E.P., A.C.E., W.J.W., and S.J.R. All authors have read and agreed to the published version of the manuscript.

Funding: DairyBio: jointly funded by Dairy Australia (Melbourne, Australia), the Gardiner Foundation (Melbourne, Australia), and Agriculture Victoria (Melbourne, Australia) funded this study and T.D.W.L.'s PhD project.

Acknowledgments: The authors thank Simone Vassiliadis for assistance with ^{1}H NMR protocol and metabolite identification, Di Mapleson, Brigid Ribaux and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance, Erika Oakes and the staff at Datagene (Bundoora, Australia) for their work coordinating this study, and the farmers who took part in this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wishart, D.S. Metabolomics: Applications to food science and nutrition research. *Trends Food Sci. Technol.* **2008**, *19*, 482–493. [CrossRef]
2. Drackley, J.K. Biology of dairy cows during the transition period: The final frontier? *J. Dairy Sci.* **1999**, *82*, 2259–2273. [CrossRef]
3. Grummer, R.R. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. *J. Anim. Sci.* **1995**, *73*, 2820–2833. [CrossRef]
4. LeBlanc, S.J.; Lissemore, K.D.; Kelton, D.F.; Duffield, T.F.; Leslie, K.E. Major advances in disease prevention in dairy cattle. *J. Dairy Sci.* **2006**, *89*, 1267–1279. [CrossRef]
5. Houle, D.; Govindaraju, D.; Omholt, S. Phenomics: The next challenge. *Nat. Rev. Gen.* **2010**, *11*, 855–866. [CrossRef]
6. Daetwyler, H.D.; Xiang, R.; Yuan, Z.; Bolormaa, S.; Vander Jagt, C.J.; Hayes, B.J.; van der Werf, J.H.J.; Pryce, J.E.; Chamberlain, A.J.; Macleod, I.M.; et al. Integration of functional genomics and phenomics into genomic prediction raises its accuracy in sheep and dairy cattle. In Proceedings of the Association for the Advancement of Animal Breeding and Genetics, Armidale, NSW, Australia, 27 October–1 November 2019; pp. 11–14.
7. Xiang, R.; Berg, I.V.D.; Macleod, I.M.; Hayes, B.J.; Prowse-Wilkins, C.P.; Wang, M.; Bolormaa, S.; Liu, Z.; Rochfort, S.J.; Reich, C.M.; et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19398–19408. [CrossRef]
8. Sun, L.W.; Zhang, H.Y.; Wu, L.; Shu, S.; Xia, C.; Xu, C.; Zheng, J.S. 1H-Nuclear magnetic resonance-based plasma metabolic profiling of dairy cows with clinical and subclinical ketosis. *J. Dairy Sci.* **2014**, *97*, 1552–1562. [CrossRef]
9. Zhang, G.; Dervishi, E.; Dunn, S.; Mandal, R.; Liu, P.; Han, B.; Wishart, D.; Ametaj, B. Metabotyping reveals distinct metabolic alterations in ketotic cows and identifies early predictive serum biomarkers for the risk of disease. *Metabolomics* **2017**, *13*, 1–15. [CrossRef]
10. Sun, Y.; Xu, C.; Li, C.; Cheng, X.; Xu, C.; Wu, L.; Zhang, H. Characterization of the serum metabolic profile of dairy cows with milk fever using (1)H-NMR spectroscopy. *Vet. Quart.* **2014**, *34*, 1–18. [CrossRef]
11. Basoglu, A.; Baspinar, N.; Tenori, L.; Licari, C.; Gulersoy, E. Nuclear magnetic resonance (NMR)-based metabolome profile evaluation in dairy cows with and without displaced abomasum. *Vet. Quart.* **2020**, *40*, 1–15. [CrossRef]
12. Hailemariam, D.; Mandal, R.; Saleem, F.; Dunn, S.M.; Wishart, D.S.; Ametaj, B.N. Identification of predictive biomarkers of disease state in transition dairy cows. *J. Dairy Sci.* **2014**, *97*, 2680–2693. [CrossRef] [PubMed]
13. Goldansaz, S.A.; Guo, A.C.; Sajed, T.; Steele, M.A.; Plastow, G.S.; Wishart, D.S. Livestock metabolomics and the livestock metabolome: A systematic review. *PLoS ONE* **2017**, *12*, e0177675. [CrossRef] [PubMed]
14. Sacco, D.; Brescia, M.A.; Sgaramella, A.; Casiello, G.; Buccolieri, A.; Ogrinc, N.; Sacco, A. Discrimination between southern Italy and foreign milk samples using spectroscopic and analytical data. *Food Chem.* **2009**, *114*, 1559–1563. [CrossRef]
15. Tenori, L.; Santucci, C.; Meoni, G.; Morrocchi, V.; Matteucci, G.; Luchinat, C. NMR metabolomic fingerprinting distinguishes milk from different farms. *Food Res. Int.* **2018**, *113*, 131–139. [CrossRef]

16. Sundekilde, U.K.; Frederiksen, P.D.; Clausen, M.R.; Larsen, L.B.; Bertram, H.C. Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. *J. Agric. Food Chem.* **2011**, *59*, 7360–7367. [CrossRef]
17. Ilves, A.; Harzia, H.; Ling, K.; Ots, M.; Soomets, U.; Kilk, K. Alterations in milk and blood metabolomes during the first months of lactation in dairy cows. *J. Dairy Sci.* **2012**, *95*, 5788–5797. [CrossRef]
18. Maher, A.D.; Hayes, B.; Cocks, B.; Marett, L.; Wales, W.J.; Rochfort, S.J. Latent biochemical relationships in the blood-milk metabolic axis of dairy cows revealed by statistical integration of 1H NMR spectroscopic data. *J. Proteome Res.* **2013**, *12*, 1428. [CrossRef]
19. Mrode, R.A. *Linear Models for the Prediction of Animal Breeding Values*, 3rd ed.; CABI: Wallingford, UK, 2014.
20. Wanichthanarak, K.; Jeamsripong, S.; Pornputtapong, N.; Khoomrung, S. Accounting for biological variation with linear mixed-effects modelling improves the quality of clinical metabolomics data. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 611–618. [CrossRef]
21. Laine, A.; Bastin, C.; Grelet, C.; Hammami, H.; Colinet, F.G.; Dale, L.M.; Gillon, A.; Vandenplas, J.; Dehareng, F.; Gengler, N. Assessing the effect of pregnancy stage on milk composition of dairy cows using mid-infrared spectra. *J. Dairy Sci.* **2017**, *100*, 2863–2876. [CrossRef]
22. Oetzel, G.R. Monitoring and testing dairy herds for metabolic disease. *Vet. Clin. North Am. Food Anim. Pr.* **2004**, *20*, 651–674. [CrossRef]
23. Ospina, P.A.; Nydam, D.V.; Stokol, T.; Overton, T.R. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation reproductive performance and milk production in transition dairy cattle in the northeastern United States. *J. Dairy Sci.* **2010**, *93*, 1596–1603. [CrossRef] [PubMed]
24. Lean, I.; Westwood, C.T.; Playford, M.C. Livestock disease threats associated with intensification of pastoral dairy farming. *N. Z. Vet. J.* **2008**, *56*, 261–269. [CrossRef] [PubMed]
25. Schwaiger, T.; Beauchemin, K.A.; Penner, G.B. Duration of time that beef cattle are fed a high-grain diet affects the recovery from a bout of ruminal acidosis: Short-chain fatty acid and lactate absorption, saliva production, and blood metabolites. *J. Anim. Sci.* **2013**, *91*, 5743–5753. [CrossRef] [PubMed]
26. Yang, Y.; Dong, G.; Wang, Z.; Wang, J.; Zhang, Z.; Liu, J. Rumen and plasma metabolomics profiling by UHPLC-QTOF/MS revealed metabolic alterations associated with a high-corn diet in beef steers. *PLoS ONE* **2018**, *13*, e0208031. [CrossRef] [PubMed]
27. Zhang, L.; Martins, A.F.; Zhao, P.; Tieu, M.; Esteban-Gómez, D.; McCandless, G.T.; Platas-Iglesias, C.; Sherry, A.D. Enantiomeric recognition of d- and l-lactate by CEST with the aid of a paramagnetic shift reagent. *J. Am. Chem. Soc.* **2017**, *139*, 17431–17437. [CrossRef] [PubMed]
28. Lees, H.J.; Swann, J.R.; Wilson, I.D.; Nicholson, J.K.; Holmes, E. Hippurate: The natural history of a mammalian–microbial cometabolite. *J. Proteome Res.* **2013**, *12*, 1527–1546. [CrossRef] [PubMed]
29. Carpio, A.; Bonilla-Valverde, D.; Arce, C.; Rodríguez-Estévez, V.; Sánchez-Rodríguez, M.; Arce, L.; Valcárcel, M. Evaluation of hippuric acid content in goat milk as a marker of feeding regimen. *J. Dairy Sci.* **2013**, *96*, 5426–5434. [CrossRef]
30. Pallister, T.; Jackson, M.A.; Martin, T.C.; Zierer, J.; Jennings, A.; Mohney, R.P.; MacGregor, A.; Steves, C.J.; Cassidy, A.; Spector, T.D.; et al. Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Sci. Rep.* **2017**, *7*, 13670. [CrossRef]
31. Liao, Y.; Hu, R.; Wang, Z.; Peng, Q.; Dong, X.; Zhang, X.; Zou, H.; Pu, Q.; Xue, B.; Wang, L. Metabolomics profiling of serum and urine in three beef cattle breeds revealed different levels of tolerance to heat stress. *J. Agric. Food Chem.* **2018**, *66*, 6926–6935. [CrossRef]
32. Yin, P.; Lehmann, R.; Xu, G. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal. Bioanal. Chem.* **2015**, *407*, 4879–4892. [CrossRef]
33. Beckonert, O.; Keun, H.C.; Ebbels, T.M.; Bundy, J.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2*, 2692–2703. [CrossRef] [PubMed]
34. Emwas, A.-H.; Luchinat, C.; Turano, P.; Tenori, L.; Roy, R.; Salek, R.M.; Ryan, D.; Merzaban, J.S.; Kaddurah-Daouk, R.; Zeri, A.C.; et al. Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: A review. *Metabolomics* **2015**, *11*, 872–894. [CrossRef] [PubMed]

35. Jobard, E.; Trédan, O.; Postoly, D.; André, F.; Martin, A.-L.; Elena-Herrmann, B.; Boyault, S. A systematic evaluation of blood serum and plasma pre-analytics for metabolomics cohort studies. *Int. J. Mol. Sci.* **2016**, *17*, 2035. [CrossRef]
36. Klein, M.S.; Almstetter, M.F.; Schlamberger, G.; Nürnberger, N.; Dettmer, K.; Oefner, P.J.; Meyer, H.H.D.; Wiedemann, S.; Gronwald, W. Nuclear magnetic resonance and mass spectrometry-based milk metabolomics in dairy cows during early and late lactation. *J. Dairy Sci.* **2010**, *93*, 1539–1550. [CrossRef] [PubMed]
37. Cozzi, G.; Ravarotto, L.; Gottardo, F.; Stefani, A.L.; Contiero, B.; Moro, L.; Brscic, M.; Dalvit, P. Short communication: Reference values for blood parameters in holstein dairy cows: Effects of parity, stage of lactation, and season of production. *J. Dairy Sci.* **2011**, *94*, 3895–3901. [CrossRef] [PubMed]
38. Luke, T.D.W.; Rochfort, S.; Wales, W.J.; Bonfatti, V.; Marett, L.; Pryce, J.E. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. *J. Dairy Sci.* **2019**, *102*, 1747–1760. [CrossRef]
39. Morales Pineyrua, J.T.; Farina, S.R.; Mendoza, A. Effects of parity on productive, reproductive, metabolic and hormonal responses of Holstein cows. *Anim. Reprod. Sci.* **2018**, *191*, 9–21. [CrossRef]
40. Davies, A.; Fearn, T. Back to basics: Calibration statistics. *Spectrosc. Eur.* **2006**, *18*, 31–32.
41. Posma, J.M.; Garcia-Perez, I.; Ebbels, T.M.D.; Lindon, J.C.; Stamler, J.; Elliott, P.; Holmes, E.; Nicholson, J.K. Optimized phenotypic biomarker discovery and confounder elimination via covariate-adjusted projection to latent structures from metabolic spectroscopy data. *J. Proteome Res.* **2018**, *17*, 1586–1595. [CrossRef]
42. Lean, I.J.; Farver, T.B.; Troutt, H.F.; Bruss, M.L.; Galland, J.C.; Baldwin, R.L.; Holmberg, C.A.; Weaver, L.D. Time series cross-correlation analysis of postparturient relationships among serum metabolites and yield variables in Holstein cows. *J. Dairy Sci.* **1992**, *75*, 1891–1900. [CrossRef]
43. Aschenbach, J.R.; Kristensen, N.B.; Donkin, S.S.; Hammon, H.M.; Penner, G.B. Gluconeogenesis in dairy cows: The secret of making sweet milk from sour dough. *IUBMB Life* **2010**, *62*, 869–877. [CrossRef] [PubMed]
44. Drackley, J.K.; Overton, T.R.; Douglas, G.N. Adaptations of glucose and long-chain fatty acid metabolism in liver of dairy cows during the periparturient period. *J. Dairy Sci.* **2001**, *84*, E100–E112. [CrossRef]
45. Van den Berg, R.A.; Hoefsloot, H.C.J.; Westerhuis, J.A.; Smilde, A.K.; van der Werf, M.J. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genom.* **2006**, *7*, 142. [CrossRef] [PubMed]
46. Emwas, A.-H.; Saccenti, E.; Gao, X.; McKay, R.; Santos, V.; Roy, R.; Wishart, D. Recommended strategies for spectral processing and post-processing of 1D 1 H-NMR data of biofluids with a particular focus on urine. *Off. J. Metab. Soc.* **2018**, *14*, 1–23. [CrossRef]
47. Fontanesi, L. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Anim. Front.* **2016**, *6*, 73–79. [CrossRef]
48. Boichard, D.; Brochard, M. New phenotypes for new breeding goals in dairy cattle. *Anim. Int. J. Anim. Biosci.* **2012**, *6*, 544–550. [CrossRef]
49. Egger-Danner, C.; Cole, J.B.; Pryce, J.E.; Gengler, N.; Heringstad, B.; Bradley, A.; Stock, K.F. Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Anim. Int. J. Anim. Biosci.* **2015**, *9*, 191–207. [CrossRef]
50. De Haas, Y.; Pryce, J.E.; Calus, M.P.; Wall, E.; Berry, D.P.; Lovendahl, P.; Krattenmacher, N.; Miglior, F.; Weigel, K.; Spurlock, D.; et al. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. *J. Dairy Sci.* **2015**, *98*, 6522–6534. [CrossRef]
51. McMurray, C.H.; Blanchflower, W.J.; Rice, D.A. Automated kinetic method for D-3-hydroxybutyrate in plasma or serum. *Clin. Chem.* **1984**, *30*, 421–425. [CrossRef]
52. Gowda, N.G.A.; Raftery, D. Quantitating metabolites in protein precipitated serum using NMR spectroscopy. *Anal. Chem.* **2014**, *86*, 5433–5440. [CrossRef]
53. Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943–948. [CrossRef] [PubMed]
54. Nielsen, N.-P.V.; Carstensen, J.M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr.* **1998**, *805*, 17–35. [CrossRef]
55. Eigenvector. *PLS Toolbox, R2017b*; Eigenvector Research Inc.: Manson, WA, USA, 2017.
56. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

57. Wei, T.; Simko, V. R Package "Corrplot": Visualization of a Correlation Matrix, (Version 0.84). Available online: https://github.com/taiyun/corrplot (accessed on 29 April 2020).
58. Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Thompson, R. *ASReml User Guide Release 4.1 Functional Specification*, *4.1*; HP1 1ES; VSN International Ltd.: Hemel Hempstead, UK, 2015; p. 67.
59. Smilde, A.; Jansen, J.; Hoefsloot, H.; Lamers, R.-J.; van Der Greef, J.; Timmerman, M. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics* **2005**, *21*, 3043–3048. [CrossRef]
60. Chong, I.-G.; Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intellig. Lab. Syst.* **2005**, *78*, 103–112. [CrossRef]

Chapter 6:

Towards a Breeding Value for Improved Metabolic Health

Chapter 6: Towards a Breeding Value for Improved Metabolic Health

The importance of maximising metabolic health in early lactation dairy cows is widely accepted, and is aptly illustrated by the fact that (1) 75% of all adverse health events and 80% of all dairy farm health expenditure occur during the first four weeks of lactation (LeBlanc et al., 2006; Lean and Degaris, 2010), and (2) many of these adverse health events are closely associated with failure of homeorhetic controls and perturbed metabolism (Drackley, 1999). To date, much of the research on improving dairy cow metabolic health has focussed on nutritional and management solutions (Overton and Waldron, 2004), however, there is increasing interest in improving dairy cow metabolic health through genetic selection (Pryce et al., 2016). While traditional genetic evaluations have enabled some genetic improvement in metabolic health, the only phenotypes available on a sufficiently large scale have been producer- and/or veterinary-recorded clinical health records. Typically, such records lack the resolution required for accurate and objective capture of the physiological and biochemical complexity that is characteristic of metabolic disorders in early lactation. More recently, genomic selection and the increasingly widespread availability of single nucleotide polymorphism (SNP) genotypes are facilitating genetic improvement in economically important traits that are lowly heritable, and difficult and costly to measure (Calus et al., 2013; Chesnais et al., 2016). The aim of the work described in this thesis was to investigate novel metabolic phenotypes for use in genomic selection for improved metabolic health in early lactation.

6.1 Review of important findings

The results presented in Chapter 2 demonstrate that, with the exception of haptoglobin, "traditional" serum metabolic profile biomarkers (BHBA, NEFA, Ca, Mg, urea, albumin, globulin and albumin to globulin ratio) are heritable traits ($0.07 \leq h^2 \leq 0.41$), and that genomic selection for more optimal biomarker concentrations is possible. Of particular interest were the favourable trends in genetic relationships observed between biomarker traits, and between biomarker GEBVs and existing health breeding values. These findings suggest that selecting for overall metabolic resilience/stability should be possible, and that selecting for improved metabolic health should improve the overall health, fertility and longevity of dairy cows. Of the biomarkers investigated, NEFA, albumin and albumin to globulin ratio (A:G) were particularly promising owing to (1) relatively high heritability estimates and genomic prediction accuracies (2) favourable genetic correlations with other biomarkers, and (3) favourable correlations with breeding values for other health traits.

In Chapter 3, the use of milk mid-infrared (MIR) spectral data to predict the serum concentrations of the aforementioned biomarkers is discussed. Prediction accuracies were

found to be moderate for serum BHBA, NEFA and urea concentrations, and poor for the remaining metabolites. Of particular interest were the results obtained from a true external validation (which used data collected in a different season, from cows of different breeds managed under a different production system), which suggested that BHBA and NEFA predictions were relatively robust. Partial least squares discriminant analysis (PLS-DA) models were used to classify animals as being affected or not affected with subclinical metabolic disorders, based on biomarker thresholds obtained from the literature. Overall, the results from this chapter indicate that while milk MIR-predictions of serum BHBA, NEFA and urea concentrations are not currently accurate enough to form the basis of individual animal diagnostic tests, or management decision making tools, their use as indicator traits in genetic selection for improved metabolic health warrants further investigation.

The final two research chapters contain description of investigation into use of proton nuclear magnetic resonance (^{1}H NMR) spectroscopy-based metabolomics to better (1) better characterise existing serum biomarkers of energy balance, and (2) identify novel intermediate metabolic phenotypes (metabotypes) for use in genetic analysis. In chapter 4, there is description of an untargeted metabolomics approach to investigate if differences exist in the metabolomic fingerprints of BHBA and NEFA, which could help to explain the weak genetic and phenotypic correlations between them. Results indicate that while BHBA and NEFA represent similar metabolic states, there are significant differences between the states, in acetate and creatine metabolism. Furthermore, 16 intermediate metabolites were identified, which were primarily intermediates of energy, phospholipid, and/or methyl donor metabolism. The ability to predict serum NEFA concentration (a "non-NMR-observable" feature) from NMR spectra was another interesting finding that warrants further investigation.

Finally, in Chapter 5, there is description of the feasibility of applying an untargeted metabolomics approach, to a dataset that is sufficiently large and diverse to be useful for identifying metabotypes for use in genetic evaluations. Results of this study indicated that herd-specific factors had large and significant effects on the serum metabolome. Furthermore, these effects could potentially influence the results of metabotype discovery by erroneously attributing herd-specific environmental effects to inter-animal variation. A method to correct ^{1}H NMR spectra for fixed environmental and physiological effects using multiple linear regression was proposed, and this enabled identification of metabotypes indicative of inter-animal variation free of confounding environmental effects.

6.2 Towards a reliable breeding value for metabolic resilience…

The findings of this thesis provide further evidence that metabolic phenotypes are likely to be of great value in the development of more accurate breeding values for improved

metabolic health. For the purposes of this discussion, let us consider a metabolic resilience index which consists of the 8 traditional biomarker traits identified in Chapter 2 as having genetic variance (BHBA, NEFA, Ca, Mg, urea, albumin, globulin and A:G). Assuming equal weighting of all traits, the average heritability of metabolic resilience would be approximately 0.23. The theoretical genomic prediction reliability of this index is approximately 0.17, which while promising and commensurate with the small reference population size (Gonzalez-Recio et al., 2014), limits its practical utility in its current form. The focus of the rest of this discussion is therefore how the findings of this thesis might be applied to develop a breeding value for metabolic resilience that is sufficiently reliable to be adopted by the Australian dairy industry.

6.2.1 How reliable does a breeding value need to be?

The FeedSaved Australian Breeding Value (ABV) provides an excellent template for the design and implementation of a breeding value for an economically important trait that is expensive and difficult to measure (Pryce et al., 2015). FeedSaved was introduced in 2015 and includes a residual feed intake (RFI) component, and a maintenance requirement component calculated from body weight estimated breeding values (EBV). Of particular interest in the context of this thesis is the RFI component of FeedSaved, which was developed using a genotyped female reference population (N = 2,036) of comparable size to the one used in Chapter 2 of this thesis (N = 1,393). The estimated genomic heritability and reliability of the RFI trait were 0.20 and 0.06, respectively. The average reliability of the FeedSaved breeding value in genotyped Holstein bulls is 0.37, which we will take as our benchmark for adoption of a breeding value by the Australian dairy industry.

6.3 How can the findings of this thesis be applied to achieve a sufficiently reliable breeding value?

The accuracy of genomic selection is influenced by several factors, including (1) the effective population size, (2) the size of the reference population, (3) the heritability of the trait, and (4) the genetic architecture of the trait (Daetwyler et al., 2008; Meuwissen, 2009; Hayes et al., 2010; Gonzalez-Recio et al., 2014). The novel metabolic phenotyping strategies investigated in this thesis offer exciting opportunities to improve the accuracy of genomic selection for improved metabolic health by:

1. Sustainably and cost-effectively increasing the size of the reference population
2. Improving heritability estimates by refining trait definitions
3. Facilitating identification of functional variants

These concepts are summarised in Figure 1.

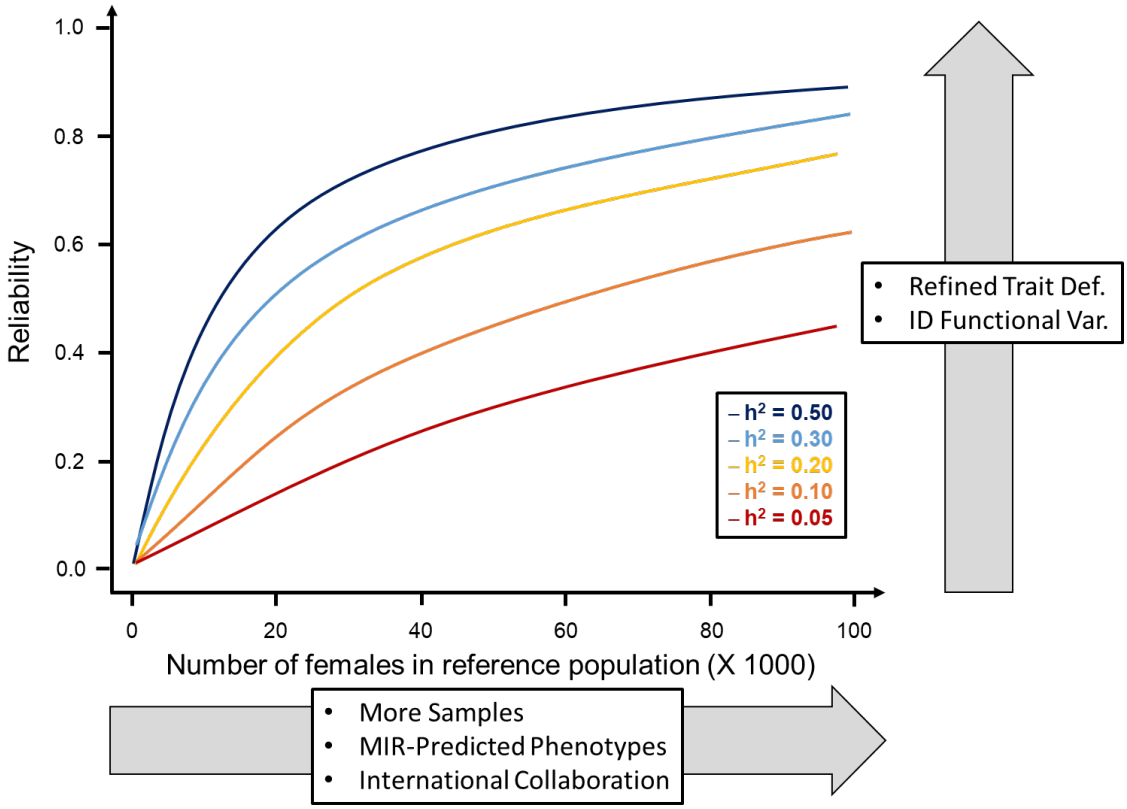


Figure 1. Opportunities to improve the accuracy of genomic selection for improved metabolic health using novel metabolic phenotyping strategies. Adapted from Gonzalez-Recio et al. (2014)

6.3.1 How can we cost-effectively and sustainably increase reference population size?

Perhaps the most obvious way to improve genomic breeding value reliability is to increase the size of the reference population. According to Gonzalez-Recio et al. (2014), for a trait with a heritability of 0.2, achieving a genomic prediction reliability of 0.37 would require a female reference population of approximately 20,000 animals. While not impossible, obtaining serum samples from 20,000 animals in early lactation is logistically challenging, costly and relatively invasive, so alternatives should be considered.

The moderately accurate yet robust MIR-predictions of serum BHBA, NEFA and urea reported in Chapter 3 offer further evidence that metabolic phenotypes predicted from milk MIR spectra are likely to be a valuable way of cost-effectively increasing the size of the reference population for some traits. For example, MIR-predicted biomarker concentrations could be used as predictor traits for serum biomarker phenotypes, in a similar way that somatic cell count (SCC) is used as a predictor trait for clinical mastitis (CM) (Mrode and Swanson, 1996; Abdelsayed et al., 2017). This approach requires a relatively strong and favourable genetic correlation between the traits (e.g. Abdelsayed et al. (2017) reported that the genetic correlation between CM and SCC in Australian dairy cattle is 0.55 ± 0.11). Promisingly, Belay et al. (2017) reported the genetic correlation between blood BHBA concentration predicted

from milk MIR spectra between 11 and 30 days in milk (DIM), and clinical ketosis, to be 0.46 ± 0.05. While lower than the genetic correlation between clinical ketosis and serum BHBA and acetone concentrations (0.68 and 0.78, respectively) (Rius-Vilarrasa et al., 2018), this is still a very promising result. Furthermore, unpublished results from work undertaken as part of this thesis show promising genetic correlations between serum and milk MIR-predicted concentrations of BHBA, NEFA and urea (0.29 ± 0.18, 0.62 ± 0.18 and 0.79 ± 0.24, respectively).

An alternative approach could be to use MIR-predicted traits as the "gold standard" phenotype. In a recent epidemiological study, Bach et al. (2019) showed that MIR-predictions of milk BHBA, blood NEFA, and relative percentages of de novo milk fatty acids are promising indicators of subsequent disease or removal from the herd. Heritability estimates for MIR-predicted milk BHBA, serum BHBA and serum NEFA concentrations are 0.14 (Koeck et al., 2014), 0.25-0.31 (Belay et al., 2017; Benedet et al., 2020), and 0.19 (Benedet et al., 2020), respectively). Given that these MIR-predicted traits are both heritable, and epidemiologically associated with similar adverse health outcomes to their "gold-standard" serum counterparts, it may make more sense to use these as the primary traits in genetic evaluations for improved metabolic health.

Another demonstrated way of increasing the size of reference populations is through data sharing and international collaboration. Again, the FeedSaved breeding value offers a perfect example, as nearly half of all the RFI phenotypes came from the UK and the Netherlands (Pryce et al., 2015). Similarly, de Haas et al. (2015) demonstrated that integrating phenotypic information for dry matter intake (DMI) from multiple countries improved the accuracy of genomic evaluations of individual countries. The objective, measurable characteristics of the metabolic phenotypes examined in this thesis make them ideal for such data sharing. It is therefore highly likely that augmenting reference populations with data from other countries will lead to similar increases in genomic prediction accuracies for metabolic health traits.

Overall, by using MIR-predicted phenotypes, and by sharing data with international collaborators, it should be possible to assemble a reference population that is large enough to allow development of a breeding value for metabolic resilience that is sufficiently reliable to be implemented by the Australian dairy industry.

6.3.2 How can refining the definitions of metabolic health traits increase heritability estimates?

It may be possible to increase heritability estimates for some of the metabolic traits discussed in this thesis by further refining trait definitions. BHBA and NEFA concentrations are perfect examples of such traits. Previous studies have shown that the heritability of both

serum and MIR-predicted BHBA and NEFA concentrations vary depending on the stage of lactation (Oikonomou et al., 2008, Koeck et al., 2014, Benedet et al., 2020). For example, Oikonomou et al. (2008) reports that the heritabilities of serum BHBA and NEFA concentrations are highest in the first week of lactation (0.40 ± 0.06 and 0.35 ± 0.05, respectively), and decrease rapidly to less than 0.20 in the fourth week of lactation (Figure 2). These differences in heritability are largely due to differences in genetic variances, which follow a similar trajectory (Figure 2) (Oikonomou et al., 2008). Furthermore, genetic correlations between metabolite concentrations in different weeks, decreased for weeks that were further apart, indicating that measurements taken at different time points represent distinct traits. Therefore, by narrowing the window of sampling to the first two weeks of lactation, it should be possible to considerably improve the heritabilities of both BHBA and NEFA reported in Chapter 2 (0.09 ± 0.04 and 0.18 ± 0.05, respectively).

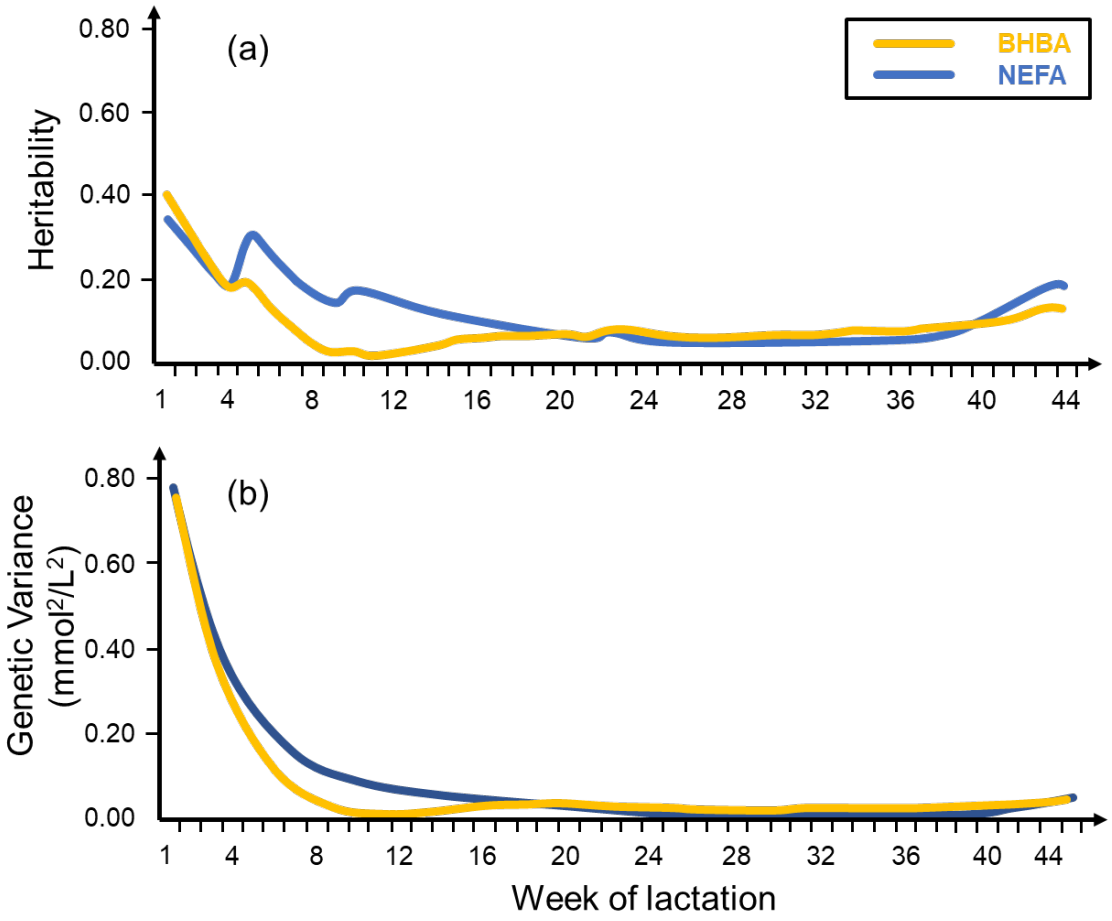


Figure 3. Estimated (a) heritability and (b) genetic variance of serum BHBA and NEFA concentrations by week of lactation in primiparous dairy cows. Adapted from Oikonomou et al. (2008)

The same is likely to be true for Ca concentration. There is evidence in the literature that there is genetic variation in clinical hypocalcaemia (Pryce et al., 2016), subclinical hypocalcaemia (Tsiamadis et al., 2016b), and serum calcium concentration (Tsiamadis et al.,

2016a). The low heritability of calcium concentration reported in Chapter 2 is almost certainly a reflection of the small number of samples taken in the high-risk, immediate periparturient period. Tsiamadis et al. (2016a) reports that the heritability of Ca concentration across the first 8 days of lactation is 0.20 (±0.02), ranging from 0.32 (±0.03) on day one, to 0.23 (±0.02) on day four. Only 256 animals in our dataset were sampled during this period, however, the estimated genomic heritability calculated using this data was 0.21 (±0.23), compared to 0.06 (±0.05) for the period between days 9 and 30 (N = 1071). As expected, both the genetic and residual variance appear to be higher in early lactation. Recent epidemiological studies undertaken by researchers at Cornell University have shown that associations between hypocalcaemic events and adverse health and production are highly dependent on both the timing and duration of hypocalcaemia (Neves et al., 2018; McArt and Neves, 2020). This work offers exciting potential trait definitions for future genetic analyses of hypocalcaemia.

Refining BHBA, NEFA and Ca trait definitions to achieve heritability estimates of 0.35, 0.30 and 0.20, respectively, would increase the overall heritability of our metabolic resilience index to approximately 0.30. This would halve the size of reference population required to achieve a genomic selection reliability of 0.36 from approximately 20,000 animals, to approximately 10,000 animals (Figure 1) (Gonzalez-Recio et al., 2014).

6.3.3 How can intermediate metabolite phenotypes be used to identify functional variants? And how might this help to improve the accuracies of genomic predictions?

One of the most exciting potential applications of metabolic phenotypes in animal breeding is the identification of functional variants through metabolite-base genome-wide association studies (mGWAS) (Fontanesi, 2016). In humans, such studies have helped to unravel the contribution of genetics to complex metabolic traits (Karsten and Christian, 2012). In livestock, there is increasing evidence that incorporating such functional genomic data can significantly improve the accuracy of genomic prediction (MacLeod et al., 2016; Daetwyler et al., 2019). Excitingly, Xiang et al. (2019) recently demonstrated that sequence variants identified from an mGWAS (using lipidomic data from 338 animals) were highly heritable, and that including these variants in SNP sets significantly increased genomic prediction accuracies for 34 complex traits in bovines.

Several recent studies have performed mGWAS on metabolic phenotypes investigated in this thesis. For example, mGWAS of serum BHBA and NEFA in early lactation dairy cows identified five candidate genes related to energy metabolism and homeostasis (Yepes et al., 2019), and Nayeri et al. (2019) identified several genomic regions associated with MIR-

predicted BHBA. It is highly likely that prioritising the variants identified in these studies will increase the accuracy of genomic predictions of BHBA and NEFA.

Another interesting finding reported by Xiang et al. (2019) was the value of intermediate phenotypes in identifying variants which have significant influences on complex traits. Intermediate phenotypes are so-called because they sit between the external phenotype and the genotype (Houle et al., 2010). For example, if we consider serum BHBA as an external phenotype, the 9 metabolites and 2 classes of protein (lipoproteins and glycoproteins) shown to be significantly correlated with BHBA in Chapter 4 can be considered intermediate phenotypes (Figure 3). Quantitative trait loci (QTL) associated with intermediate phenotypes often have a larger "signal to noise" ratio compared to QTL associated with external phenotypes, making them potentially more valuable for identifying true causal variants (Xiang et al., 2019). This is consistent with a recent GWAS study that demonstrated that metabolic clustering (a relatively "broad" ternary trait) is highly polygenic and regulated by many small effects (Atashi et al., 2020).

Furthermore, intermediate phenotypes offer more potential traits for mGWAS studies. For example, unpublished data from an mGWAS performed as a part of this thesis, using imputed whole genome sequence data, identified three genomic regions significantly ($-\log_{10}(p)$ > 6) associated with NEFA concentration. Subsequent GWAS analyses performed on three of the intermediate metabolites identified in Chapter 4 revealed significant associations with a further six genomic regions.

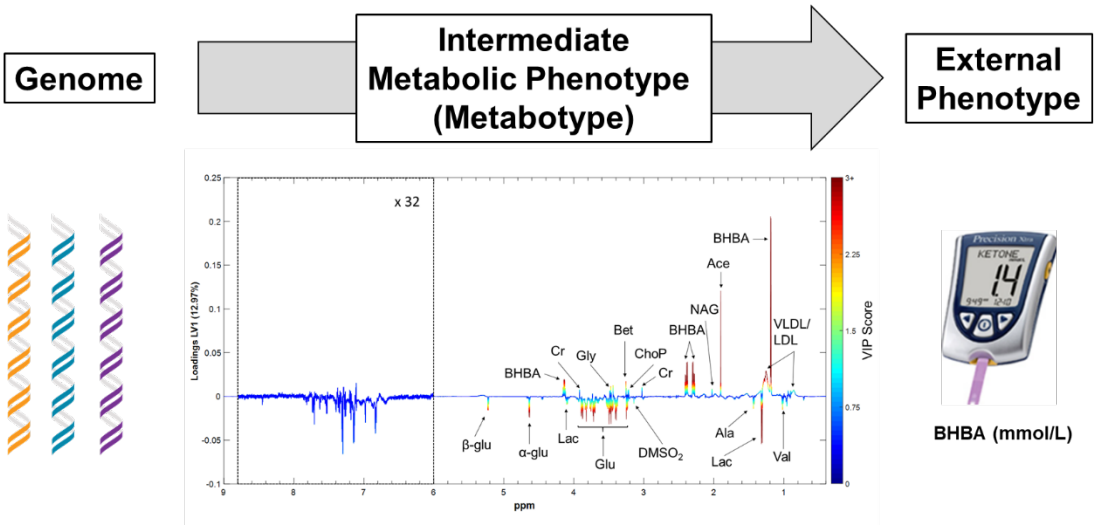


Figure 3. Intermediate metabolic phenotype (metabotype) of serum BHBA concentration, identified using ^{1}H NMR-based metabolomics. α-Glu = α glucose, β-Glu = β glucose, Ace = acetate, Ala = alanine, Bet = betaine, BHBA = β hydroxybutyrate, Cr = creatine, $DMSO_2$ = dimethyl sulfone, Glu = glucose, Gly = glycine, Ile = isoleucine, Lac = lactate, Leu = leucine, NAG = N-acetyl glycoprotein, ChoP = phosphocholine, Pyr = pyruvate, Val = valine, LDL = low density lipoprotein; VLDL = very low density lipoprotein.

Overall, it is highly likely that functional genomic data identified by mGWAS studies will facilitate the identification of variants associated with metabolic health traits. These variants can then be prioritised, either using statistical approaches such as BayesRC (MacLeod et al., 2016), or by creating custom variant sets (Xiang et al., 2019) that can be included on custom SNP chips such as the AgVic50kXT (Illumina, San Diego, CA, USA). This should ultimately increase genomic prediction accuracies for complex traits such as metabolic health.

6.4 What comes next?

6.4.1 More samples…

The findings of this thesis have helped narrow the sampling window for blood collections. Future sampling should focus on the first two weeks of lactation for energy balance traits, and the first four days of lactation for calcium traits. Even if heritability estimates can be improved through refined trait definitions, many more samples will be required to achieve a reliable breeding value for metabolic resilience. Furthermore, many of the genetic correlations reported in Chapter 2 were not significantly different from zero, and it is hoped that increasing the number of samples will lead to development of better understanding of the genetic architecture of traits.

6.4.2 Continued refinement of trait definitions

Perhaps the biggest challenge I encountered throughout this PhD project, was how exactly does one define a holistic metabolic health trait? While I hope the findings of this thesis have made a substantial contribution to answering this question, there is still much work to be done. In Australia, the dairy industry is working hard to improve the flow of animal health data from farms and veterinary clinics to genetic researchers through the implementation of a new online data repository, DataVat (DataGene, 2020). Such data will be essential in helping to define and validate metabolic health traits. Another obstacle is the lack of epidemiological studies on metabolic disorders in the Australian dairy herd. Such studies are required to (1) assess the prevalence of metabolic disorders, and thereby determine their economic cost, (2) identify appropriate thresholds for existing biomarkers, and (3) establish associations between biomarkers (both existing and novel), and adverse health, production and reproductive outcomes. This work is central to further our understanding of the genetic architecture of metabolic health traits, and therefore what traits should be included in any index.

Fertility has recently been identified as one of the most important breeding objectives for Australian dairy farmers. Given the strong epidemiological links between metabolic health and fertility, it may be that fertility data can be used as a form of validation for metabolic phenotypes. However, given that the risk of pregnancy in hyperketonaemic (BHBA $\geq$ 1.0

mmol/L) animals is only 16% less than in non-hyperketonaemic animals (Ospina et al., 2010a), this approach is likely to miss a lot of the important adverse effects of subclinical disease, such as reduced milk production and quality, and increased risk of other diseases (Ospina et al., 2010a; Ospina et al., 2010b; McArt et al., 2013).

Another option could be to define a successful transition period as the absence of any disorder, clinical or subclinical, during the transition period. Such a multi-trait model could take into consideration all available health data including clinical disease records, existing metabolic phenotypes (serum, milk and MIR-predicted phenotypes) and somatic cell count (SCC). This could help identify what Dr Bill Wales of Agriculture Victoria Research refers to as "invisible cows" – those that successfully make it through lactation year after year with no problems.

6.4.3 Metabolomic studies

Metabolomics was recently described as the being the "cornerstone of the next generation phenotyping approaches that are needed to refine and improve trait descriptions" in livestock (Fontanesi, 2016). Exciting opportunities exist to use metabolomic approaches to identify objective, measurable metabotypes associated with more complex trait definitions such as the aforementioned "successful transition" or "invisibility".

While extremely promising, novel biomarkers require extensive validation and epidemiological studies to confirm associations with adverse animal health and production outcomes. For this reason, there is also great value in continuing to better characterise the biochemical and genetic parameters of existing biomarkers, and thereby build on the already extensive body of epidemiological and physiological research in the literature. As more samples are collected, there will be great value in undertaking detailed metabolomic phenotyping on a subset of animals identified as being "extreme", such as those with extreme values of traditional metabolic profile biomarker concentrations, and those with clinical diseases. Increasing the size and diversity (both phenotypic and genetic) of metabolomic reference populations should enable identification of more metabotypes that are truly representative of genetic differences in metabolic resilience. The approach presented in Chapter 5 should make increasing the size of the reference population easier, by allowing integration of metabolomic data from multiple sources (within and between countries) without confounding from fixed environmental effects.

The large and significant effects of herd-specific factors on the serum metabolome also warrant further investigation. In particular, serum lactate concentrations varied significantly between farms. In chapter 5, we hypothesized that these differences may be due to dietary and/or management differences leading to ruminal acidosis, however, these differences could also reflect differences in pre-analytical sample handling. After collection, serum remains biologically active, and continued anaerobic metabolism by blood cells can lead to increases in

lactate concentration (Teahan et al. 2006). Watkins et al. (2011) reported that lactate concentrations in human whole blood samples were significantly ($P < 0.05$) higher in samples that had been stored at room temperature for 30 minutes compared to those in samples that were analysed immediately after collection. Similarly, Teahan et al. (2006) demonstrated that variations in clotting time and prolonged serum-clot contact time lead to changes in energy metabolite concentrations, in particular those of lactate. Interestingly, unpublished statistical analysis of the data used in Chapter 5 reveals that lactate concentration has a very large and significant effect on NMR spectra even when herd-specific factors (which include differences in pre-analytical sample handling) are accounted for (Table 1). Given the unique logistical challenges of large-scale, on-farm sample collection, further work is required for us to develop better understanding of the impacts of differences in pre-analytical sample handling on the bovine serum metabolome.

Table 1. Results of multiple linear regression models of principal component (PC) scores derived from PCA of ^{1}H NMR spectra, against days in milk (DIM), parity (1 to 4 +), herd/sample collection date (SCD) (each visit to a herd was treated as an independent event to account for differences in feed and sample handling), and lactate concentration. Conditional Wald F statistics (F-con) and corresponding P values describe the magnitude and statistical significance of each fixed effect, respectively, assuming that the effects of remaining predictor variables have been accounted for.

	PC1 (47.64%)		PC2 (15.59%)		PC3 (7.45%)	
Fixed Effect	**F-con[1]**	**P Value**	**F-con**	**P Value**	**F-con**	**P Value**
DIM	2.23	<0.001	3.05	<0.001	1.55	0.035
Parity	-	-	-	-	-	-
Lactate	95.23	<0.001	194.95	<0.001	18.38	<0.001
Herd/SCD	25.6	<0.001	10.71	<0.001	3.36	<0.001

Further work is also warranted to investigate the use of ^{1}H NMR spectroscopy to quantify lipoproteins and glycoproteins. There is also likely to be great value in using more sensitive metabolomic techniques for metabotypes discovery, such as liquid chromatography mass-spectroscopy.

6.4.4 Functional genomic studies

Finally, perhaps the most exciting future work to follow on from this thesis will be the search for QTL associated with metabolic health. Recent updates to the bovine reference genome (Hayes and Daetwyler, 2019) and improvements in imputation methods have increased the power of GWAS to identify causal variants. It is hoped that the phenotypes investigated in this thesis will form the basis of exciting future work in this area.

6.5 Concluding statement

I hope that the findings of this thesis will contribute to improving the health and welfare of dairy cattle by offering a preliminary "metabolic phenotype framework" for use in genetic selection for improved metabolic health during early lactation.

References

Abdelsayed, M., M. Haile-Mariam, and J. E. Pryce. 2017. Genetic parameters for health traits using data collected from genomic information nucleus herds. J. Dairy Sci. 100:9643-9655. 10.3168/jds.2017-12960

Atashi, H., M. Salavati, J. De Koster, M. A. Crowe, G. Opsomer, and M. Hostens. 2020. Genome-wide association for metabolic clusters in early-lactation Holstein dairy cows. J. Dairy Sci. 103. 10.3168/jds.2019-17369

Bach, K. D., D. M. Barbano, and J. A. A. McArt. 2019. Association of mid-infrared-predicted milk and blood constituents with early-lactation disease, removal, and production outcomes in Holstein cows. J. Dairy Sci. 102:10129-10139. 10.3168/jds.2019-16926Belay, T. K., M. Svendsen, Z. M. Kowalski, and T. Ådnøy. 2017. 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. 100:6298-6311. 10.3168/jds.2016-12458

Benedet, A., A. Costa, M. De Marchi, and M. Penasa. 2020. Heritability estimates of predicted blood β-hydroxybutyrate and nonesterified fatty acids and relationships with milk traits in early-lactation Holstein cows. J. Dairy Sci. 103. 10.3168/jds.2019-17916

Calus, M. P. L., Y. de Haas, M. Pszczola, and R. F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. Animal. 7:183-191. 10.1017/S1751731112001450

Chesnais, J. P., T. A. Cooper, G. R. Wiggans, M. Sargolzaei, J. E. Pryce, and F. Miglior. 2016. Using genomics to enhance selection of novel traits in North American dairy cattle. J. Dairy Sci. 99:2413-2427. 10.3168/jds.2015-9970

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395. 10.1371/journal.pone.0003395

Daetwyler, H. D., R. Xiang, Z. Yuan, S. Bolormaa, C. J. Vander Jagt, B. J. Hayes, J. H. J. van der Werf, J. E. Pryce, A. J. Chamberlain, I. M. Macleod, and M. E. Goddard. 2019. Integration of functional genomics and phenomics into genomic prediction raises its accuracy in sheep and dairy cattle. Pages 11-14 Proc. Assoc. Advmt. Anim. Breed. Genet., Armidale, NSW, Australia.

DataGene. 2020. DataVat. Accessed June 26, 2020. http://uat.datavat.com.au/

de Haas, Y., J. E. Pryce, M. P. Calus, E. Wall, D. P. Berry, P. Lovendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsegge, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. J. Dairy Sci. 98:6522-6534. 10.3168/jds.2014-9257

Drackley, J. K. 1999. Biology of dairy cows during the transition period: the final frontier? J. Dairy Sci. 82:2259-2273. 10.3168/jds.S0022-0302(99)75474-3

Fontanesi, L. 2016. Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. Anim. Front. 6:73-79. 10.2527/af.2016-0011

Gonzalez-Recio, O., M. P. Coffey, and J. E. Pryce. 2014. On the value of the phenotypes in the genomic era. J. Dairy Sci. 97:7905-7915. 10.3168/jds.2014-8125

Hayes, B. J. and H. D. Daetwyler. 2019. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. Annu. Rev. Anim. Biosc. 7:89-102. 10.1146/annurev-animal-020518-115024

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet. 6:e1001139. 10.1371/journal.pgen.1001139

Houle, D., D. Govindaraju, and S. Omholt. 2010. Phenomics: The next challenge. Nat. Rev. Gen. 11:855-866. 10.1038/nrg2897

Karsten, S. and G. Christian. 2012. Genetic variation in metabolic phenotypes: study designs and applications. Nat. Rev. Gen.13:759. 10.1038/nrg3314

Koeck, A., J. Jamrozik, F. S. Schenkel, R. K. Moore, D. M. Lefebvre, D. F. Kelton, and F. Miglior. 2014. Genetic analysis of milk β-hydroxybutyrate and its association with fat-to-protein ratio, body condition score, clinical ketosis, and displaced abomasum in early first lactation of Canadian Holsteins. J. Dairy Sci. 97:7286-7292. 10.3168/jds.2014-8405

Lean, I. J. and P. J. Degaris. 2010. Transition Cow Management. A review for nutritional professionals, veterinarians and farm advisers. Dairy Australia, Melbourne, Australia.

LeBlanc, S. J., K. D. Lissemore, D. F. Kelton, T. F. Duffield, and K. E. Leslie. 2006. Major advances in disease prevention in dairy cattle. J. Dairy Sci. 89:1267-1279. 10.3168/jds.S0022-0302(06)72195-6

MacLeod, I. M., P. J. Bowman, C. J. Vander Jagt, M. Haile-Mariam, K. E. Kemper, A. J. Chamberlain, C. Schrooten, B. J. Hayes, and M. E. Goddard. 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genom. 17. 10.1186/s12864-016-2443-6

McArt, J. A. A. and R. C. Neves. 2020. Association of transient, persistent, or delayed subclinical hypocalcemia with early lactation disease, removal, and milk yield in Holstein cows. J. Dairy Sci. 103:690-701. 10.3168/jds.2019-17191

McArt, J. A. A., D. V. Nydam, G. R. Oetzel, T. R. Overton, and P. A. Ospina. 2013. Elevated non-esterified fatty acids and β-hydroxybutyrate and their association with transition dairy cow performance. Vet. J. 198:560-570. 10.1016/j.tvjl.2013.08.011

Meuwissen Theo, H. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35. 10.1186/1297-9686-41-35

Mrode, R. A. and G. J. T. Swanson. 1996. Genetic and statistical properties of somatic cell count and its suitability as an indirect means of reducing the incidence of mastitis in dairy cattle. Anim. Breed. Abstr. 64:847-857.

Nayeri, S., F. Schenkel, A. Fleming, V. Kroezen, M. Sargolzaei, C. Baes, A. Cánovas, J. Squires, and F. Miglior. 2019. Genome-wide association analysis for β-hydroxybutyrate concentration in Milk in Holstein dairy cattle. BMC Genet. 20. 10.1186/s12863-019-0761-9

Neves, R. C., B. M. Leno, K. D. Bach, and J. A. A. McArt. 2018. Epidemiology of subclinical hypocalcemia in early-lactation Holstein dairy cows: The temporal associations of plasma calcium concentration in the first 4 days in milk with disease and milk production. J. Dairy Sci. 101:9321-9331. 10.3168/jds.2018-14587

Oikonomou, G., G. E. Valergakis, G. Arsenos, N. Roubies, and G. Banos. 2008. Genetic profile of body energy and blood metabolic traits across lactation in primiparous holstein cows. J. Dairy Sci. 91:2814-2822. 10.3168/jds.2007-0965

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010a. Associations of elevated nonesterified fatty acids and β-hydroxybutyrate concentrations with early lactation

reproductive performance and milk production in transition dairy cattle in the northeastern United States. J. Dairy Sci. 93:1596-1603. 10.3168/jds.2009-2852

Ospina, P. A., D. V. Nydam, T. Stokol, and T. R. Overton. 2010b. Evaluation of nonesterified fatty acids and β-hydroxybutyrate in transition dairy cattle in the northeastern United States: Critical thresholds for prediction of clinical diseases. J. Dairy Sci. 93:546-554. 10.3168/jds.2009-2277

Overton, T. R. and M. R. Waldron. 2004. Nutritional management of transition dairy cows: strategies to optimize metabolic health. J. Dairy Sci. 87:E105-E119. 10.3168/jds.S0022-0302(04)70066-1

Pryce, J. E., O. Gonzalez-Recio, G. Nieuwhof, W. J. Wales, M. P. Coffey, B. J. Hayes, and M. E. Goddard. 2015. Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. J. Dairy Sci. 98:7340-7350. 10.3168/jds.2015-9621

Pryce, J. E., K. L. Parker Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B. Heringstad, C. Egger-Danner, K. F. Stock, A. J. Bradley, and J. B. Cole. 2016. Invited review: Opportunities for genetic improvement of metabolic diseases. J. Dairy Sci. 99:6855-6873. 10.3168/jds.2016-10854

Rius-Vilarrasa, E., W. F. Fikse, E. Carlén, J.-Å. Eriksson, J. Pöso, U. S. Nielsen, and G. P. Aamand. 2018. Improved genetic evaluation of health traits using metabolic biomarkers in Nordic dairy cattle. No 53 in Proc. Interbull Meet., Auckland, NZ. Interbull, Uppsala, Sweden.

Teahan O, Gamble S, Holmes E, Waxman J, Nicholson JK, Bevan C, Keun HC. 2006. Impact of analytical bias in metabonomic studies of human blood serum and plasma. Anal Chem 78:4307– 4318.

Tsiamadis, V., G. Banos, N. Panousis, M. Kritsepi-Konstantinou, G. Arsenos, and G. E. Valergakis. 2016a. Genetic parameters of calcium, phosphorus, magnesium, and potassium serum concentrations during the first 8 days after calving in Holstein cows. J. Dairy Sci. 99:5535-5544. 10.3168/jds.2015-10787

Tsiamadis, V., G. Banos, N. Panousis, M. Kritsepi-Konstantinou, G. Arsenos, and G. E. Valergakis. 2016b. Genetic parameters of subclinical macromineral disorders and major clinical diseases in postparturient Holstein cows. J. Dairy Sci. 99:8901-8914. 10.3168/jds.2015-10789

Xiang, R., I. v. D. Berg, I. M. Macleod, B. J. Hayes, C. P. Prowse-Wilkins, M. Wang, S. Bolormaa, Z. Liu, S. J. Rochfort, C. M. Reich, B. A. Mason, C. J. Vander Jagt, H. D. Daetwyler, M. S. Lund, A. J. Chamberlain, and M. E. Goddard. 2019. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. Proc. Natl. Acad. Sci. U. S. A. 116:19398-19408. 10.1073/pnas.1904159116

Yepes, F. A. L., D. V. Nydam, S. Mann, L. Caixeta, J. A. A. McArt, T. R. Overton, J. J. Wakshlag, and H. J. Huson. 2019. Longitudinal phenotypes improve genotype association for hyperketonemia in dairy cattle. Animals. 9:1059. 10.3390/ani9121059

Appendix 1: Co-Author Contributions to Published Research Articles

Timothy David William Luke made significant leading contributions to all publications listed hereafter.

Chapter 2: Genomic Prediction of Serum Biomarkers of Health In Early Lactation

T. D. W. Luke (TL), T. T. T. Nguyen (TN), S. Rochfort (SR), W. J. Wales (WW). C. M. Richardson (CR), M. Abdelsayed (MA) and J. E. Pryce (JP)

Conceptualization: TL, JP, WW, & SR

Sample acquisition: TL

Animal ethics approval: TL & WW

Methodology: TL, TN, CR, MA & JP

Data analysis: TL, TN & JP

Funding & resource acquisition: JP

Project administration: SR, WW & JP

Supervision: WW, SR & JP

Writing & visualisation: TL

Writing (review & editing): All authors

Chapter 3: Metabolic Profiling of Early Lactation Dairy Cows Using Milk Mid-Infrared Spectra

T. D. W. Luke (TL), S. Rochfort (SR), W. J. Wales (WW). V. Bonfatti (VB), L. Marett (LM) and J. E. Pryce (JP)

Conceptualization: TL, JP, WW, & SR

Sample acquisition: TL & LM

Animal ethics approval: TL, LM & WW

Methodology: TL, VB, SR & JP

Data analysis: TL, SR & JP

Funding & resource acquisition: JP & WW

Project administration: WW, SR & JP

Supervision: WW, SR & JP

Writing & visualisation: TL

Writing (review & editing): All authors

Chapter 4: A Tale of Two Biomarkers: Untargeted ^{1}H NMR Metabolomic Fingerprinting of BHBA and NEFA in Early Lactation Dairy Cows

T. D. W. Luke (TL), J. E. Pryce (JP), W. J. Wales (WW). and S. Rochfort (SR)

Conceptualization: TL, JP, WW, & SR

Sample Acquisition: TL

Animal Ethics Approval: TL & WW

Methodology: TL & SR

Laboratory analysis: TL

Data analysis: TL & SR

Funding & resource acquisition: JP & SR

Project administration: JP, WW & SR

Supervision: JP, WW & SR

Writing & visualisation: TL

Writing (review & editing): All authors

Chapter 5: Use of Large and Diverse Datasets for ^{1}H NMR Serum Metabolic Profiling of Early Lactation Dairy Cows

T. D. W. Luke (TL), J. E. Pryce (JP), A. C. Elkins (AE), W. J. Wales (WW). and S. Rochfort (SR)

Conceptualization: TL, JP, WW, & SR

Sample Acquisition: TL

Animal Ethics Approval: TL & WW

Methodology: TL, AE & SR

Laboratory analysis: TL & AE

Data analysis: TL & SR

Funding & resource acquisition: JP & SR

Project administration: JP, WW & SR

Supervision: JP, WW & SR

Writing & visualisation: TL

Writing (review & editing): All authors

Appendix 2: Chapter 4 Supplementary Material

Table S1. ^{1}H NMR chemical shifts (δ) and multiplicity of metabolites in bovine serum run in deuterated water (D_2O). Clearly observed resonances are indicated in bold text**. s**, singlet; d, doublet; dd, doublet of a doublet; m, multiplet; t, triplet. The right two columns show the direction of the relationship with serum β-hydroxybutyrate (BHBA) and non-esterified fatty acid (NEFA) concentrations determined by colorimetric assays. * = tentative identification.

	Metabolite	Chemical shift (δ) and multiplicity	BHBA	NEFA
1	cholate*	**0.70 (m),** 0.91 (m), 0.96 (m), 1.43 (m), 1.87 (m), 2.10 (m), 2.22 (m), 3.65 (m), 4.06 (t)	-	↓
2	LDL/VLDL	**0.86 (m), 1.25 (m)**	↑	↑
3	leucine	**0.94 (d)**, **0.95 (d)**, 1.66 (m), 1.66 (m), 1.73 (m), 3.72 (m)	-	↑
4	isoleucine	**0.93 (t)**, **1.00 (d)**, 1.24 (m), 1.45 (m), 1.45 (m), 3.66 (d)	-	↓
5	valine	**0.98 (d)**, **1.03 (d)**, 2.26 (m), 3.60 (d)	↓	↓
6	β-hydroxybutyrate	**1.20 (d)**, **2.31 (m), 2.41 (m), 4.16 (m)**	↑	↑
7	lactate	**1.31 (d),** 4.31 (q)	↓	↓
8	alanine	**1.46 (d),** 3.77 (q)	↓	↓
9	acetate	**1.9 (s)**	↑	↓
10	N-acetyl glycoprotein	**2.03 (m)**	↑	↑
11	pyruvate	**2.46 (s),** 7.65 (s)	-	↓
12	citrate	**2.52 (d), 2.66 (d)**	-	-
13	creatine	**3.02 (s), 3.92 (s)**	↑	↓
14	phosphocreatine	**3.03 (s)**, **3.93 (s)**	-	-
15	dimethyl sulfone	**3.14 (s)**	↓	-
16	choline	**3.19 (s)**, 3.50 (m), 4.05 (m)	-	-
17	phosphocholine	**3.21 (s)**, 3.58 (t), 4.17 (m)	↑	↑
18	betaine	**3.25 (s)**, 3.89 (s)	↑	-
19	methanol	**3.34 (s)**		
20	glucose	**3.23 (dd)**, **3.40 (m), 3.46 (m), 3.52 (dd), 3.73 (m), 3.82 (m), 3.89 (dd), 4.63 (d)**, **5.22 (d)**	↓	↓
21	glycine	**3.50 (s)**	↑	↑
22	β-glucose	**4.63 (d)**	↓	↓
23	α-glucose	**5.22 (d)**	↓	↓
24	3-phenyllactate*	**2.87 (dd)**, 3.09 (dd), 4.26 (dd), **7.31 (m), 7.39 (m)**	-	-
25	hippurate	3.96 (d), **7.54 (m), 7.62 (m), 7.83 (dd)**	-	-
26	formate	**8.44 (s)**	-	-

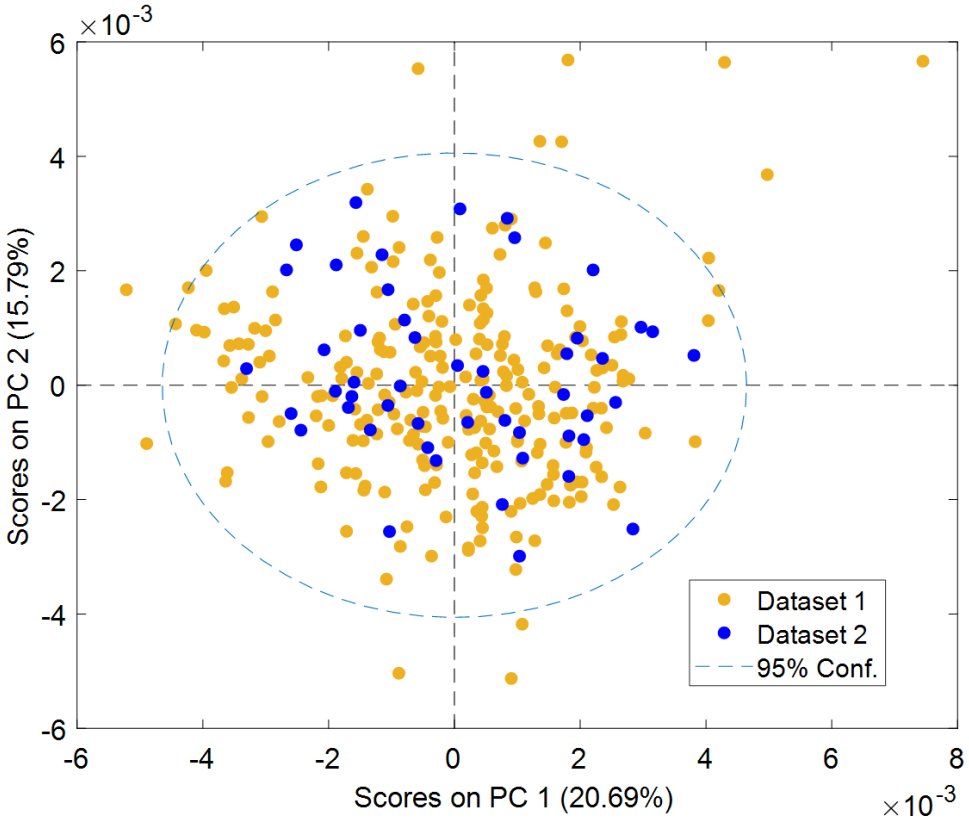
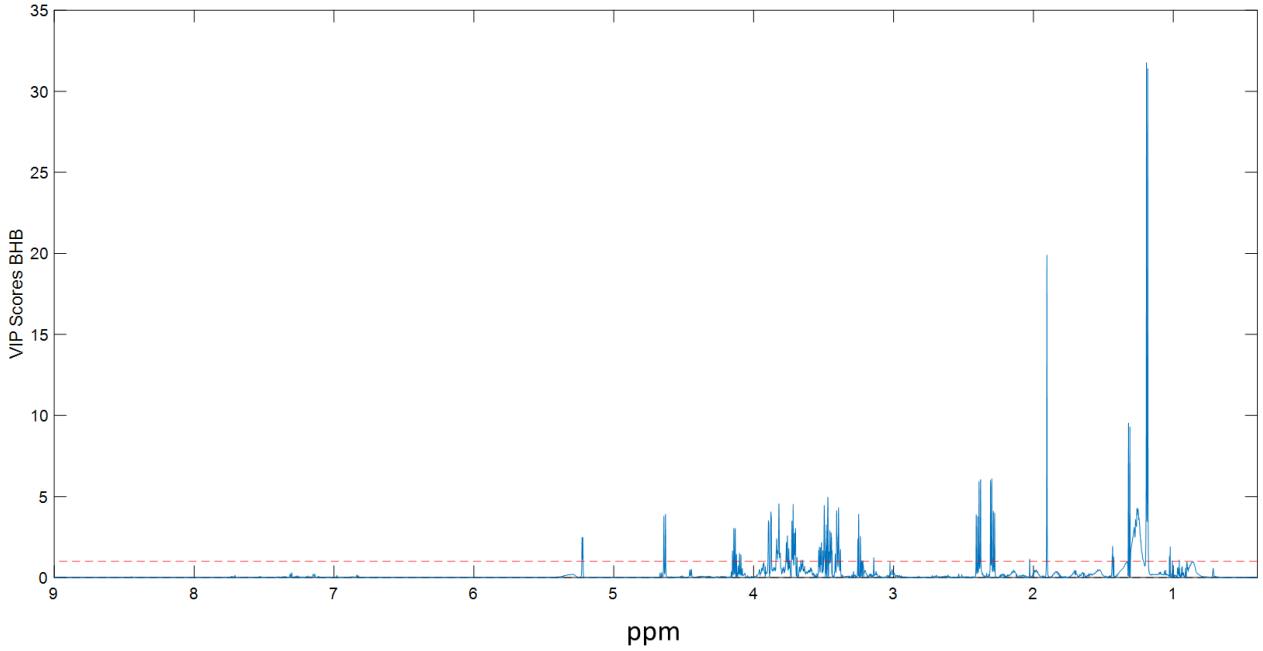


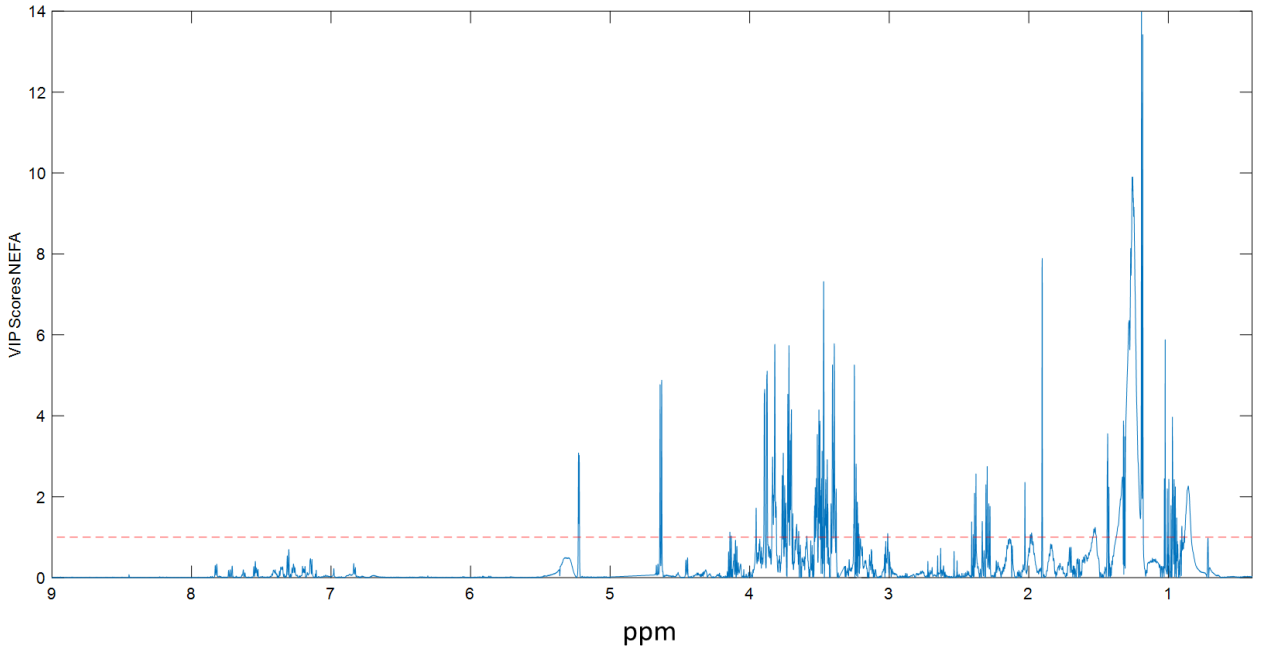
Figure S1. Results of PCA of ^{1}H NMR spectra of serum obtained from 298 dairy cows in early lactation from the Ellinbank research farm (Dataset 1, N = 248) and a commercial dairy farm in Tasmania (Dataset 2, N = 50).

Table S2. Results of ANOVA-simultaneous component analysis (ASCA) of ^{1}H NMR spectra of bovine serum (N = 298). Effect describes the relative influence of each variable (Herd, Age and days in milk (DIM)) on each spectra. P-value is derived from permutation testing (1000 iterations).

Variable	PCs	Effect	P-Value
Herd	1	0.47	0.33
Age	10	4.72	0.04
DIM	20	10.17	0.09
Residuals	-	86.06	-



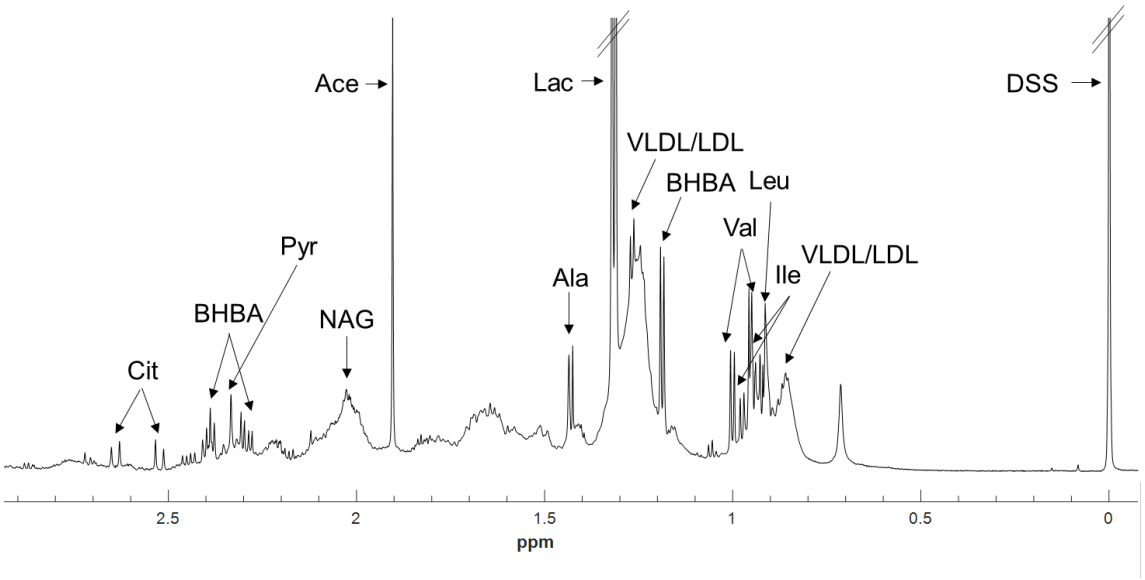
(a)

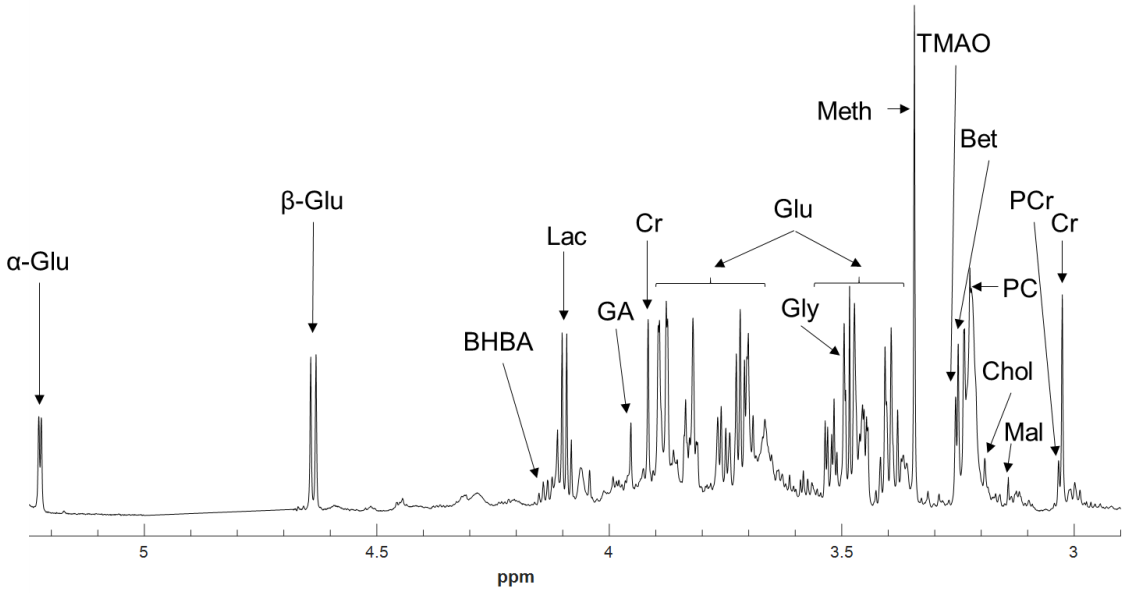


(b)

Figure S2. Variable importance of projection (VIP) scores derived from orthogonal partial least squares (OPLS) regression of ^{1}H NMR spectra of serum obtained from 298 dairy cows in early lactation, against (a) BHBA concentration and (b) NEFA concentration.

Appendix 3: Chapter 5 Supplementary Material





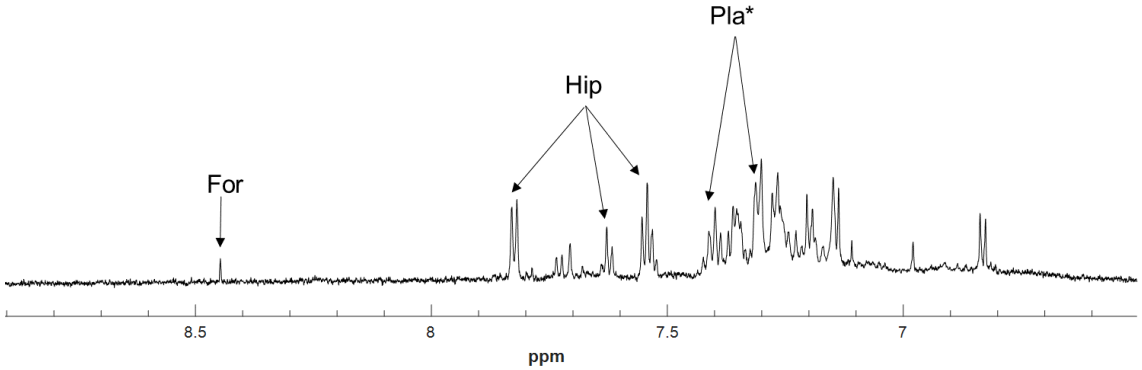


Figure S1. Representative 700MHz ^{1}H NMR spectrum (δ -0.1 to 9.0) of serum obtained from a Holstein-Friesian cow in early lactation. The δ 0.4 to 2.9, δ 2.9 to 5.25 and δ 6.5 to 9.0 regions have been magnified for clarity purposes. α-Glu = α glucose, β-Glu = β glucose, Ace = acetate, Ala = alanine, Bet = betaine, BHBA = β hydroxybutyrate, Cit = citrate, Chol = choline, Cr = creatine, For = formate, GA = glycolate, Glu = glucose, Gly = glycine, Hip = hippurate, Ile = isoleucine, Lac = lactate, Leu = leucine, Mal = malonate, Meth = methanol, NAG = N-acetyl glycoprotein, PC = phosphocholine, PCr = phosphocreatine, PLa* = 3-phenyllactate, Pyr = pyruvate, TMAO = Trimethylamine N-oxide, Val = valine, VLDL/LDL = Very low density lipoprotein and low density lipoprotein. * = tentative identification.

Table S1. ^{1}H NMR chemical shifts (δ) and multiplicity of metabolites in bovine serum run in deuterated water (D_2O). Clearly observed resonances are indicated in bold text**. s**, singlet; d, doublet; dd, doublet of a doublet; m, multiplet; t, triplet.

Label	Metabolite	Chemical shift (δ) and multiplicity
Leu	leucine	**0.94 (d)**, **0.95 (d)**, 1.66 (m), 1.66 (m), 1.73 (m), 3.72 (m)
Ile	isoleucine	**0.93 (t)**, **1.00 (d)**, 1.24 (m), 1.45 (m), 1.45 (m), 3.66 (d)
Val	valine	**0.98 (d)**, **1.03 (d)**, 2.26 (m), 3.60 (d)
BHBA	β-hydroxybutyrate	**1.20 (d)**, **2.31 (m), 2.41 (m), 4.16 (m)**
Lac	lactate	**1.31 (d),** 4.31 (q)
Ala	alanine	**1.46 (d),** 3.77 (q)
Ace	acetate	**1.9 (s)**
Pyr	pyruvate	**2.46 (s),** 7.65 (s)
Cit	citrate	**2.52 (d), 2.66 (d)**
Cr	creatine	**3.02 (s), 3.92 (s)**
PCr	phosphocreatine	**3.03 (s)**, **3.93 (s)**
Mal	malonate	**3.11 (s)**
Chol	choline	**3.19 (s)**, 3.50 (m), 4.05 (m)
PC	phosphocholine	**3.21 (s)**, 3.58 (t), 4.17 (m)
Bet	betaine	**3.25 (s)**, 3.89 (s)
TMAO	trimethylamine N-oxide	**3.25 (s)**
Meth	methanol	**3.34 (s)**
Glu	glucose	**3.23 (dd)**, **3.40 (m), 3.46 (m), 3.52 (dd), 3.73 (m), 3.82 (m), 3.89 (dd), 4.63 (d)**, **5.22 (d)**
Gly	glycine	**3.50 (s)**
GA	glycolate	**3.93 (s)**
β-Glu	β-glucose	**4.63 (d)**
α-Glu	α-glucose	**5.22 (d)**
PLa	3-phenyllactate*	**2.87 (dd)**, 3.09 (dd), 4.26 (dd), **7.31 (m), 7.39 (m)**
Hip	hippurate	3.96 (d), **7.54 (m), 7.62 (m), 7.83 (dd)**

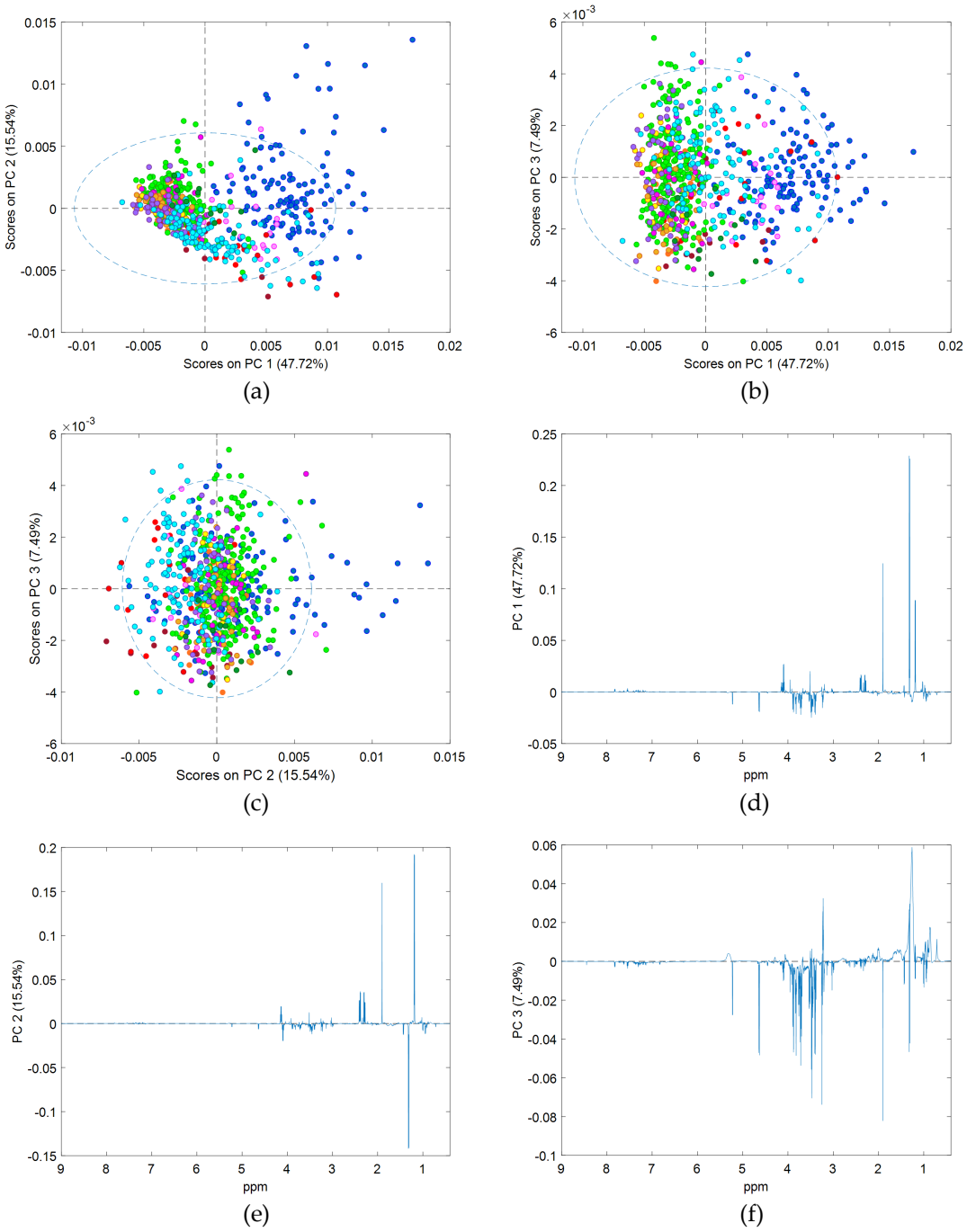


Figure S2. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for weeks in milk using linear regression; (a) PC 1 vs PC 2 scores, (b) PC 1 vs PC 3 scores (c) PC 2 vs PC 3 scores (d) PC 1 loadings (e) PC 2 loadings and (f) PC 3 loadings plots. Scores plots are coloured by farm of origin.

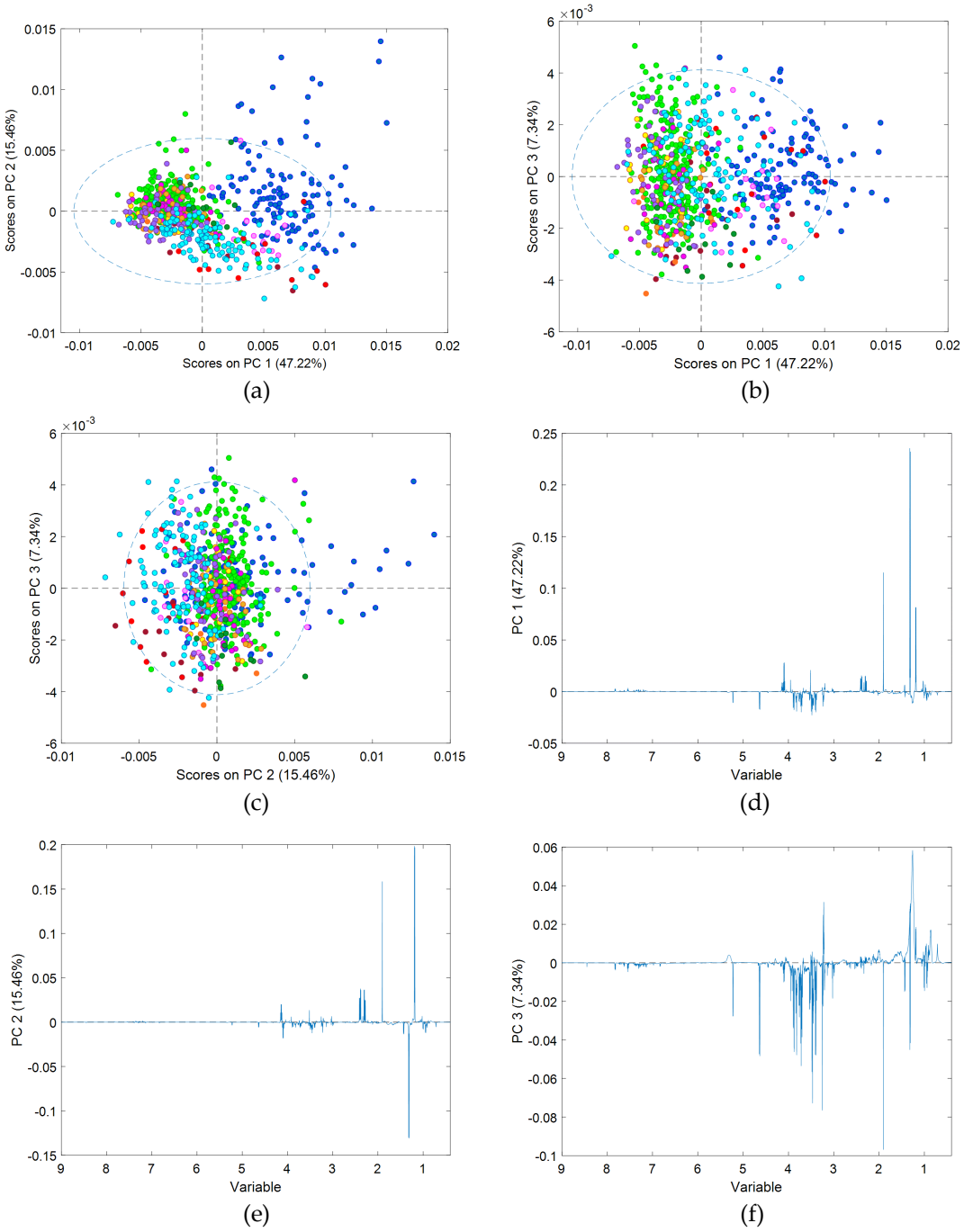


Figure S3. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for Parity using linear regression; (a) PC 1 vs PC 2 scores, (b) PC 1 vs PC 3 scores (c) PC 2 vs PC 3 scores (d) PC 1 loadings (e) PC 2 loadings and (f) PC 3 loadings plots. Scores plots are coloured by farm of origin.

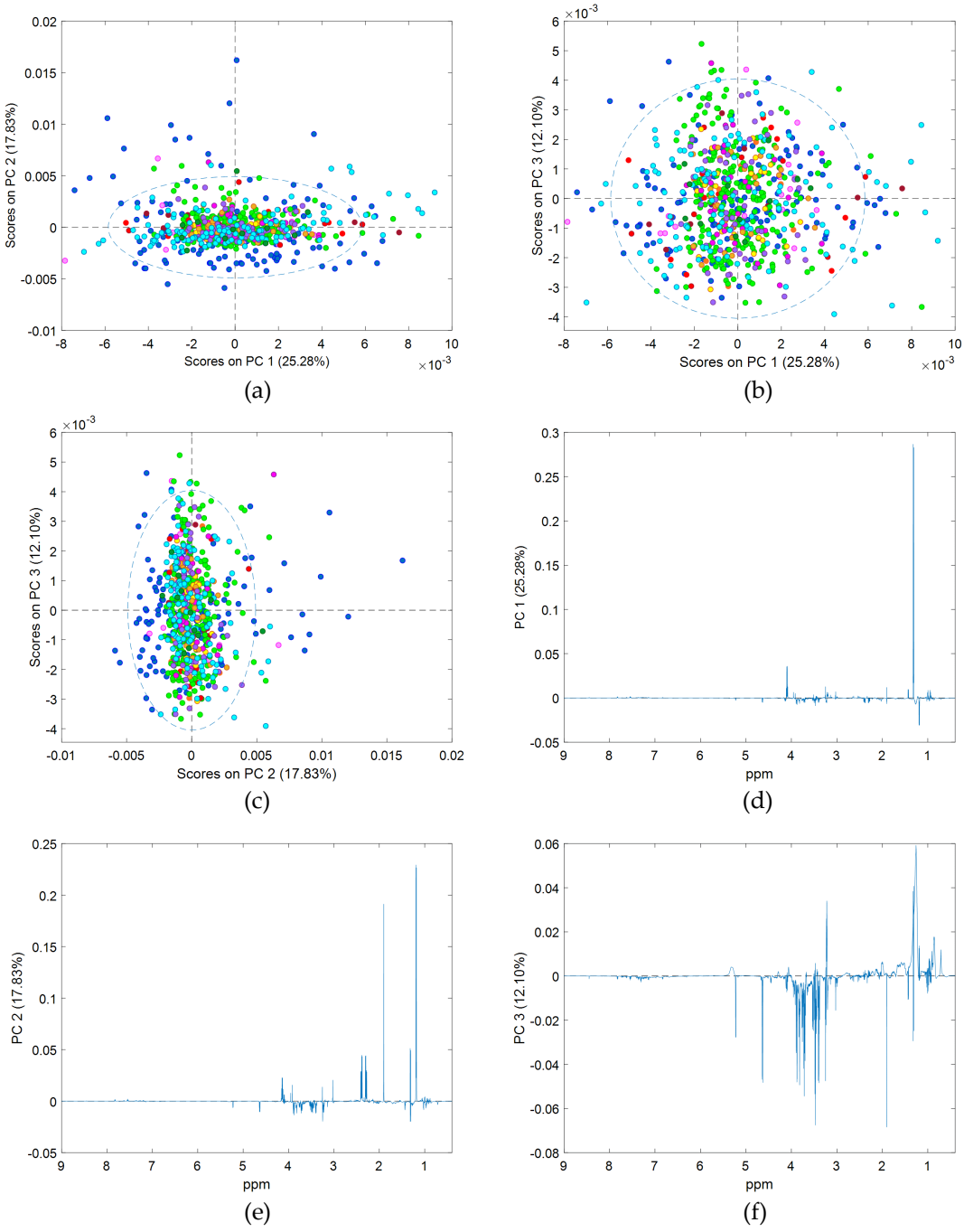
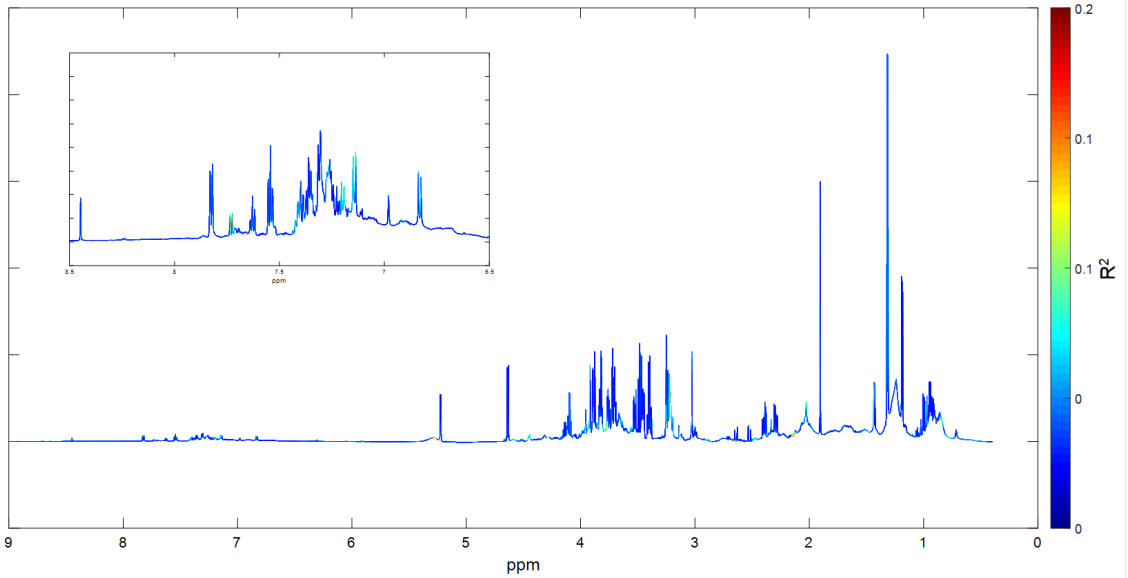
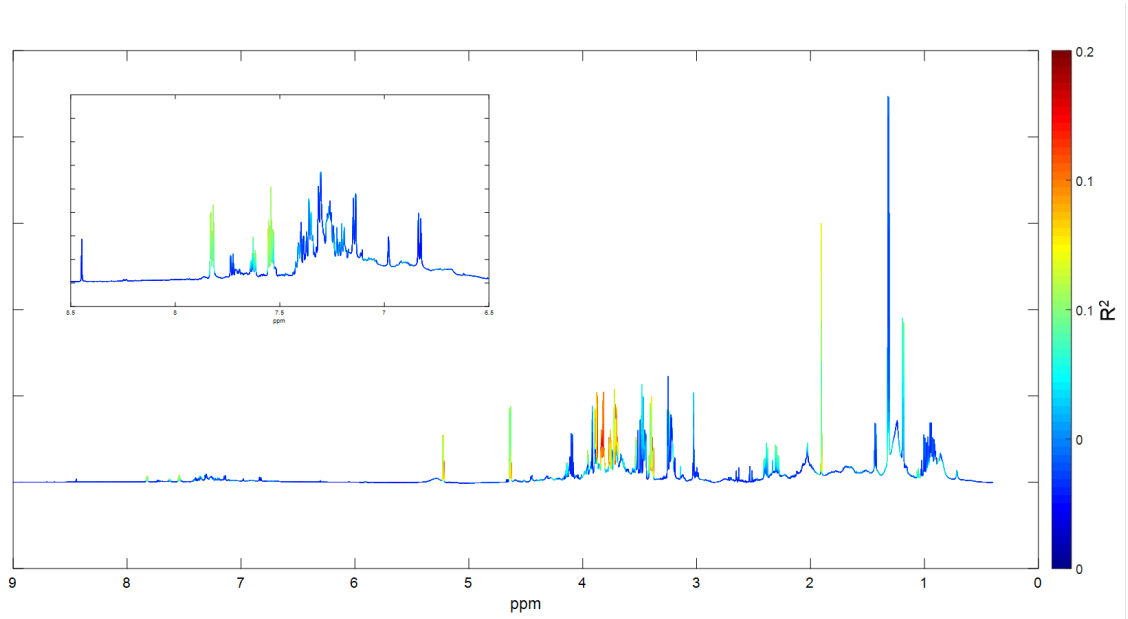


Figure S4. Results of PCA of 707 ^{1}H NMR spectra of serum obtained from dairy cows in early lactation, corrected for Herd using linear regression; (a) PC 1 vs PC 2 scores, (b) PC 1 vs PC 3 scores (c) PC 2 vs PC 3 scores (d) PC 1 loadings (e) PC 2 loadings and (f) PC 3 loadings plots. Scores plots are coloured by farm of origin.



(a)



(b)

Figure S5. Average ^{1}H NMR spectrum of bovine serum. Colour-coding represents the percentage of variation in the signal at each chemical shift intenisty that can be explained by (a) WIM and (b) Parity. The δ 6.5 to 8.5 region has been magnified for clarity purposes.

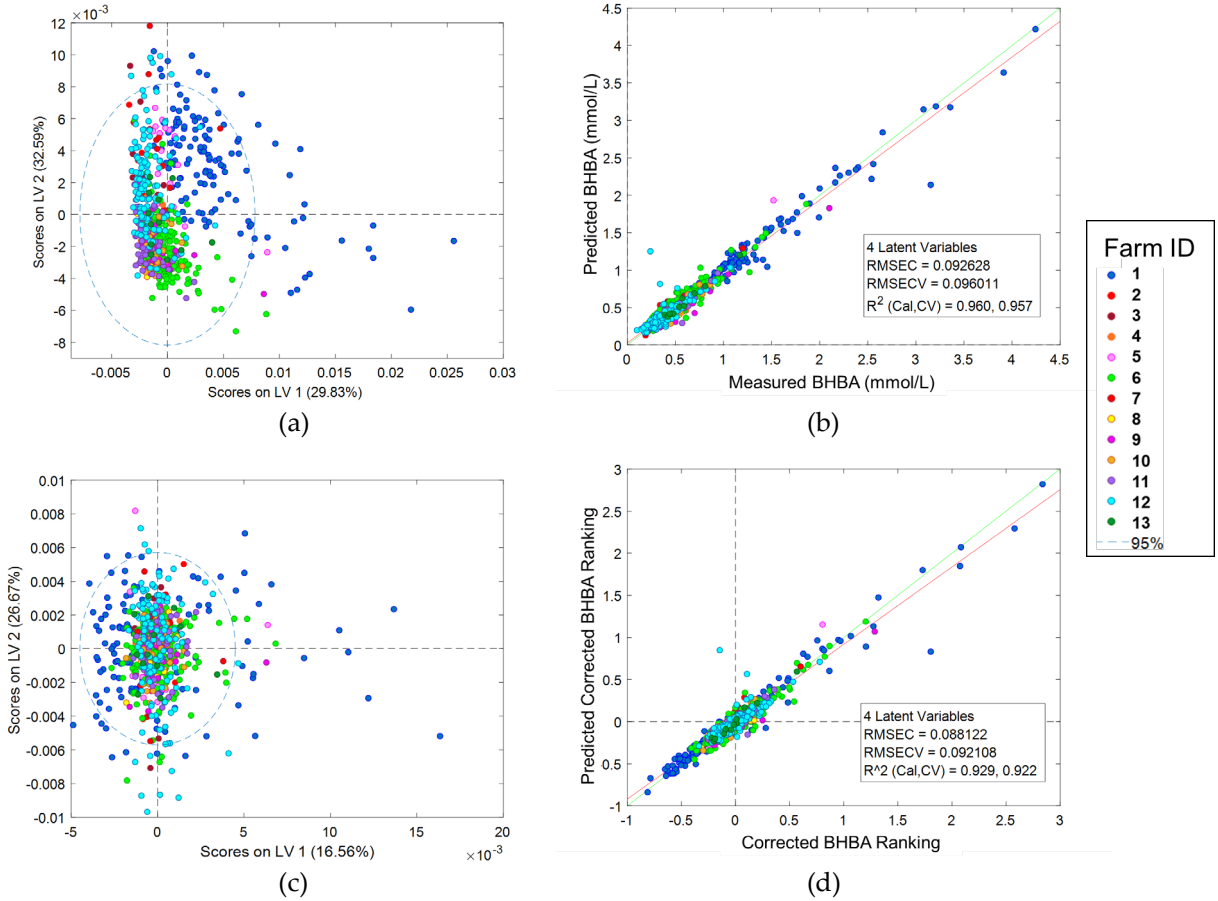


Figure S6. Results of OPLS regressions of serum BHBA concentration against ^{1}H NMR spectrum of bovine serum (n = 707): (a) LV1 vs LV2 scores for uncorrected data (b) CV predicted vs measured BHBA (c) LV1 vs LV2 scores for data corrected for all fixed effects (d) CV predicted vs measured corrected BHBA ranking. Plots are colored by Herd.

Table S2. Results of ANOVA-simultaneous coponent analysis (ASCA) of uncorrected ^{1}H NMR spectra of bovine serum (n = 707), and spectra that have been corrected using linear regression for (1) all fixed effects, (2) weeks in milk (WIM), (3) parity, and (4) herd of origin. Effect describes the relative influence of each variable (weeks in milk (WIM), Parity and Herd) on each spectral dataset. P-value is derived from permutation testing (50 iterations).

	Uncorrected		All Fixed Effects		WIM		Parity		Herd	
Variable	Effect	P-Value	Effect	P-Value	Effect	P-Value	Effect	P-Value	Effect	P-Value
WIM	1.37	0.02	0.00	1.00	0.00	0.02	1.55	0.02	1.68	0.02
Parity	4.10	0.02	0.00	1.00	4.24	0.02	0.00	0.02	3.30	0.02
Herd	43.99	0.02	0.00	1.00	43.60	0.02	41.72	0.02	0.00	1.00

Appendix 4: List of Published Research Articles Contributed to During Candidature

Bonfatti, V., S. A. Turner, B. Kuhn-Sherlock, **T. D. W. Luke**, P. N. Ho, C. V. C. Phyn, and J. E. Pryce. 2019. Prediction of blood beta-hydroxybutyrate content and occurrence of hyperketonemia in early-lactation, pasture-grazed dairy cows using milk infrared spectra. J. Dairy Sci. 102:6466-6476. 10.3168/jds.2018-15988

Ho, P. N., V. Bonfatti, **T. D. W. Luke**, and J. E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460-10470. 10.3168/jds.2019-16412

Luke, T.D.W., V. Russo, S. Rochfort, B. Wales, and J. Pryce. 2018. Mid-infrared spectroscopy of milk as a tool to predict subacute ruminal acidosis. J. Anim. Sci. 96:502-502. 10.1093/jas/sky404.1096

Pryce, J. E., T. T. T. Nguyen, P. N. Ho, **T. D. W. Luke**, S. Rochfort, W. J. Wales, P. J. Moate, L. C. Marett, G. Nieuwhof, M. Abdelsayed, M. Axford, M. Shaffer, and M. Haile-Mariam. 2019. Applying next generation phenotyping strategies for genetic gain in dairy cattle. Proc. Assoc. Advmt. Anim. Breed. Genet. 23:390-393.