


SHORT COMMUNICATION

Open Access



# Using mid-infrared spectroscopy to increase GWAS power to detect QTL associated with blood urea nitrogen

Irene van den Berg<sup>1\*</sup> , Phuong N. Ho<sup>1</sup>, Tuan V. Nguyen<sup>1</sup>, Mekonnen Haile-Mariam<sup>1</sup>, Timothy D. W. Luke<sup>1</sup> and Jennie E. Pryce<sup>1,2</sup>

## Abstract

Blood urea nitrogen (BUN) is an indicator trait for urinary nitrogen excretion. Measuring BUN level requires a blood sample, which limits the number of records that can be obtained. Alternatively, BUN can be predicted using mid-infrared (MIR) spectroscopy of a milk sample and thus records become available on many more cows through routine milk recording processes. The genetic correlation between MIR predicted BUN (MBUN) and BUN is 0.90. Hence, genetically, BUN and MBUN can be considered as the same trait. The objective of our study was to perform genome-wide association studies (GWAS) for BUN and MBUN, compare these two GWAS and detect quantitative trait loci (QTL) for both traits, and compare the detected QTL with previously reported QTL for milk urea nitrogen (MUN). The dataset used for our analyses included 2098 and 18,120 phenotypes for BUN and MBUN, respectively, and imputed whole-genome sequence data. The GWAS for MBUN was carried out using either the full dataset, the 2098 cows with records for BUN, or 2000 randomly selected cows, so that the dataset size is comparable to that for BUN. The GWAS results for BUN and MBUN were very different, in spite of the strong genetic correlation between the two traits. We detected 12 QTL for MBUN, on bovine chromosomes 2, 3, 9, 11, 12, 14 and X, and one QTL for BUN on chromosome 13. The QTL detected on chromosomes 11, 14 and X overlapped with QTL detected for MUN. The GWAS results were highly sensitive to the subset of records used. Hence, caution is warranted when interpreting GWAS based on small datasets, such as for BUN. MBUN may provide an attractive alternative to perform a more powerful GWAS to detect QTL for BUN.

## Background

The dairy industry is under increasing pressure to improve its sustainability and reduce its environmental footprint. Reducing urinary nitrogen excretion from grazing dairy cattle would lead to a reduced environmental impact [1]. While it would be difficult to collect a large enough dataset with urinary nitrogen excretion records to allow genetic selection, related indicator traits are more readily available. Blood urea nitrogen (BUN) is a biomarker for urinary nitrogen excretion [2].

However, direct measures of BUN level remain challenging to obtain because collecting blood samples routinely may not be feasible on most dairy farms. Alternatively, BUN can be predicted using mid-infrared (MIR) spectroscopy of a milk sample [3, 4]. Previous studies have reported genetic correlations between MIR predicted BUN (MBUN) and BUN that range from 0.90 [5] to 0.98 [6]. Hence, for the purpose of genetic analyses, MBUN and BUN can be considered as the same trait. Including MBUN in a reference population for genomic prediction can increase the prediction accuracy for BUN [6]. To further increase the accuracy of genomic prediction, genome-wide association studies (GWAS) could be used to select sequence variants associated with

\*Correspondence: [irene.vandenberga@agriculture.vic.gov.au](mailto:irene.vandenberga@agriculture.vic.gov.au)

<sup>1</sup> AgriBio, Centre of AgriBioscience, Agriculture Victoria, 5 Ring Road, Bundoora, VIC 3082, Australia

Full list of author information is available at the end of the article



BUN. Including sequence variants among the prediction markers can increase the accuracy of genomic prediction [7–9]. Similar to genomic prediction, large datasets are required to increase the power of GWAS. While the power of a GWAS for BUN may be limited due to the difficulties in obtaining a sufficient number of samples, MBUN may provide an alternative to detect quantitative trait loci (QTL) associated with BUN. While several GWAS have been published for milk urea nitrogen (MUN) [10–12], a trait that is highly related to BUN [5, 13], to our knowledge, no GWAS has been published for either BUN or MBUN. Therefore, the objective of our study was to perform GWAS for BUN measured from blood samples and MBUN predicted from milk samples using spectroscopy, compare these two GWAS and detect QTL for both traits, and compare the detected QTL with previously reported QTL for MUN.

## Methods

Details of the BUN and MBUN phenotypes, and the genotypes used for this analysis have been previously described [5, 14]. We used one record per cow, with records from 2098 cows for BUN and from 18,120 cows for MBUN. The cows with BUN records were a subset of the cows with MBUN records. The majority of these records were from Holstein animals, but Jersey, Australian Red, Ayrshire and crossbred animals were also included in the dataset to maximise mapping power and precision [15]. While the number of Australian Red and Ayrshire animals was very small, they were included in the dataset because 987 of the crossbreds were part Australian Red and/or Ayrshire. Table 1 shows a breakdown of the number of records, their mean and standard deviation per breed for each trait. BUN was derived from blood samples, and MBUN from milk samples, according to the protocols described in Luke et al. [3] and Ho et al. [4].

For all cows, high-density genotypes from the Bovine high-density (HD) Genotyping BeadChip and imputed

whole-genome sequence data were available. The genotyping and imputation pipelines are described in van den Berg et al. [14]. Briefly, cows were genotyped with low- to medium-density single nucleotide polymorphism (SNP) panels. Raw genotypes were filtered based on the GenCall score, and then imputed to the BovineSNP50K BeadChip using a mixed Holstein and Jersey imputation reference population. Genotypes were subsequently further imputed to HD and finally to whole-genome sequence. The reference populations used for imputation to 50K, HD and sequence data included 14,722, 2700 and 4190 *Bos taurus* cattle, respectively. The latter corresponded to Run8 of the 1000 Bull Genomes Project [16, 17]. Missing genotypes in the reference whole-genome sequence data were imputed using the Beagle software v.4.1 [18], and only bi-allelic variants with an allele count of at least 3 and a Beagle  $R^2$  higher than 0.9 were retained. Imputation to 50K and HD was done using the Fimpute software v.3 [19], whereas imputation to sequence level was carried out with the Minimac4 software [20]. All variants were mapped to the ARS-UCD1.2 reference genome [21]. After filtering on minor allele frequency ( $MAF \geq 0.005$ ) and Minimac imputation  $R^2 (\geq 0.4)$ , 15,625,438 SNPs were retained for subsequent GWAS.

HD genotypes (717,463 SNPs) were used to construct a genomic relationship matrix (GRM) and perform a principal component analysis (PCA) using the GCTA tool [22]. Scores for the first principal component (PC1) were included in the analyses to account for differences between the Holstein and Jersey breeds as described in van den Berg et al. [5]. PC1 was highly correlated (0.9997) with ADMIXTURE ancestry fractions [5]. Previous estimates [5] of fixed effects and covariates (the PC1, test month, herd-year-season, days in milk and age) were used to adjust phenotypes before performing the GWAS. GWAS was done using the GCTA tool [22], including the GRM (based on HD) for the following scenarios:

**Table 1** Number of records (N), mean and standard deviation (SD) per breed for blood urea nitrogen (BUN) and MIR predicted BUN (MBUN)

Breed	BUN			MBUN		
	N	Mean	SD	N	Mean	SD
Holstein	1569	5.6	2.1	12,660	4.9	1.9
Jersey	59	4.2	1.3	1857	5.8	1.6
Australian Red	2	5.3	0.5	95	6.9	2.2
Ayrshire	0	–	–	12	6.3	1.9
Crossbred	468	5.2	2.6	3496	4.8	2.1
All	2098	5.5	2.2	18,120	5.0	1.9

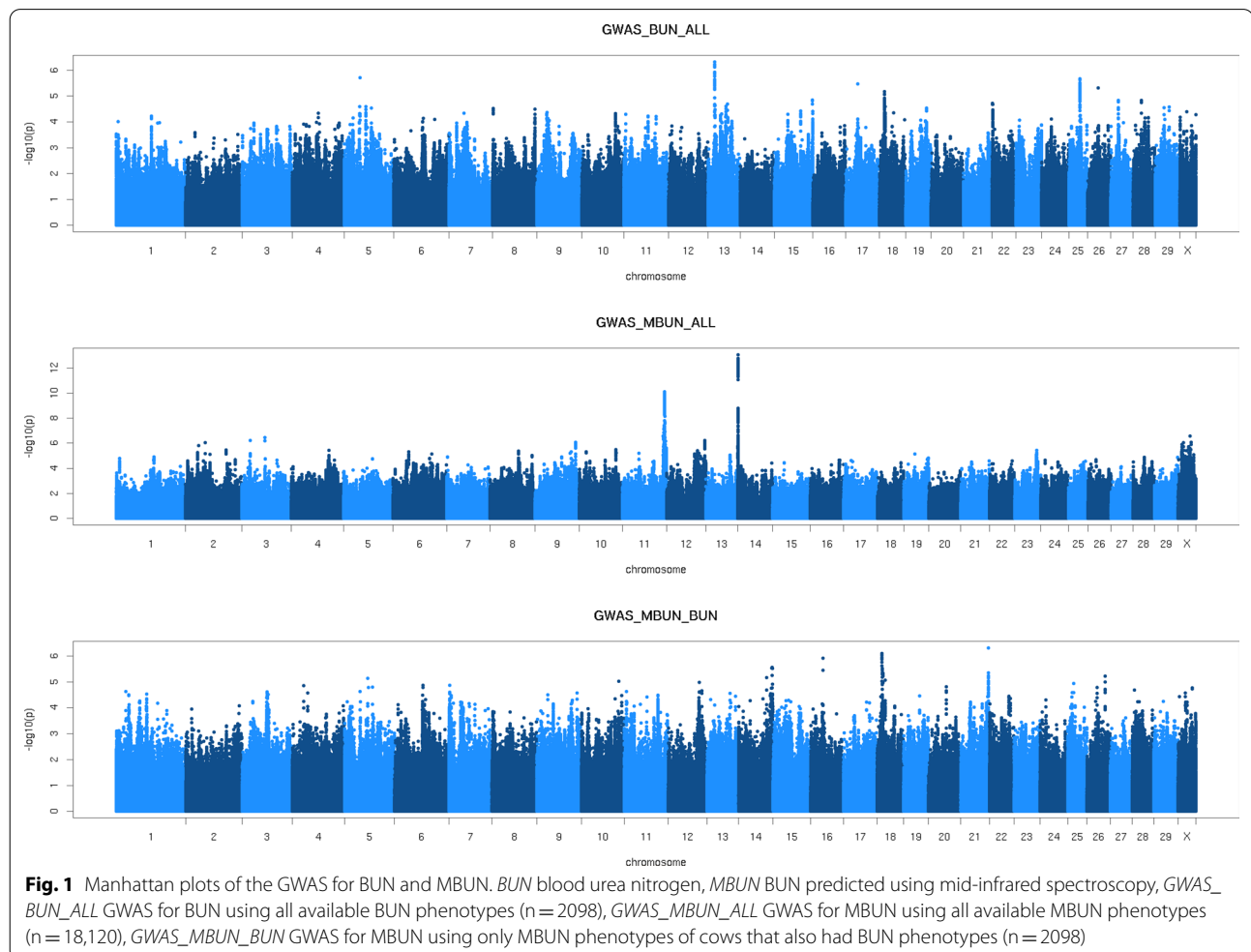
1. GWAS\_BUN\_ALL: GWAS for BUN, using all 2098 cows with BUN records;
2. GWAS\_MBUN\_ALL: GWAS for MBUN, using all 18,120 cows with MBUN records;
3. GWAS\_MBUN\_BUN: GWAS for MBUN, using only the records from the 2098 cows with BUN records;
4. GWAS\_MBUN\_2K, using the records from the 2000 randomly selected cows (repeated 5 times).

Because of the reduced dataset, GWAS\_BUN\_ALL, GWAS\_MBUN\_BUN and GWAS\_MBUN\_2K were carried out only for variants with a  $MAF \geq 0.05$ . We considered that all the variants with a  $p\text{-value} \leq 10^{-6}$  were significant, and calculated the false discovery rate (FDR) as  $FDR = (n\text{Variants} \times 10^{-6}) / n\text{Significant}$ , where  $n\text{Variants}$  is the total number of variants in the GWAS and  $n\text{Significant}$  is the number of variants in the GWAS with a  $p\text{-value} \leq 10^{-6}$ . QTL intervals were subsequently defined by grouping variants that were separated by less than 1 Mb in the same QTL interval. Genomic inflation factors of the GWAS were

estimated using the “estlambda” function in the GenABEL R package [23]. To test the association between previously reported dairy cattle QTL [24, 25] and the QTL detected for BUN and MBUN, we repeated the GWAS for which QTL associated with BUN or MBUN were located in the same region as previously reported QTL including the previously reported QTL as covariates.

## Results and discussion

In total, 640 and 5 significant variants were detected for MBUN and BUN, respectively, corresponding to an FDR of 0.02 and 2.27, respectively. Additional file 1: Figure S1 shows the Q-Q-plots of the observed and expected p-values in the GWAS for BUN and MBUN. Genomic inflation factors were equal 0.98 and 1.07 for BUN and MBUN, respectively. The much larger number of records for MBUN (18,120) than for BUN (2098) resulted in an increased power to detect QTL for MBUN compared to BUN (Fig. 1). Table 2 lists the QTL that were detected for each trait. Only one QTL on chromosome 13 was



**Table 2** Positions of potential QTL detected for BUN and MBUN

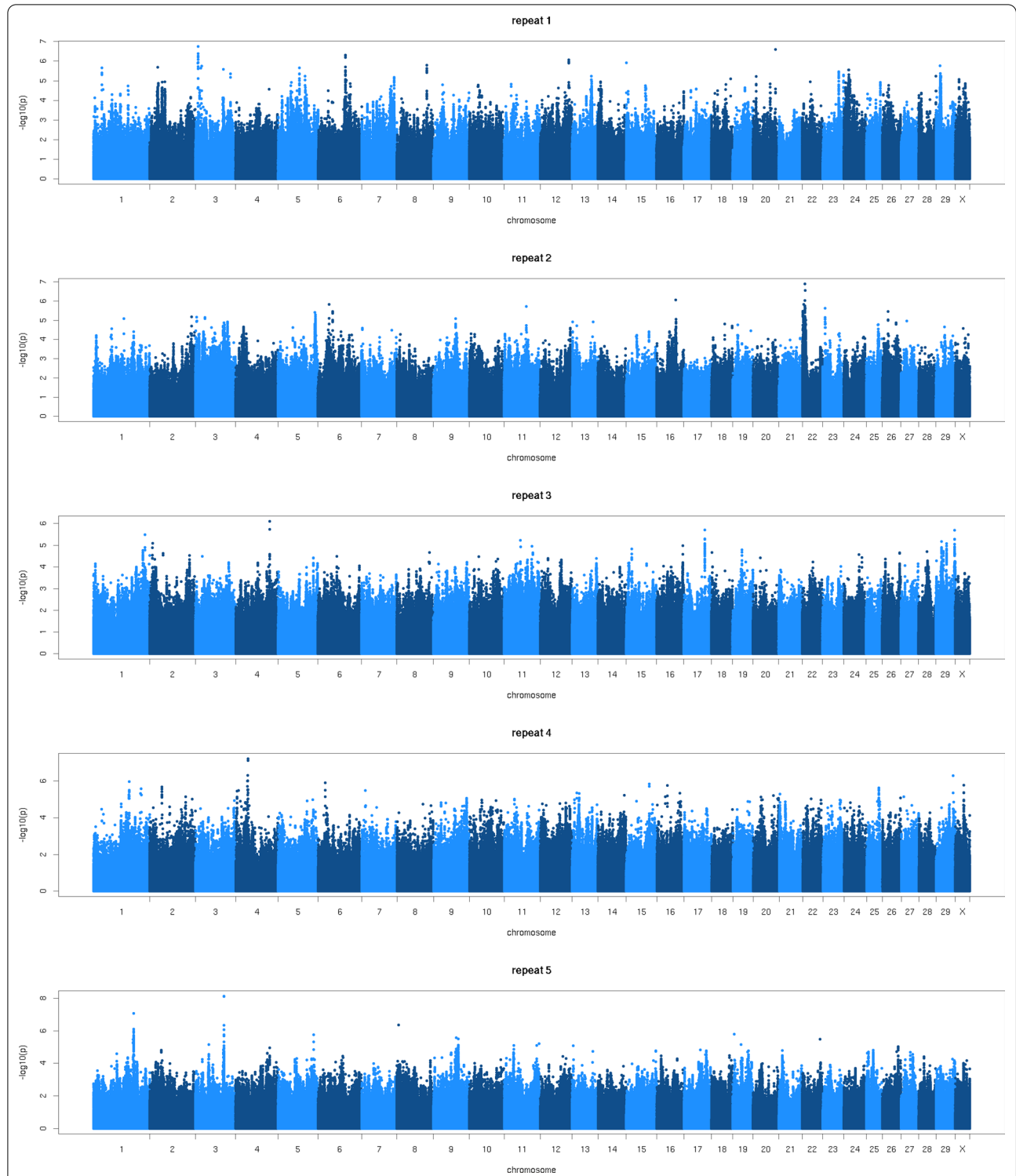
Chr	Pos (bp)	Trait	$P_{BUN}$	$P_{MBUN}$	Annotation	Genes	Start (bp)	End (bp)	N
2	47,485,307	MBUN	$6.5 \times 10^{-1}$	$9.4 \times 10^{-7}$	Intron	<i>EPC2</i>	47,485,307	47,485,307	1
3	18,190,277	MBUN	$3.1 \times 10^{-2}$	$6.3 \times 10^{-7}$	Upstream	<i>CRCT1</i>	18,190,277	18,190,277	1
3	55,238,179	MBUN	$1.2 \times 10^{-3}$	$3.5 \times 10^{-7}$	Intergenic	<i>PKN2-ENSBTAG00000051499</i>	55,238,179	55,257,795	33
9	97,711,991	MBUN	$3.0 \times 10^{-2}$	$8.4 \times 10^{-7}$	Intron	<i>PRKN</i>	97,711,991	97,711,991	1
11	99,897,676	MBUN	$4.1 \times 10^{-1}$	$2.7 \times 10^{-7}$	Intron	<i>ENSBTAG00000054738</i>	99,897,676	99,916,846	6
11	103,271,858	MBUN	$7.7 \times 10^{-1}$	$7.7 \times 10^{-11}$	Synonymous	<i>GLT6D1</i>	101,401,373	103,606,366	382
11	105,110,370	MBUN	$8.0 \times 10^{-1}$	$6.8 \times 10^{-7}$	Intron	<i>RXRA</i>	105,110,370	105,110,370	1
12	84,682,464	MBUN	$2.8 \times 10^{-1}$	$6.0 \times 10^{-7}$	Intergenic	<i>IRS2-RF00001</i>	84,682,431	84,699,022	3
13	15,837,206	BUN	$4.8 \times 10^{-7}$	$6.6 \times 10^{-2}$	Intergenic	<i>ENSBTAG00000048047-GATA3</i>	15,835,519	15,839,775	5
14	631,698	MBUN	$1.5 \times 10^{-1}$	$8.6 \times 10^{-14}$	Upstream	<i>BOP1</i>	512,818	1,278,273	209
X	31,752,695	MBUN	$2.7 \times 10^{-2}$	$9.0 \times 10^{-7}$	Intergenic	<i>ENSBTAG00000018311-IDS</i>	31,752,695	31,752,695	1
X	100,768,475	MBUN	$4.2 \times 10^{-1}$	$2.7 \times 10^{-7}$	Intergenic	<i>MAOA-PPP1R2C</i>	100,768,475	100,768,475	1
X	107,587,617	MBUN	$3.1 \times 10^{-1}$	$8.7 \times 10^{-7}$	Intergenic	<i>MAGEB16-ENSBTAG00000040406</i>	107,587,617	107,587,617	1

BUN, blood urea nitrogen; MBUN, BUN predicted using mid-infrared spectroscopy; chr, chromosome; pos, position of the most significant variant associated with the QTL; trait, trait for which the QTL is significant; bp, base pair according to the ARS-UCD1.2 annotation;  $P_{BUN}$ , p-value in the GWAS for BUN;  $P_{MBUN}$ , p-value in the GWAS for MBUN; annotation, annotation of the most significant variant; gene, gene in which the most significant region was located or, if the most significant variant was intergenic, the genes between which the most significant variant was located, start, start of the QTL interval; end, end of the QTL interval; N, number of variants with  $p \leq 10^{-6}$  in the QTL interval

detected for BUN (15,837,206 bp,  $P_{BUN} = 4.8 \times 10^{-7}$ ). The variant associated with this QTL was an intergenic variant, located between the *ENSBTAG00000048047* and *GATA binding protein 3 (GATA3)* genes, and was not significant in the GWAS for MBUN ( $P_{MBUN} = 0.07$ ). Given the extremely high FDR for BUN, this QTL is likely a false positive.

We detected 12 QTL for MBUN, located on chromosomes 2, 3, 9, 11, 12, 14 and X. The most significant variant associated with MBUN was located at 631,698 bp on chromosome 14 ( $p = 8.6 \times 10^{-14}$ ), upstream of the *BOP1 ribosomal biosynthesis factor (BOP1)* gene. Several previous studies reported *BOP1* as a candidate gene associated with milk production traits [26–28]. The QTL interval on chromosome 14 encompassed the *diacylglycerol O-acyltransferase homolog 1 (DGAT1)* gene, a well-known causal gene for milk production traits in dairy cattle [24]. After including the causal variant for the *DGAT1* QTL in the model as a fixed effect, none of the remaining variants on chromosome 14 were significant (see Additional file 2: Figure S2). Hence, *DGAT1* was in high LD with all variants within the peak and is likely the causal variant that underlies the QTL. The second most significant QTL was a synonymous variant in the *glycosyltransferase 6 domain containing 1 (GLT6D1)* gene, located at 103,271,858 bp on chromosome 11. *GLT6D1* is associated with periodontitis in humans [29]. The QTL interval on chromosome 11 also included the *progesterone-associated endometrial protein (PAEP)* gene, a candidate gene for milk production traits in dairy cattle [25], and the *alpha 1-3-N-acetylgalactosaminyltransferase and alpha*

*1-3-galactosyltransferase (ABO)* gene, which has been reported as a candidate gene for protein yield in dairy cattle [30] and determines human blood type [31]. The most significant variant in the GWAS for MBUN was located in an intron of the *PAEP* gene at 103,262,933 bp. When including this variant as a fixed effect in the model, the large peak in the area disappeared (see Additional file 3: Figure S3), and none of the previously reported top variants for the QTL were significant anymore, which indicates that rather than the three QTL detected on chromosome 11, we detected only one QTL that is in high LD with *PAEP*. When the variant in *PAEP* was included as a fixed effect in the model, the only remaining significant variant was an intron in the *vav guanine nucleotide exchange factor 2 (VAV2)* gene located at 104,765,599 bp with a p-value of  $9.7 \times 10^{-7}$ . Ariyaratne et al. [10] reported a QTL for MUN in the same region on chromosome 11. MUN and MBUN had a genetic correlation of 0.77 in the dataset used in the current study [5], hence QTL for MBUN and MUN were expected to overlap. Both MBUN and MUN are derived using MIR spectroscopy data of a milk sample, which may contribute to the strong genetic correlation between MBUN and MUN and similarity in the GWAS results. In a GWAS for MUN using the same individuals as in our current analysis, QTL for MUN were detected that overlapped with QTL for MBUN on chromosomes 11, 14 and X [14]. On chromosome 11, the variant located at 103,271,858 bp was the most significant variant on this chromosome for both MBUN and MUN, with a p-value of  $5.4 \times 10^{-16}$  for MUN. The variants that were detected for MBUN at 631,698 bp



**Fig. 2** Manhattan plots of the GWAS for MBUN using different subsets of records. *MBUN* blood urea nitrogen (BUN) predicted using mid-infrared spectroscopy. The five repeats show GWAS carried out using MBUN phenotypes from 2000 randomly selected cows

on chromosome 14 and 107,587,617 bp on chromosome X have p-values of  $1.2 \times 10^{-21}$  and  $1.1 \times 10^{-6}$  for MUN, respectively (unpublished observations). MBUN had moderate genetic correlations with fat yield (0.28) and fat percentage (0.35) [5], which may explain the overlap between QTL detected for MBUN and the well-known milk production QTL on chromosomes 11 and 14 [24, 25]. However, given that the majority of the significant peaks identified in the GWAS for MBUN were associated with milk production, the GWAS results for any MIR-predicted trait may be heavily biased by milk characteristics. Several GWAS for MIR wave numbers have detected major QTL on chromosomes 3 [32], 11 [32–34] and 14 [32, 33, 35, 36] in the same regions where we detected QTL for MBUN.

The GWAS results for BUN and MBUN were very different, which was surprising given the genetic correlation of 0.90 between the two traits [5]. None of the QTL detected for MBUN were close to significance in the GWAS for BUN. Hence, it is possible that, in spite of the strong genetic correlation between MBUN and BUN, some of the variation in MBUN picked up in the GWAS is related to the variation of the MIR spectrum rather than to variation in BUN. Alternatively, the differences between the GWAS for MBUN and BUN may be due to the smaller subset of animals that had BUN records. To explore this, we repeated the GWAS for MBUN using only the animals with BUN phenotypes (GWAS\_MBUN\_BUN, Fig. 1). Using this dataset the peaks that we detected in the GWAS\_MBUN\_ALL analysis disappeared, which strongly suggests that the GWAS results were very sensitive to the set of individuals used. The dataset that we analysed contained multiple breeds, including crossbreds, hence the difference in GWAS results between subsets could also be due to differences in breed composition.

To test if the differences between GWAS\_MBUN\_ALL and GWAS\_MBUN\_BUN were caused by (1) particular characteristics of the subset of animals with BUN records or (2) by the smaller sample size, we carried out a GWAS with records from 2000 randomly selected cows (GWAS\_MBUN\_2K), and as shown in Fig. 2, the results differed between datasets. None of the GWAS\_MBUN\_2K detected any of the larger QTL detected in the GWAS\_MBUN\_ALL. In each of the GWAS\_MBUN\_2K, small peaks with minimum p-values of approximately  $10^{-6}$  to  $10^{-8}$  were identified, but they were at different positions in each of the datasets. This implies that the GWAS results obtained by using small datasets (in this case, around 2000 cows) should be interpreted with caution, and larger datasets may be required to detect peaks that are less sensitive to the particular set of data analysed. A MIR predicted trait for which records

can easily be generated for a large number of animals, such as MBUN, may provide an attractive alternative to perform a more powerful GWAS for hard-to-measure traits, such as BUN. Since there is a strong genetic correlation between MBUN and BUN, they can be considered as the same trait [5, 6], and hence the QTL detected for MBUN could be interpreted as QTL for BUN. This indicates the potential of using MIR equations from a breeding perspective. However, although Ho et al. [4] reported a comparable prediction accuracy from herd/animal independent validation and herd-year by herd-year validation, further research is required to develop prediction equations for MBUN that can be transferred across environments.

## Conclusions

The GWAS results for BUN and MBUN were very different, in spite of the strong genetic correlation between the two traits. We detected 12 QTL for MBUN, located on chromosomes 2, 3, 9, 11, 12, 14 and X. The QTL on chromosomes 11, 14 and X overlapped with previous QTL detected for milk production traits and/or MUN. We detected one QTL for BUN using a dataset of about 2000 cows that was located on chromosome 13. However, when we repeated the GWAS for MBUN on smaller (2000) subsets of the dataset i.e. so that their size was comparable to that for BUN, the GWAS results were very sensitive to the subset of records used. Hence, using approximately 2000 cow phenotypes as was done for the GWAS for BUN may not be sufficient for accurate QTL detection, and caution is warranted when interpreting GWAS results based on small datasets. Based on the strong genetic correlation between MBUN and BUN that was estimated in previous studies, MBUN may provide an attractive alternative to perform a more powerful GWAS to detect QTL for BUN.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-022-00719-5>.

**Additional file 1: Figure S1.** Quantile–Quantile (Q–Q) plots of expected and observed p-values for the GWAS of blood urea nitrogen (BUN) and BUN predicted using mid-infrared spectroscopy (MBUN).

**Additional file 2: Figure S2.** GWAS for MBUN (blood urea nitrogen predicted using mid-infrared spectroscopy) on chromosome 14 including the causal variant for *DGAT1* as covariate.

**Additional file 3: Figure S3.** GWAS for MBUN (blood urea nitrogen predicted using mid-infrared spectroscopy) on chromosome 11 including an intron variant in *PAEP* as covariate.

## Acknowledgements

This study was part of the DairyBio program. DairyBio is jointly funded by Dairy Australia (Melbourne, Australia) and Agriculture Victoria (Melbourne,

Australia) and The Gardiner Foundation (Melbourne, Australia). The authors thank Di Mapleson, Brigid Ribaux from Agriculture Victoria Research, and the staff at Ellinbank Dairy Research Centre (Ellinbank, Australia) for their technical expertise and assistance in measuring BUN. We also thank staff from Datagene (Bundoora, Australia) for their work in sampling and on-farm coordinating role in this study, and the farmers who participated in this project. Hico Pty Ltd (Maffra, Victoria, Australia) and TasHerd Pty Ltd (Hadspen, Australia) are gratefully acknowledged for collecting milk samples and providing MIR spectral data. Regional Laboratory Services (Benalla, Victoria, Australia) are thanked for measuring the concentrations of blood urea used in this study. The authors would like to extend their gratitude to Dr. Bolormaa Sunduimijid and Dr. Iona MacLeod for imputation, and Dr. Coralie Reich, Brett Mason and Dr. Amanda Chamberlain, for genotyping, and the farmers and DataGene for access to data used in this study. We thank partners from the 1000 bull genomes project for the data access. We acknowledge Dr. Paul Stothard and the team at the University of Alberta for collating annotation information of sequence variants used in this study.

#### Author contributions

IB designed the study, carried out the GWAS, and wrote the manuscript. TL provided BUN phenotypes. PH derived MBUN phenotypes. TN imputed the sequence data. MH and JP helped design the study. All authors read and approved the final manuscript.

#### Funding

This research was part of the DairyBio program, which is jointly funded by Dairy Australia (Melbourne, Australia), Agriculture Victoria (Melbourne, Australia) and The Gardiner Foundation (Melbourne, Australia).

#### Availability of data and materials

Data requests should be directed to DataGene Ltd (Melbourne, Australia) as custodians of data on Australian dairy cows. Research-related requests for access to the data may be accommodated on a case-by-case basis.

#### Declarations

##### Ethics approval and consent to participate

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).

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>AgriBio, Centre of AgriBioscience, Agriculture Victoria, 5 Ring Road, Bundoora, VIC 3082, Australia. <sup>2</sup>School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia.

Received: 16 December 2021 Accepted: 5 April 2022

Published online: 18 April 2022

#### References

- O'Callaghan P, Kelly-Quinn M, Jennings E, Antunes P, O'Sullivan M, Fenton O, et al. The environmental impact of cattle access to watercourses: a review. *J Environ Qual*. 2019;48:340–51.
- Kohn RA, Dinneen MM, Russek-Cohen E. Using blood urea nitrogen to predict nitrogen excretion and efficiency of nitrogen utilization in cattle, sheep, goats, horses, pigs, and rats. *J Anim Sci*. 2005;83:879–89.
- Luke TDW, Rochfort S, Wales WJ, Bonfatti V, Maret L, Pryce JE. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. *J Dairy Sci*. 2019;102:1747–60.
- Ho PN, Luke TDW, Pryce JE. Validation of milk mid-infrared spectroscopy for predicting the metabolic status of lactating dairy cows in Australia. *J Dairy Sci*. 2021;104:4467–77.
- van den Berg I, Ho PN, Haile-Mariam M, Beatson PR, O'Connor E, Pryce JE. Genetic parameters of blood urea nitrogen and milk urea nitrogen concentration in dairy cattle managed in pasture-based production systems of New Zealand and Australia. *Anim Prod Sci*. 2021;61:1801–10.
- van den Berg I, Ho PN, Luke TDW, Haile-Mariam M, Bolormaa S, Pryce JE. The use of milk mid-infrared spectroscopy to improve genomic prediction accuracy of serum biomarkers. *J Dairy Sci*. 2021;104:2008–17.
- Brøndum RF, Su G, Janss L, Sahana G, Guldborndtsen B, Boichard D, et al. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J Dairy Sci*. 2015;98:4107–16.
- van den Berg I, Boichard D, Lund MS. Sequence variants selected from a multi-breed GWAS can improve the reliability of genomic predictions in dairy cattle. *Genet Sel Evol*. 2016;48:83.
- Van Raden PM, Tooker ME, O'Connell JR, Cole JB, Bickhart DM. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet Sel Evol*. 2017;49:32.
- Ariyaratne HBPC, Correa-Luna M, Blair HT, Garrick DJ, Lopez-Villalobos N. Identification of genomic regions associated with concentrations of milk fat, protein, urea and efficiency of crude protein utilization in grazing dairy cows. *Genes (Basel)*. 2021;12:456.
- Bouwman AC, Schopen GCB, Bovenhuis H, Visser MHPW, van Arendonk JAM. Genome-wide scan to detect quantitative trait loci for milk urea nitrogen in Dutch Holstein-Friesian cows. *J Dairy Sci*. 2010;93:3310–9.
- Cecchinato A, Ribeca C, Chessa S, Cipolat-Gotet C, Maretto F, Casellas J, et al. Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal*. 2014;8:1062–70.
- Gustafsson AH, Palmquist DL. Diurnal variation of rumen ammonia, serum urea, and milk urea in dairy cows at high and low yields. *J Dairy Sci*. 1993;76:475–84.
- van den Berg I, Ho PN, Nguyen TV, Haile-Mariam M, MacLeod IM, Beatson PR, et al. GWAS and genomic prediction of milk urea nitrogen in Australian and New Zealand dairy cattle. *Genet Sel Evol*. 2022;54:15.
- van den Berg I, Boichard D, Lund MS. Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. *J Dairy Sci*. 2016;99:8932–45.
- Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 2014;46:858–65.
- Hayes BJ, Daetwyler HD. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annu Rev Anim Biosci*. 2019;7:89–102.
- Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 2018;103:338–48.
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–7.
- Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience*. 2020;9:giaa021.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23:1294–6.
- Grisart B, Farnir F, Karim L, Cambisano N, Kim JJ, Kvasz A, et al. Genetic and functional confirmation of the causality of the *DGAT1* K232A quantitative trait nucleotide in affecting milk yield and composition. *Proc Natl Acad Sci USA*. 2004;101:2398–403.

25. Raven LA, Cocks BG, Kemper KE, Chamberlain AJ, Vander Jagt CJ, Goddard ME, et al. Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. *Mamm Genome*. 2016;27:81–97.
26. Oliveira HR, Lourenco DAL, Masuda Y, Misztal I, Tsuruta S, Jamrozik J, et al. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. *J Dairy Sci*. 2019;102:9995–10011.
27. Sanchez MP, Govignon-Gion A, Croiseau P, Fritz S, Hozé C, Miranda G, et al. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genet Sel Evol*. 2017;49:68.
28. Cai Z, GuldbRANDtsen B, Lund MS, Sahana G. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. *BMC Genet*. 2019;20:15.
29. Schaefer AS, Richter GM, Nothnagel M, Manke T, Dommisch H, Jacobs G, et al. A genome-wide association study identifies *GLT6D1* as a susceptibility locus for periodontitis. *Hum Mol Genet*. 2010;19:553–62.
30. MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*. 2016;17:144.
31. Yamamoto F, Clausen H, White T, Marken J, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. *Nature*. 1990;345:229–33.
32. Zaalberg RM, Janss L, Buitenhuis AJ. Genome-wide association study on Fourier transform infrared milk spectra for two Danish dairy cattle breeds. *BMC Genet*. 2020;21:9.
33. Benedet A, Ho PN, Xiang R, Bolormaa S, De Marchi M, Goddard ME, et al. The use of mid-infrared spectra to map genes affecting milk composition. *J Dairy Sci*. 2019;102:7189–203.
34. Rutten MJM, Bovenhuis H, Heck JML, van Arendonk JAM. Predicting bovine milk protein composition based on Fourier transform infrared spectra. *J Dairy Sci*. 2011;94:5683–90.
35. Wang Q, Bovenhuis H. Genome-wide association study for milk infrared wavenumbers. *J Dairy Sci*. 2018;101:2260–72.
36. Tiplady KM, Lopdell TJ, Reynolds E, Sherlock RG, Keehan M, Johnson TJJ, et al. Sequence-based genome-wide association study of individual milk mid-infrared wavenumbers in mixed-breed dairy cattle. *Genet Sel Evol*. 2021;53:62.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

