



Identifying cows with subclinical mastitis by bulk single nucleotide polymorphism genotyping of tank milk

G. Blard, Z. Zhang, W. Coppieters, and M. Georges¹

Unit of Animal Genomics, Department of Animal Production, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34),
1 Avenue de l'Hôpital, 4000-Liège (Sart Tilman), Belgium

ABSTRACT

Mastitis remains the most important health issue in dairy cattle. Improved methods to identify cows developing subclinical mastitis would benefit farmers. We herein describe a novel method to determine the somatic cell counts (SCC) of individual cows by bulk genotyping a sample of milk from the milk tank with panels of genome-wide single nucleotide polymorphisms (SNP). We developed a simple linear model to estimate the contribution of individual cows to the genomic DNA present in the tank milk from 1) the known genotypes of individual cows for the interrogated SNP and 2) the ratio of SNP alleles in the tank milk. Using simulations, we estimate that 3,000, 50,000, and 700,000 SNP are sufficient to accurately ($R^2 > 0.98$) estimate individual SCC in tanks containing milk from 25, 100, and 500 cows, respectively. Using actual data, we demonstrate that the SCC of 21 cows can be estimated with a coefficient of determination of 0.60 using approximately 9,000 SNP. The proposed method increases the value of the proposition of SNP genotyping individual cows for genomic selection purposes.

Key words: mastitis, single nucleotide polymorphism, bulk genotyping, tank milk

INTRODUCTION

Mastitis is generally regarded as the most important health issue in dairy farming. Costs to the farmer result from treatment, decreased milk yield and value, and culling. In the Netherlands, the annual cost of mastitis to farmers has been estimated at approximately 80 euros per cow per year (Hogeveen et al., 2011).

One often distinguishes clinical from subclinical mastitis. Mastitis is said to be clinical when resulting in either milk or udder anomalies that are manifest to the

farmer. Subclinical mastitis is defined as the presence of microorganisms in combination with elevated SCC in the milk; 200,000 cells/mL is often used as the SCC cut-off value (Reneau, 1986; Dohoo and Leslie, 1991; Schukken and Kremer, 1996). Subclinical mastitis is costly on its own because of its negative effect on milk yield, which decreases log linearly with SCC (Tyler et al., 1989).

At present, the control of mastitis is primarily driven by the detection and treatment of clinical mastitis. More effective detection of subclinical mastitis would be a valuable addition to mastitis control. In most circumstances, SCC for individual cows are only measured periodically (typically every 4–6 wk), precluding close monitoring of the udder health status of individual cows. Advanced versions of milking robots will precociously detect changes in conductivity or even SCC, yet automatic milking has been associated with a decrease, rather than an improvement, in udder health (Hovinen and Pyörälä, 2011). Thus, novel approaches to monitor SCC of individual cows could be a benefit to the sector.

We herein propose a method that allows determination of SCC for individual milking cows (and, hence, detection of cows with subclinical mastitis) by genotyping a sample of milk from the farm's tank for a panel of SNP. At present, 3 favored bovine SNP panels are available, including approximately 3,000 (**3K**), 50,000 (**50K**), and 700,000 (**700K**) SNP distributed across the genome. All of these are developed and commercialized by Illumina Inc. (San Diego, CA; www.illumina.com). The proposed method assumes that individual genotypes for the same SNP are available for all cows on the farm. As genomic selection is becoming common practice in dairy cattle, an increasing proportion of the cow population is being SNP genotyped. As an example, in the spring of 2011, more than 85,000 Holstein-Friesian animals had been genotyped with one of the SNP arrays mentioned above in the United States alone. The majority of these were females and their proportion was rapidly increasing with time. We anticipate that farms in which all cows will be SNP genotyped will be commonplace in the near future.

Received November 23, 2011.

Accepted March 13, 2012.

¹Corresponding author: michel.georges@ulg.ac.be

MATERIALS AND METHODS

Method

Assume a dairy farm with n cows and that the volume of milk contributed by cow i to the milk tank is known (as it is often in reality) and equals v_i . Next, assume that the SSC per liter of milk contributed to the tank by cow i is c_i and that all cows on the farm have been genotyped for an array interrogating m SNP. Assume that g_{ij} is an indicator variable for the genotype of cow i for SNP j (taking a value of 0 for genotype AA, 0.5 for genotype AB, and 1 for genotype BB). Also, assume that a sample of DNA extracted from the milk tank has been genotyped with the same array, and that the estimated proportion of allele B of SNP j in the milk sample is f_j . The proposed method stems from the realization that one can determine the proportion of genomic DNA (and, hence, somatic cells) contributed by each cow to the genomic DNA present in the tank milk by confronting the SNP genotypes of individual cows with the ratios of SNP alleles in the tank milk. More specifically, the proportion of somatic cells contributed by cow i to the tank (pc_i), can be estimated from a set of m linear equations of the form

$$f_j = \sum_{i=1}^n pc_i \times g_{ij} + \varepsilon_j$$

by minimizing

$$SSE = \sum_{j=1}^m \varepsilon_j^2.$$

where ε_j is the error term in estimating f_j , and SSE is the sum of squared errors for all SNP. From the obtained pc_i values, one can then determine the SCC per liter for cow i relative to the rest of the herd (rc_i) as

$$rc_i = pc_i / pv_i = c_i / \bar{c},$$

where pv_i is the known proportion of the tank's volume contributed by cow i and \bar{c} is the average SCC per liter in the herd. If \bar{c} is known (it can be measured in the milk tank), the absolute SCC per liter for cow i can be computed from rc_i .

Simulated Data

We simulated farms with 25, 100, and 500 cows. The cows' daily milk production (L/d) was assumed normally distributed with a mean of 30 L and standard deviation 0.2 L. Somatic cell counts per liter were assumed to be proportional ($\times 8 \times 10^6$) to a chi-squared

distribution (2 df). We assumed the use of SNP arrays with 3,000, 50,000, or 700,000 SNP, corresponding closely to presently available commercial products. Allelic frequencies were assumed to have a uniform distribution. Cows were assumed to have genotypes for each SNP (in practice, missing values would be filled in by imputation). Estimates of B-allele proportions ($0 < f < 1$) in the tank milk were assumed normally distributed around the true frequency with residual standard error of 0.05.

Actual Data

We collected a sample of milk from a tank containing known quantities of milk (mean: 29.5 L; SD: 5.3 L) contributed the same day by 21 cows that had been previously genotyped using either of 2 custom-made ~50K Illumina arrays (Charlier et al., 2008). Individual SCC were determined on the same day for each cow using a Fossomatic FC instrument (Foss A/S, Hillerød, Denmark). The DNA was extracted from the milk sample using standard procedures and genotyped with the Illumina BovineSNP50 genotyping BeadChip (Matukumalli et al., 2009). Estimates of B-allele frequencies were directly obtained from the BeadStudio software package (Illumina Inc.). We also had SNP genotypes for 19 cows from the same farm that did not contribute milk to the tank.

To calibrate the relationship between B-allele frequency as computed with BeadStudio, and actual B-allele proportion (f_j in the Methods section), we took advantage of DNA samples available for 95 Dutch Holstein-Friesian samples that had previously been genotyped with the Illumina BovineSNP50 genotyping BeadChip (Matukumalli et al., 2009). The DNA concentrations were estimated as the average of 2 independent fluorometric measurements performed using PicoGreen (Life Technologies, Carlsbad, CA) according to the instructions of the manufacturer. We mixed equal volumes of DNA solutions from the 95 sires and genotyped the resulting DNA pool using the Illumina BovineSNP50 genotyping BeadChip. The relationship between 1) BeadStudio-derived B-allele frequency and 2) B-allele proportion computed from the known SNP genotypes and DNA concentrations of the sires is shown in Supplementary Figure 1 (available online at <http://www.journalofdairyscience.org/>) for 45,248 SNP with complete genotype data across the 95 sires. The correlation between both measures was 0.98, yet departure from linearity was obvious. We, therefore, adjusted the relation by fitting local (i.e., for each SNP) linear regressions based on 3,000 left- and 3,000 right-sided neighboring SNP (corresponding to frequency ranges

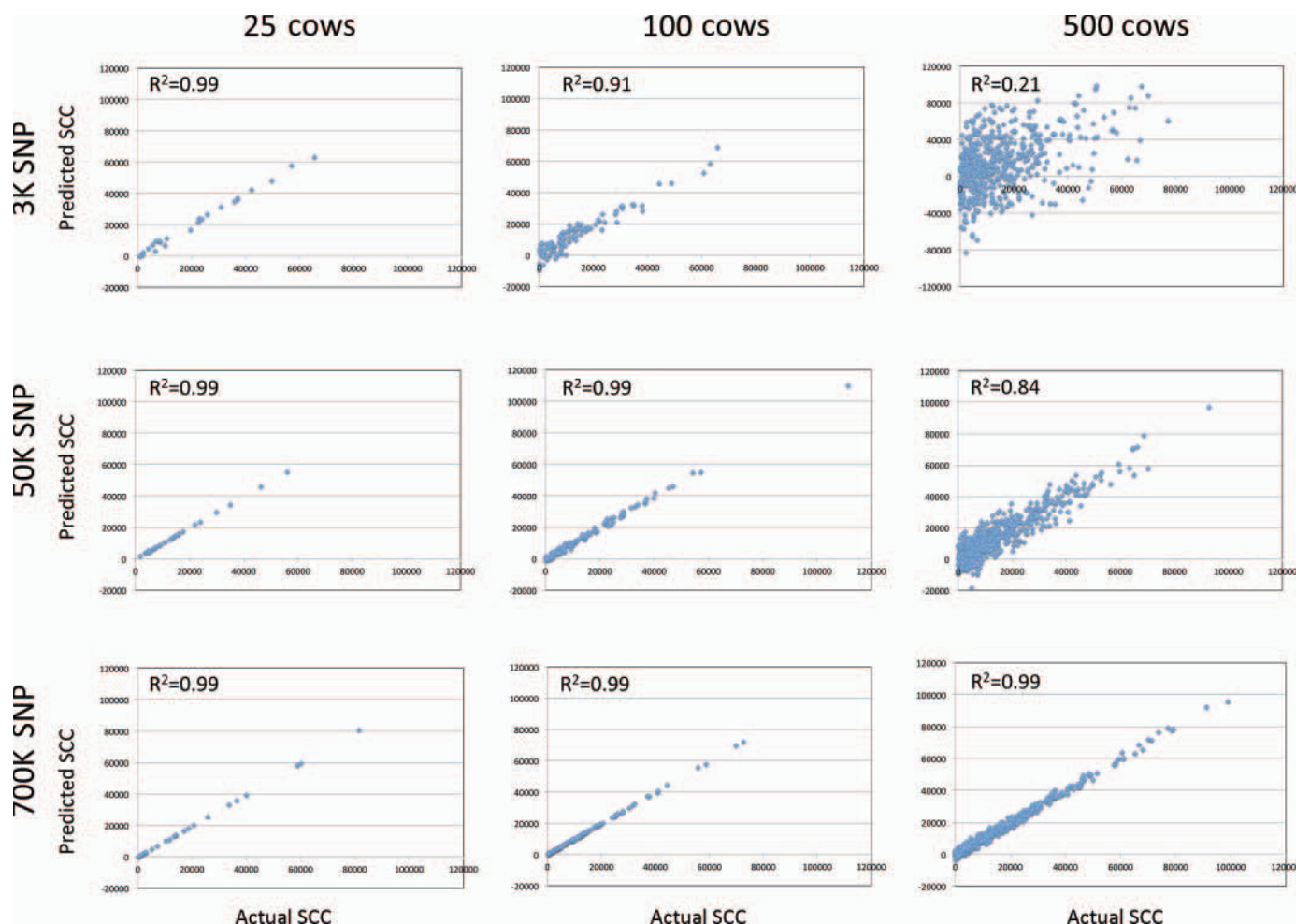


Figure 1. Simulated data. Representative examples of the relationship between actual SCC and predictions based on SNP genotyping of a sample from the tank milk. Evaluated scenarios consider 25, 100, and 500 cows genotyped for 3,000 (3K), 50,000 (50K), or 700,000 (700K) SNP. Color version available in the online PDF.

of $\sim 7.5\%$). BeadStudio-measured B-allele frequencies were converted to adjusted f_j using the SNP-specific β_0 (intercept) and β_1 (slope) estimators (Supplementary Figure 1, available online at <http://www.journalofdairyscience.org/>). Mean square errors averaged 0.005, corresponding to a residual standard deviations of 0.07 (i.e., comparable to the values used in the simulations). Statistical analyses were conducted in R software (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Simulated Data

Figure 1 shows representative examples confronting actual and estimated SCC. Predictions were very accurate with coefficient of determination >0.98 (i.e., R^2

= proportion of the variance of true SCC accounted for by predictions) with 3K or more SNP for 25 cows, 50K or more SNP for 100 cows, and 700K SNP for 500 cows. Predictors appeared unbiased under all tested conditions. Thus, our simulations indicated that bulk genotyping of tank milk could be effective for SCC monitoring of individual cows, including identification of cows with increased SCC, indicative of subclinical mastitis.

We explored the possibility of determining whether specific cows did or did not contribute milk to the tank; we evaluated the statistical significance of individual cow effects (pc_i) using a standard t -test. Supplementary Figure 2A (available online at <http://www.journalofdairyscience.org/>) shows representative examples of frequency distribution of the statistical significance $[-\log(p)]$ of the pc_i for cows that did contribute milk to the tank. For approximately 80% of the cows, P -

values were ≤ 0.0001 when using arrays interrogating $\geq 3K$ SNP for herds with 25 cows, $\geq 50K$ SNP for herds with 100 cows, and $\geq 700K$ SNP for herds with 500 cows. As expected, P -values were strongly and inversely correlated with the proportion of SCC contributed by a given cow to the milk tank (data not shown). By comparison, the frequency distribution of corresponding P -values for cows that did not contribute milk to tank (added 1 at a time in the model), largely followed the uniform distribution expected under the null hypothesis, despite a slight excess of low P -values (2-fold excess of P -values < 0.01 , including 3-fold excess of P -values < 0.001 ; Supplementary Figure 2B, available online at <http://www.journalofdairyscience.org/>).

Actual Data

We selected 8,696 SNP (1) that were interrogated by the 3 SNP panels used and (2) for which genotype information was complete for all 40 analyzed cows (21 that did contribute milk to the tank and 19 that did not). Supplementary Figure 3 (available online at <http://www.journalofdairyscience.org/>) shows the distribution of minor allele frequencies in the actual data set for these 8,696 SNP. The distribution was fairly uniform, hence comparable to the simulated SNP sets.

Figure 2 shows the relation between measured SCC and SSC estimated from the adjusted B-allele proportions (f_j) in the pooled milk sample. The correlation was highly significant ($P < 0.0001$; $R^2 = 0.60$), yet lower than in equivalent simulated data sets. We then evaluated our ability to distinguish cows contributing milk to the tank from those that were not. The P -values associated with pc_i were $\leq 10^{-4}$ for all cows contributing milk, except the one with lowest SCC (10,000/mL; $P = 0.006$). However, when adding cows that did not contribute milk to the tank one-by-one in the linear model, the P -value of the corresponding regression coefficients (pc_i) was ≤ 0.0025 (i.e., a Bonferroni-corrected 5% threshold) for approximately one-third of the cows (Supplementary Figure 4, available online at <http://www.journalofdairyscience.org/>).

DISCUSSION

In this work, we present a novel approach for determining SCC for individual cows by genotyping a sample of milk from the milk tank (i.e., without having to perform a separate measurement for each animal). The method presupposes that individual SNP genotypes are available for each cow. This will increasingly correspond to reality as (1) genotyping costs continue to drop and (2) applications of genomic selection extend to cows.

The method proved effective using both simulated and actual data. Not unexpectedly, for comparable scenarios in terms of number of cows and SNP, the coefficient of determination was considerably better with the simulated than with the actual data. Several factors could contribute to this discrepancy, including (1) inaccuracies in counting somatic cells in actual samples; (2) absence of linkage disequilibrium between simulated SNP, leading to an overestimation of the informativeness of the simulated SNP sets; and (3) relatedness between actual but not simulated cows. For all of these reasons, application of the proposed method in the field would certainly benefit from maximizing the number of SNP used to compensate for propagation of inaccuracies from various sources.

The SNP genotypes of the individual cows could either be actual genotypes or—assuming that the SNP panel used on the tank milk is not identical to that used on the cows—imputed genotypes. Methods for genotype imputation are well established and particularly effective in livestock, as close relatives are often genotyped, allowing exploitation of Mendelian and within-family linkage information in addition to population-wide linkage disequilibrium information. Alternatively, it might be possible to impute allelic frequencies for the tank milk at missing SNP positions from the allelic frequencies measured at flanking markers and from the known linkage disequilibrium structure in the herd of interest. The effect of genotype (frequency) imputation on the precision of the proposed method will have to be evaluated.

In Belgium, SCC is typically evaluated approximately 10 times per year at a cost of approximately €25 per cow per year. In a farm with 100 cows, this amounts to approximately €2,500 per year. At the pres-

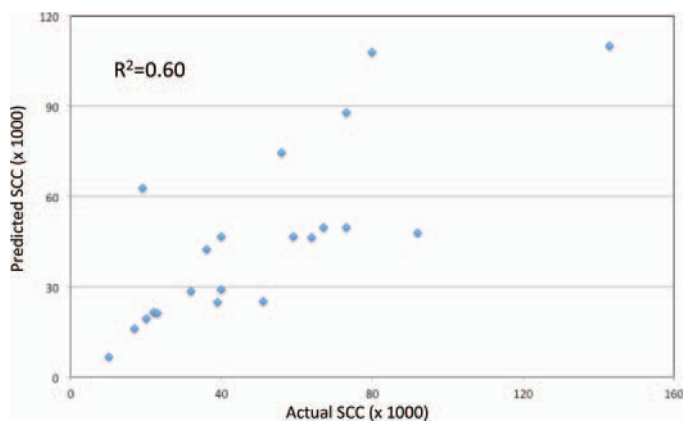


Figure 2. Actual data. Relationship between measured SCC and SCC estimated by SNP genotyping milk from a tank including milk from 21 cows. Color version available in the online PDF.

ent price, this corresponds to >50 analyses with a 50K SNP array, and >15 analyses with a 700K SNP array. Thus, the argument can be made that, even today, the proposed methodology is price competitive.

Analyzing DNA extracted from tank milk will become increasingly useful and cost effective as other applications are becoming available. Monitoring the milk's microbiome, including detection of putative pathogens, by means of high-throughput sequencing of, for instance, 16S rRNA amplicons is certainly one application that is virtually immediately practical.

We explored the possibility of determining whether specific cows did or did not contribute milk to the tank. This might, for instance, be useful to monitor the respect of exclusion of milk from treated cows to the tank. Related applications have previously been evaluated in the context of human genetics (Homer et al., 2008). In this study, we used the statistical significance of the cow-specific pc_i regression coefficients (measured using a standard t -test) as an indicator of the presence or absence of milk of a specific cow in the pool. Although specificity and sensitivity appeared satisfactory with the simulated data, this was not the case with actual data. The reasons underlying this discrepancy could be multiple and are presently being examined.

In summary, we herein describe the principles and demonstrate the feasibility of a novel approach to determine SCC of individual cows that has the potential to be a valuable addition to the arsenal of methods to control mastitis including subclinical.

ACKNOWLEDGMENTS

We are grateful to Erik Mullaart and Cattle Improvement Cooperative (CRV) for providing us with the milk samples and SNP genotypes used to demonstrate the

feasibility of the method on actual data. We are grateful to Frédéric Rollin, Christian Hanzen, and Leonard Théron (Department of Clinical Sciences of Livestock, Faculty of Veterinary Medicine, University of Liège) for fruitful discussions. This work was funded by the Unit of Animal Genomics (University of Liège, Liège, Belgium).

REFERENCES

- Charlier, C., W. Coppieters, F. Rollin, D. Desmecht, J. S. Agerholm, N. Cambisano, E. Carta, S. Dardano, M. Dive, C. Fasquelle, J.-C. Frennet, R. Hanset, X. Hubin, C. Jorgensen, L. Karim, M. Kent, K. Harvey, B. R. Pearce, P. Simon, N. Tama, H. Nie, S. Vandeputte, S. Lien, M. Longeri, M. Fredholm, R. J. Harvey, and M. Georges. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.* 40:449–454.
- Dohoo, L. R., and K. E. Leslie. 1991. Evaluation of changes in somatic cell counts as indicators of new intramammary infections. *Prev. Vet. Med.* 10:225–237.
- Homer, N., S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 4:e1000167.
- Hogeveen, H., K. Huijps, and T. J. G. M. Lam. 2011. Economic aspects of mastitis: New developments. *N. Z. Vet. J.* 59:16–23.
- Hovinen, M., and S. Pyörälä. 2011. Udder health of dairy cows in automatic milking. *J. Dairy Sci.* 94:547–562.
- Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350.
- Reneau, J. K. 1986. Effective use of dairy herd improvement somatic cell count in mastitis control. *J. Dairy Sci.* 69:1708–1720.
- Schukken, Y. H., and W. D. J. Kremer. 1996. Monitoring udder health: Objectives, materials and methods. Pages 351–360 in *Herd Health and Production Management in Dairy Practice*. Wageningen Press, Wageningen, the Netherlands.
- Tyler, J. W., M. C. Thurmond, and L. Lasslo. 1989. Relationship between test-day measurements of somatic cell count and milk production in California dairy herds. *Can. J. Vet. Res.* 53:182–187.