J. Dairy Sci. 96:7306–7317
http://dx.doi.org/10.3168/jds.2013-6818


# Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms

**H. A. Mulder,*†[1] R. E. Crump,† M. P. L. Calus,† and R. F. Veerkamp†**
*Animal Breeding and Genomics Centre, Wageningen University, PO Box 338, 6700 AH Wageningen, the Netherlands
†Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, the Netherlands

## ABSTRACT

In recent years, it has been shown that not only is the phenotype under genetic control, but also the environmental variance. Very little, however, is known about the genetic architecture of environmental variance. The main objective of this study was to unravel the genetic architecture of the mean and environmental variance of somatic cell score (SCS) by identifying genome-wide associations for mean and environmental variance of SCS in dairy cows and by quantifying the accuracy of genome-wide breeding values. Somatic cell score was used because previous research has shown that the environmental variance of SCS is partly under genetic control and reduction of the variance of SCS by selection is desirable. In this study, we used 37,590 single nucleotide polymorphism (SNP) genotypes and 46,353 test-day records of 1,642 cows at experimental research farms in 4 countries in Europe. We used a genomic relationship matrix in a double hierarchical generalized linear model to estimate genome-wide breeding values and genetic parameters. The estimated mean and environmental variance per cow was used in a Bayesian multi-locus model to identify SNP associated with either the mean or the environmental variance of SCS. Based on the obtained accuracy of genome-wide breeding values, 985 and 541 independent chromosome segments affecting the mean and environmental variance of SCS, respectively, were identified. Using a genomic relationship matrix increased the accuracy of breeding values relative to using a pedigree relationship matrix. In total, 43 SNP were significantly associated with either the mean (22) or the environmental variance of SCS (21). The SNP with the highest Bayes factor was on chromosome 9 (Hapmap31053-BTA-111664) explaining approximately 3% of the genetic variance of the environmental variance of SCS. Other significant SNP explained less than 1% of the genetic variance. It can be concluded that fewer genomic regions affect the environmental variance of SCS than the mean of SCS, but genes with large effects seem to be absent for both traits.

**Key words:** genome-wide association, environmental variance, genomic selection, somatic cell score

## INTRODUCTION

Mastitis is one of the most costly diseases in dairy cattle husbandry. Recording of mastitis is limited and the heritability of mastitis resistance is low. Therefore, direct genetic selection for mastitis resistance is difficult. Mastitis leads, however, to increased SCC, which is easy to measure during routine milk recording and has a higher heritability than mastitis. Therefore, SCS (log-transformed SCC) is often used as an index trait for mastitis, and in some national genetic evaluations, SCC peaks are used as predictor traits for mastitis (de Haas et al., 2008). Moreover, elevated SCC has a negative economic value, because milk factories give price penalties to farmers if bulk tank SCC is above a certain threshold (Veerkamp et al., 1998). Therefore, SCC is also a breeding goal trait in itself. Occurrence of mastitis leads to elevated SCC or peaks and, as such, also increases the lactation variance of SCS per cow (Green et al., 2004; de Haas et al., 2008; Urioste et al., 2012). Therefore, selection for reduced standard deviation or variance of SCS are alternatives that could be used to improve mastitis resistance. Urioste et al. (2012) showed that the heritability of standard deviation of SCS was 0.14. Another way to model the variability in SCS is to model genetic differences in environmental variance between cows using a double hierarchical generalized linear model (**DHGLM**), as shown by Rönnegård et al. (2013). Rönnegård et al. (2013) showed that the environmental variance of SCS is heritable. The amount of genetic variance in the size of the environmental variance is, however, small. In addition, Rönnegård et al. (2013) showed that the variance of SCS has a negative economic value in its own right because of price penalties for too-high SCC by milk factories.

Received March 15, 2013.
Accepted July 18, 2013.
[1]Corresponding author: Han.mulder@wur.nl

**Table 1.** Summary statistics of SCC (×1,000 cells/mL) and SCS [SCS = $\log_2$ (SCC)]

| Trait | Number of records | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|
| SCC | 46,353 | 133 | 411 | 1 | 14,678 |
| SCS | 46,353 | 5.73 | 1.65 | 0.00 | 13.84 |

Several studies have performed searches for QTL for mastitis or SCS (Rupp and Boichard, 2003; Khatkar et al., 2004). Wijga et al. (2012) found 2 SNP associated with lactation standard deviation of SCS (i.e., on chromosome 6 and 18). The one on chromosome 18 was also significant for the lactation average of SCS. So far, no studies have estimated genome-wide associations for measures of environmental variance for SCS or for any other traits in dairy cattle (e.g., obtained from double hierarchical generalized linear models; Rönnegård et al., 2010). Individual differences in environmental variance may be considered as differences in micro-environmental sensitivity, which might be partly under genetic control. In the case of SCS, higher variance in SCS can be caused by a higher incidence of mastitis (i.e., due to the bimodal distribution of SCS for healthy and mastitic cows). Studies that have searched for QTL of environmental variance investigated traits in *Arabidopsis thaliana* (Shen et al., 2012), pigs (Yang et al., 2011), and chickens (Rönnegård and Valdar, 2011; Wolc et al., 2012), but so far, such studies have not been performed in dairy cattle. In addition, accuracies of pedigree-based EBV or direct genomic values (**DGV**) for environmental variance of SCS are unknown.

The main objective of this study was to unravel the genetic architecture of mean and environmental variance of SCS by identifying genome-wide associations for mean and environmental variance of SCS in dairy cows and by quantifying the accuracy of DGV. In addition, genetic parameters for mean and environmental variance of SCS were estimated using both a genomic relationship matrix (**GRM**) and a numerator or pedigree relationship matrix (**NRM**). Note that the name environmental variance was used because we are interested in the genetic architecture of the true environmental variance and the residual variance of the models used in our study provide an estimate of that true environmental variance.

## MATERIALS AND METHODS

### Data

The data used in the present study originated from Teagasc Moorepark Dairy Production Research Center (Oak Park, Carlow, Ireland), Scottish Agricultural College (Edinburgh, UK), Wageningen UR Livestock Research (Lelystad, the Netherlands), and Swedish University of Agricultural Science (Uppsala, Sweden). On 1,642 first-lactation Holstein heifers, SCC and SCS [SCS = $\log_2$ (SCC)] data of 46,353 test-day records were available, as well as genotype data. These cows had between 1 and 45 test days, with a median of 30 test days. For estimation of genetic parameters, we used all of these cows. For estimation of genome-wide associations, we used records of cows with more than 5 test days, which resulted in 1,563 cows, and the pedigree contained 9,368 animals. Summary statistics of the phenotypic data used are in Table 1.

All used animals with phenotypic information were genotyped with the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) containing 54,001 SNP. Single nucleotide polymorphisms that fulfilled the following criteria were included in the analysis: (1) gene calling (GC) score >0.20 and genetic testing (GT) score >0.55, (2) call rate >95%, (3) minor allele frequency >0.01 in each country, and (4) no extreme deviation from Hardy-Weinberg equilibrium (i.e., $\chi^2$ <600; $P = 1.7 \times 10^{-132}$). After the quality control, 37,590 SNP remained. Checks for Mendelian inconsistencies between pedigree and SNP data were performed for all genotyped parent-offspring pairs and among sibs (Calus et al., 2011). Missing genotypes were imputed using BEAGLE software (Browning and Browning, 2007). Chromosome number and locations of the SNP on the BovineSNP50 were obtained from the UMD3.0 bovine genome assembly from the University of Maryland (College Park). The phenotypic and genotypic data as well as the edits were basically the same as used in van Binsbergen et al. (2012), except that the number of records and the number of animals was slightly different due to not all animals having SCC records.

### Prediction of (Genomic) Breeding Values and Genetic Parameters

A DHGLM was used to estimate breeding values and genetic parameters for mean and environmental variance of test-day SCS using either NRM or GRM. The DHGLM was described by Rönnegård et al. (2010) and the implementation in ASReml (Gilmour et al., 2006) was used here. A univariate model was used iteratively on both the mean and environmental variance of SCS. Rönnegård et al. (2010) used for the variance model a