J. Dairy Sci. 97:2481–2486
http://dx.doi.org/10.3168/jds.2013-7149

Short communication: Validation of somatic cell score–associated loci identified in a genome-wide association study in German Holstein cattle

Hamdy Abdel-Shafy,[1] Ralf H. Bortfeldt, Monika Reissmann, and Gudrun A. Brockmann[2]
Breeding Biology and Molecular Genetics, Department for Crop and Animal Sciences, Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany

ABSTRACT

Recently, we identified 6 genomic loci affecting daughter yield deviations (DYD) for somatic cell score (SCS) in a genome-wide association study (GWAS) performed with German Holstein bulls. In the current study, we tested if these loci were associated with SCS in cows using their own performance data. The study was performed with 1,412 German Holstein cows, of which 483 were daughters of 71 bulls that had been used in the GWAS. We tested 10 single nucleotide polymorphisms (SNP) representing 6 genomic regions that were associated with DYD for SCS in bulls. All tested SNP were significant in cows. Seven of them, located on *Bos taurus* autosomes (BTA) 6, 13, and 19, had the same direction of effect as those previously reported in the bull population. The most significant associations were detected on BTA6 and BTA19, accounting for 1.8% of the total genetic variance. The major allele of the 2 SNP on BTA6 and the minor allele of the 2 SNP on BTA19 were favorable for lower SCS. The differences between the homozygous genotype classes were up to 15,000 cells/mL. The verification of SNP associated with SCS in this study provides further evidence for the functional role of the linked genomic regions for immune response and contributes to identification of causative mutations. In particular, SNP with minor frequency of the favorable allele possess high potential to reduce SCS in German Holstein cattle by selection.
Key words: validation study, single nucleotide polymorphism, mastitis resistance, candidate gene

Short Communication

Beside management strategies, genetic improvement is an important tool to control mastitis in dairy farms. Because clinical mastitis and its indicator SCS have low heritabilities of about 0.10 and 0.16 in German Holstein cattle, respectively, traditional methods of selection have limited success (Martin et al., 2010; Hinrichs et al., 2011). Alternatively, selection based on genomic information is expected to be more accurate, to cost less, and to become sustainable practice (Meuwissen et al., 2001). However, because most SNP used for genomic selections are linked to unknown causative mutations, linkage disequilibrium (**LD**) must be adjusted from time to time because of recombination events. To circumvent re-evaluation of SNP effects and to understand the biological mechanism of gene variants, it is desirable to know the causative mutations. An essential step is the accurate genomic mapping of genes contributing to SCS and clinical mastitis.

Previously, we identified 16 SNP on 6 chromosomal regions that were significantly associated with daughter yield deviations (**DYD**) of SCS in a genome-wide association study (**GWAS**) in German Holstein bulls (H. Abdel-Shafy, R. H. Bortfeldt, J. Tetens, and G. A. Brockmann, unpublished data). The significant SNP were located on BTA 5, 6, 13, 18, and 19, and on the X chromosome. Among these associated loci, the chromosomal region on BTA6 between 85.5 and 88.1 Mb (position is based on UMD3.1 assembly; http://www.ensembl.org) was also significant in US Holstein cattle (Cole et al., 2011), which provides evidence for the importance of this locus in the regulation of SCS in dairy cattle.

Although GWAS are powerful to identify genomic loci with high precision, false-positive associations due to confounding effects are considered a limiting factor. The most important effect regarding this issue is population stratification (Goddard and Hayes, 2009). However, even after correction for population stratification, some false results might still remain. Therefore, validation studies are important to test whether the LD between a causative mutation and a SNP is the same between populations and is not affected by the population sub-structure (Chanock et al., 2007). For validation, animals with recently recorded data that were not used in the initial discovery study—for example, cows—can be used. In contrast to DYD of bulls, performance data

Received June 17, 2013.
Accepted December 16, 2013.
[1]Present address: Department of Animal Production, Faculty of Agriculture, Cairo University, 6 El-Gamma street, 12613 Giza, Egypt.
[2]Corresponding author: gudrun.brockmann@agrar.hu-berlin.de

of cows allows us to test additive and dominance effects (Lush, 1935). Therefore, the objectives of this study were to validate associations obtained from a GWAS for DYD of SCS in bulls (H. Abdel-Shafy, R. H. Bortfeldt, J. Tetens, and G. A. Brockmann, unpublished data) in a cow population and to test the mode of inheritance of associated loci.

The study was carried out with 1,412 German Holstein cows descending from 284 sires; 71 of these bulls were used in the initial GWAS, in which we detected significant SNP for DYD of SCS (H. Abdel-Shafy, R. H. Bortfeldt, J. Tetens, and G. A. Brockmann, unpublished data). The average number of cows per sire was 5 with a minimum of 1 (122 cases) to a maximum of 80 (1 case). The cows were born between 1996 and 2005 and calved between 1999 and 2010, with more than 87% calving after 2003. The age at first calving for these cows ranged from 620 to 1,250 d, with an average of 840 d. The cows were kept on 3 dairy farms in northern Germany under similar management conditions. Records from monthly test-days of SCC were provided by the center of national breeding evaluation (VIT: Vereinigte Informationssysteme Tierhaltung, Verden, Germany). To ensure a homogeneous data set, only the first 3 lactations were used. Furthermore, records between 5 and 305 DIM and lactations with a minimum of 3 test-days were accepted. Using these criteria, the data set included 1,412 cows with 36,813 records. Among these cows, 1,401, 1,370, and 1,156 animals finished the first, second, and third lactations, respectively. The SCC in the final data set was transformed using a logarithmic transformation function (Ali and Shook, 1980) to obtain a distribution (SCS) that is closer to normal: SCS $= \log_2(\text{SCC}/100) + 3$.

Genomic DNA was extracted from lymphocytes using the NucleoSpin Blood Quick Pure Kit (Macherey-Nagel, Düren, Germany). Cows were genotyped for 10 SNP in 6 regions on BTA5, 6, 13, 18, 19, and chromosome X (Table 1), which were associated with DYD for SCS in our previous GWAS with bulls (H. Abdel-Shafy, R. H. Bortfeldt, J. Tetens, and G. A. Brockmann, unpublished data). Genotyping was performed with allele-specific primers (Table 1) in PCR assays as described previously (Kreuzer et al., 2013); the genotyping rate was 98.6%.

To test the association between single SNP and SCS, a mixed model was applied using the MIXED procedure of SAS (version 9.3 SAS Institute Inc., Cary, NC). The model considered the lactation curve fit of Ali and Schaeffer (1987). Test-day records within lactation of a cow were treated as repeated measurements. The Schwarz Bayesian information criterion was used to select an appropriate covariance structure in the repeated statement (Schwarz, 1978). Because of unequal distances between test-days, the spatial power covariance structure was used as an appropriate fit for the data set. The final model was

$$\begin{aligned} Y_{ijklmnop} = \mu + G_i + AC_j + b_{k1}\,(DIM/c) \\ + b_{k2}\,(DIM/c)^2 + b_{k3}\log(c/DIM) + b_{k4}\,[\log(c/DIM)]^2 \\ + b_l\,(MY) + p_m + hys_n + s_o + \varepsilon_{ijklmnop}, \end{aligned}$$

where $Y_{ijklmnop}$ is the test-day record of SCS, μ is the overall mean of observations, G_i is the fixed effect of the SNP genotypes, AC_j is the fixed effect of age at calving, b_{k1} to b_{k4} are the regression coefficients associated with the fixed lactation function, where DIM is days in milk and c is a constant set to 305 (Ali and Schaeffer, 1987), and b_l is the regression coefficient associated with the fixed effect of milk yield (MY; Jamrozik and Schaeffer, 2010). Whereas the random environmental effect between consecutive test-day records within each lactation of a cow was included in the term of p_m, the environmental differences between consecutive lactations were considered as a random effect in the term of hys_n [hys is a combined effect of herd (h), year (y), and season (s) of calving, where seasons were defined as calendar quarters: January to March, April to June, July to September, and October to December]. We accounted for population stratification by fixing the sire s_o as a random effect, and $\varepsilon_{ijklmnop}$ is the residual error.

Single nucleotide polymorphisms were considered significant at a threshold of $\alpha \leq 0.05$ after Bonferroni correction if the nominal P-value $\times$ 10 ≤ 0.05, where 10 is the number of tested SNP. Post hoc, effects among genotype classes were tested for significance using a Tukey-Kramer test as implemented in SAS (SAS Institute Inc.). In addition, additive and dominance effects were tested. For estimation of the genetic variance explained by significant SNP, we used the formula $2\beta^2 f(1 - f)$, where β denotes the regression coefficient; that is, the effect of the locus per copy of the variant, and f denotes the frequency of the variant (Park et al., 2010). We used 5′ and 3′ flanking regions of 1 Mb around significant SNP to search for candidate genes that could be linked to the significant SNP. Gene search was performed via the Ensembl database (UMD3.1 Ensembl database build 73; http://www.ensembl.org).

The minor allele frequencies of all tested SNP ranged from 0.14 to 0.41 in the cow population, which were almost the same as in the bull population (ranging from 0.14 to 0.43; Table 2). All 10 tested SNP were significantly associated with SCS after Bonferroni correction ($P \leq 0.005$). These SNP were located on BTA5,