A Whole Genome Association Study to Detect Loci Associated with Somatic Cell Score in Dairy Cattle

S. Wijga[*], J.W.M. Bastiaansen[*], E. Wall[†], E. Strandberg[‡], Y. de Haas[§], D.P. Berry[**], and H. Bovenhuis[*]

Introduction

One of the primary characteristics of a robust dairy cow is its disease resistance. Mastitis is considered one of the most prevalent and costly production diseases in the current dairy industry (Seegers et al. (2003)). Mastitis resistance has become part of the breeding objective in many countries (Rupp and Boichard (2003)), for economical reasons as well as to improve animal welfare (Schulman et al. (2009)). However, only a few countries routinely record mastitis, making direct selection for resistance against mastitis difficult. Further, the heritability for mastitis is generally low. Somatic cell count (SCC), which is usually log-transformed to somatic cell score (SCS), is often used in selection as an indirect measure of mastitis (Rupp and Boichard (2003)). SCC is recorded routinely in most milk recording systems. Moreover, the heritability of SCC is higher than that of mastitis and moderate to high positive genetic correlations between SCC and the occurrence of mastitis exist. An unfavorable genetic correlation between mastitis resistance and production traits (Rupp and Boichard (2003)), has made SCC a difficult trait to genetically improve (Baes et al. (2009)), despite the large genetic variation that exists in SCC (Rupp and Boichard (2003)). Breeding for increased mastitis resistance could benefit from using genomic information.

The recent discovery of thousands of single nucleotide polymorphisms (SNPs) in livestock genomes, forming dense marker maps, and a concurrent strong reduction in genotyping costs has created new opportunities for the use of marker data (Daetwyler (2009)), allowing for genome-wide association studies (Hirschorn and Daley (2005)). The aim of the present study was to identify SNPs associated with SCS.

Material and methods

Animals and phenotypes. The present study used first lactation records on 1,933 cows from 4 European Holstein dairy cattle research populations. These populations were located in the Netherlands (n=590), Ireland (n=546), Scotland (n=653) and Sweden (n=144). Cows with lactation lengths of at least 150 days and that had at least 10 test-day records were included

[*]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, 6709PG, The Netherlands
[††]Sustainable Livestock Systems Group, Scottish Agricultural College, Bush Estate, Penicuik, Midlothian, EH26 0PH, United Kingdom
[‡]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden
[§]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, 6708WC, The Netherlands
[**]Teagasc, Moorepark Dairy Production Research Centre, Fermoy, Co. Cork, Ireland

in the analyses. For each cow the lactation-average SCC was calculated from her test-day records. Lactation-average SCC is generally used for genetic improvement of udder health, however it ignores variation in the pattern of SCC levels during lactation (de Haas, (2003)). The standard deviation of test-day SCC (SCC-SD) (Urioste et al. (2010)) better reflects differences in patterns. Lactation-average SCC was converted to SCS, where $SCS = -\log_2\left(SCC/10^5\right)+3$ (Rupp and Boichard (2003)). SCC-SD was log-converted the same way into SCS-SD.

The phenotypic information on the 4 dairy cattle populations was combined and adjusted for the fixed environment of country by herd-year-season of calving. Seasons were defined as calendar quarters. Adjacent seasons with fewer than five individuals were combined. Residuals from the model were retained to be used as phenotypes for the association analyses.

Genotypes and genotype quality assurance Cows were genotyped using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA). Quality control was performed on the genotype data, for which criteria set by Hayes et al. (2009) were used as a guideline. SNPs were included in the dataset if they met the following criteria: 1) minor allele frequency > 1% in each country and > 5% in the complete dataset; 2) the percentage of missing genotypes for a SNP across samples was < 5%; 3) GTscore and GCscore were > 0.55 and > 0.20, respectively; and 4) SNP did not show a strong deviation from Hardy Weinberg equilibrium (Hardy Weinberg χ^2 values < 600). GT- and GCscores are measures of genotyping quality at a SNP-across-animal and animal-by-SNP level, respectively. Furthermore, only animals with SNP call-rates > 95% were retained.

Statistical analyses. Data were analysed performing a basic one-degree of freedom allelic test of association with genotype on adjusted phenotypes. Calculations were performed using the software package PLINK (version 1.07, Purcell et al. (2007)). A false discovery rate (FDR) adjustment, set at a cut-off value of 0.05, was performed, which is available in the R package 'qvalue' (Storey and Tibshirani (2003)).

Results and discussion

The initial dataset consisted of 1,933 cows with first lactation records. However, due to selection criteria that were set, 1,525 and 1,523 animals were retained for the analyses of SCS and SCS-SD, respectively. The association analyses were performed with 35,373 and 35,374 SNPs for SCS and SCS-SD respectively. The $-log_{10}$ of the P-values obtained with the analyses were plotted against their chromosomal position (Figure 1).

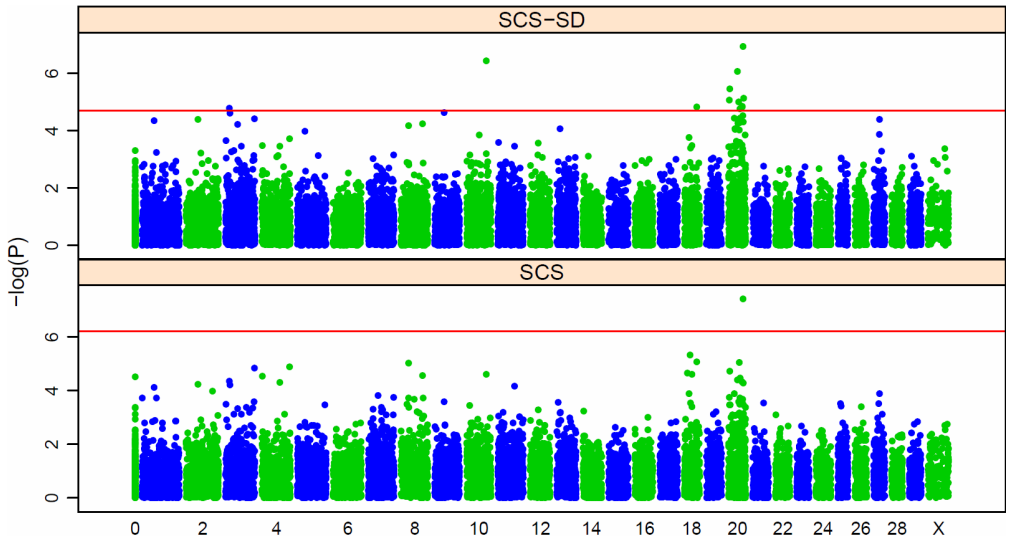

Figure 1: *-log$_{10}$* P-values from single SNP analyses. Chromosomes are arranged from left to right from chromosome 0 (unassigned SNPs) to chromosome X. SNPs above the red horizontal line passed the 0.05 False Discovery Rate threshold.

One SNP passed the 0.05 FDR threshold for SCS. This SNP was located on chromosome 20. Multiple studies have identified regions containing QTL underlying genetic variation for SCS on almost all bovine chromosomes (Khatkar et al. (2004)). SNPs associated with SCS on chromosome 20 have previously been reported by Ashwell et al. (2004).

For SCS-SD, significance levels of 12 SNPs passed the 0.05 FDR threshold. Nine of these SNPs were located on chromosome 20, 1 SNP was located on chromosome 18, 1 was located on chromosome 10 and 1 on chromosome 3. On chromosome 3, 10 as well as chromosome 18 QTL affecting SCS have been reported in previous studies (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index, Hu & Reecy (2007)). The SNP associated with SCS was also associated with SCS-SD.

More SNPs were found to be associated with SCS-SD than with SCS. Lactation-average SCS might not be able to differentiate between SCS from healthy and diseased animals (Madsen et al. (2008)). SCS-SD is more sensitive to individual test-day SCC values that are high in diseased animals. Results by Boettcher et al. (2007) suggest that SCS in healthy and diseased animals are different traits. Therefore, differences found in the present study might reflect differences in genetic background between SCS and SCS-SD, where SCS refers to the baseline SCC during lactation and SCS-SD, accounting for the variation in the curve, might be a reflection of the immune reactivity.

Conclusion

The dataset used in the present study contains phenotypic information from 4 dairy cattle populations from 4 different countries. A strong point of these data is that they are based on research populations. Due to frequent recording on these farms a large number of test-days are available for individual cows, with up to 52 test-days per cow. These frequent recordings increase the probability of detecting cases of mastitis. These data were used for a genome-wide association study, which detected significant associations with SCS and SCS-SD. 12 SNPs were found associated with SCS-SD. One of these SNPs was also found to be associated with SCS.

Acknowledgements

This work was carried out as part of the RobustMilk project that is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services.

References

Ashwell, M., Heyen, D., and Sonstegard, T. (2004). *J. Dairy Sci.*, 87:468-475.

Baes, C., Brand, B., and Mayer, M. (2009). *J. Dairy Sci.*, 92:4046-4054.

Boettcher, P., Caraviello, D., and Gianola, D. (2007). *J. Dairy Sci.*, 90:435-443.

Daetwyler, H., Schenkel, F., Sargolzaei, M. (2009). *J. Dairy Sci.*, 91:3225-3236.

de Haas, Y. (2004). *J. Dairy Sci.*, 87:95-105.

Hayes, B., Bowman, P., and Chamberlain, A. (2009). *J. Dairy Sci.*, 92:433-443.

Hirschhorn, J. and Daly, M. (2005). *Nat. Rev. Genet.*, 6:95-108.

Hu, Z. and Reecy, J. (2007). *Mamm. Genome.*, 18:1-4.

Khatkar, M., Thomson, P., Tammen, I. (2004). *Genet. Sel. Evol.*, 36:163-190.

Madsen, P., Shariati, M., and Ødegård, J. (2008). *J. Dairy Sci.*, 91:4355-4364.

Purcell, S., Neale, B., Todd-Brown, K. *et al.* (2007). *Am. J. Hum. Genet.*, 81:559-575.

Rupp, R. and Boichard, D. (2003). *Vet. Res.*, 34:671-688.

Schulman, N., Sahana, G., Iso-Touru, T. *et al.* (2009). *Anim. Genet.*, 40:509-515.

Seegers, H., Fourichon, C., and Beaudeau, F. (2003). *Vet. Res.*, 34:475-491.

Storey, J. and Tibshirani, R. (2003). In *Proc National Academy of Sciences*, volume 100, pages 9440–9445.

Urioste, J., Franzén, J., and Strandberg, E. (2010). In *Proc 9th WCGALP*, Leipzig. 1-6 August.