Genome Wide Association Study for Milk Production and Fat to Protein Ratio in Dairy Cattle

J.W.M. Bastiaansen[*], H. Bovenhuis[*], S. Wijga[*], S. Mc Parland[†], E. Wall[‡], E. Strandberg[§], and R.F. Veerkamp[¶]

Introduction

High-throughput genotyping has made it possible to move from linkage studies to whole genome association studies. This is expected to result in higher power to detect QTL with smaller effects as well as more accurate estimates of QTL location. So far a limited number of Genome Wide Association Studies (GWAS) have been published in dairy cattle. In general, cattle GWAS have been done with genotypes and estimated breeding values or daughter yield deviations on progeny tested bulls. GWAS based on EBVs of bulls are very powerful but they are limited to routinely recorded traits. More detailed phenotypic recording takes place on experimental farms, however, the number of records on one farm is usually small. The cost of genotyping a large number of samples continues to go down rapidly, and collection of accurate phenotypes are now the limiting factor for GWAS.

One of the aims of the RobustMilk project (http://www.robustmilk.eu) is to develop genomics tools to aid selection for robustness and milk quality. For this study, RobustMilk partners in 4 different countries combined high quality phenotypic data from their experimental herds, as well as high density genotypes, to allow a GWAS for dairy cow robustness. The RobustMilk dataset provides specific advantages because it consists of individual cow phenotypes with more frequent recording of milk production as well as traits not routinely collected like feed intake and energy balance. Here we report the first results from our GWAS for milk production (MILK) and the deviation of fat to protein ratio (FPR) in early lactation from the mean FPR across the lactation.

Material and methods

Animals and phenotypes. First lactation weekly milk test-day records were combined for 1,933 Holstein cows that were located in The Netherlands (n = 590), Scotland (n = 653),

[*] Animal Breeding and Genomics Centre, Wageningen University, Wageningen, 6709PG, The Netherlands
[†] Teagasc, Moorepark Dairy Production Research Centre, Fermoy, Co. Cork, Ireland
[‡] 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

Sweden (n = 144) and Ireland (n = 546). In total, 77,268 milk, 70,613 fat and 70,671 protein test-day records were available. The Wilmink exponential curve (Wilmink, 1987) was fitted to individual cow test-day records to calculate 305-day yields. In this study, 305 day milk kilograms (MILK), fat percentage (FAT%), and protein percentage (PROT%) were calculated on cows that had at least 10 test-day records with at least one record before day 75 and at least one record after day 150 of lactation. In addition, cows with 305-day yields of less than 3000 kg milk, 100 kg fat or 100 kg protein were discarded.

Energy Balance. FPR in the milk during early lactation was used as a measure of energy balance. FPR was calculated for all test-days within weeks 2 to 5 of lactation for cows that had at least 2 test-day records for FAT% and PROT% within this 4 week period. FPRdev was calculated for each test-day by subtracting the FPR over the complete lactation from the FPR on each test-day. The mean FPRdev and the maximum FPRdev during weeks 2 to 5 were calculated. The maximum FPRdev was used in subsequent association analyses.

Genotypes. Cows were genotyped with the Illumina BovineSNP50 BeadChip (illumina Inc., San Diego, CA) containing 54,001 SNP markers. Only cows with genotype call rates >95% were retained. Lower than 95% call rate indicates poor DNA quality which would affect accuracy of the called genotypes. Markers were included for further analyses if they fulfilled all of the following quality criteria : 1) GCscore > 0.20 and GTscore > 0.55; 2) genotype call rate > 0.95% within cows from country; 3) MAF > 0.01 in each country as well as > 0.05 in the complete dataset; and 4) no extreme deviation from Hardy Weinberg Equilibrium (χ^2 < 600). The GCscore and GTscore are quality measures on the genotype calls from the genotyping assay.

Adjusted data. Phenotypes from the 4 different sources were combined and regressed on systematic environmental effects using the following model:

$$Y_{ij} = CHYS_j + \beta_1 age_i + \beta_2 age_i^2 + e_{ij} \quad (1)$$

where $CHYS$ was a combination of country C and herd H in which the record was produced together with the year Y and season S of calving of the cow producing the record. Seasons were defined as a calendar quarters and adjacent calendar quarters were merged where necessary to obtain $CHYS$ class size $\geq$5. A linear age and quadratic age^2 effect were fitted. Non significant effects (P > 0.10) were removed.

Association Analysis. Single marker association analyses were performed with residuals from model 1. Association analyses were undertaken using the software package PLINK (v1.07) (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell *et al.*, 2007) with default settings. P values for association were obtained for all markers that passed quality control. A false discovery rate (FDR) adjustment was performed according to Storey and Tibshirani (2003) as implemented in the R package qvalue. An FDR cutoff value of 0.05 was applied.

Results and discussion

Cows experience a negative energy balance (NEB) during the first weeks of lactation when demands for milk production are high and the cow is trying to return to homeostasis. The NEB corresponds with increased levels of fatty acids in blood during weeks 2 to 5 of lactation

(van Knegsel *et al.*, 2007). During this period of NEB an associated increase in the ratio of fat to protein (FPR) has been observed in the milk, attributable mainly to the mobilization of body fat. The deviation (FPRdev) of FPR in early lactation from FPR over the complete lactation was considered to be an improved measure over FPR because FPRdev removes the natural variation in FPR between cows, variation that is not related to NEB in early lactation (de Vries and Veerkamp, 2000).

Across all cows 305 day average FPR was 1.19 and the mean and maximum FPR in weeks 2 to 5 of lactation were 1.30 and 1.45 respectively. From the 1,933 cows in the dataset, 1,823 passed the selection criteria for MILK and 1,728 passed the selection criteria for FPRdev. Genotyping was not successful for 2.8% of cows which meant 1,766 and 1,674 cows were included in association analyses for MILK and FPRdev respectively.

All model parameters $CHYS$, age and age^2 had an effect on 305 day MILK (P < 0.10). Both $CHYS$ and age had an effect on FPRdev (P < 0.10) but age^2 was not significant (P > 0.10) and was removed from the model. Association analyses were performed with 35,444 and 35,397 markers respectively for MILK and FPRdev. The $-log_{10}$ of P values were plotted for all markers against their genome position (Figure 1).

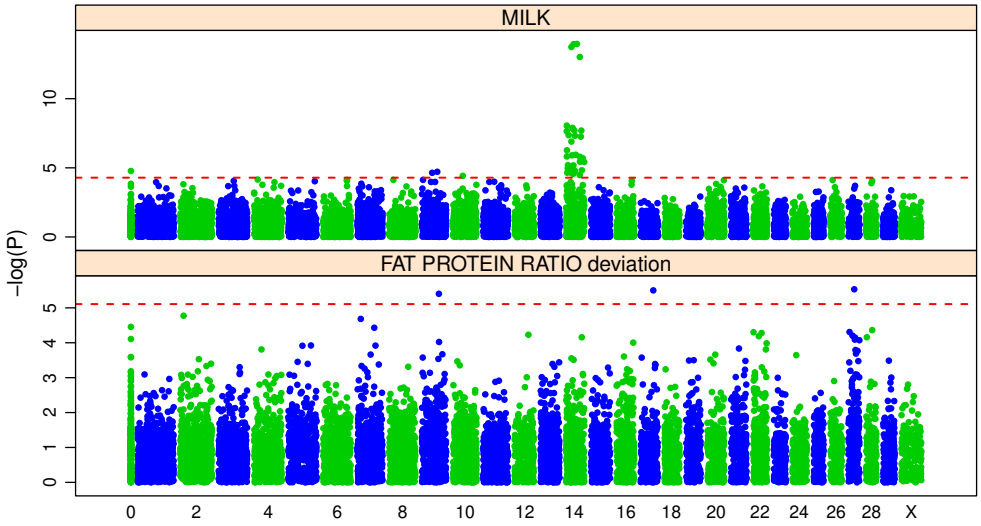

Figure 1: $-log_{10}$ of P values from single marker analyses with MILK (top) and FPRdev (bottom). Chromosomes are shown from chromosome 0 (unassigned markers) to chromosome X. Markers above the horizontal line survive a 0.05 False Discovery Rate.

For MILK, 36 markers passed the 0.05 FDR threshold of which 32 markers were located on chromosome 14, 2 on chromosome 9, 1 on chromosome 10 and 1 marker with FDR < 0.05 was not assigned to a chromosome. Milk yield QTL have been reported on all chromosomes of cattle (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) (Hu and Reecy, 2007). Highly significant results were also found for FAT% and PROT% with markers on chromosome 14

where a major gene for milk fat content was known to be located (results not shown). Markers with significant associations with MILK were found along the whole length of chromosome 14 while the known major gene is located at one extreme of the chromosome. The structure of the dataset may have caused LD to exist over longer distances than expected but this requires further investigation.

Three markers passed the 0.05 FDR threshold for FPRdev, one each on chromosomes 9, 17 and 27. Nkrumah *et al.* (2007) documented evidence of a QTL for feed intake on chromosome 17 in beef cattle. QTL for MILK and FAT% have been found on all chromosomes and may affect FPRdev because higher levels of MILK or FAT% are expected to decrease the cow's energy balance. The traits MILK and FPRdev showed a small positive correlation of 0.15 in this dataset.

Chromosome 9 has 2 significant markers for MILK at megabase (MB) positions 45.5 and 68.1 as well as one significant marker for FPRdev at MB position 75.4. Linkage disequilibrium, r^2, between these 3 markers ranged from 0.00 to 0.06 indicating that they detect separate associations.

More detailed analyses of the associated markers and their genomic context will be needed to assess how they are involved in energy balance during early lactation. Additional information on traits like feed intake and bodyweights for the cows in this study will be used.

Conclusion

Significant associations were detected for MILK and FPRdev in the RobustMilk dataset with a model that simply adjusts for fixed differences between means of contemporary groups and age at calving. Associations were found for MILK, as expected on chromosome 14, and also on chromosomes 9 and 10. Three markers, on chromosomes 9, 17 and 27, were associated with FPRdev which was used as a measure of NEB.

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

de Vries, M. J. and Veerkamp, R. F. (2000). *J. Dairy Sci.*, 83:62–69.

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

Nkrumah, J. D., Sherman, E. L., Li, C. *et al.* (2007). *J. Anim Sci.*, 85:3170–3181.

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

Storey, J. D. and Tibshirani, R. (2003). *Proc. Nat. Acad. Sci.*, 100:9440–9445.

van Knegsel, A., van den Brand, H., Dijkstra, J. *et al.* (2007). *J. Dairy Sci.*, 90:3397–3409.

Wilmink, J. (1987). *Livest. Prod. Sci.*, 17:1 – 17.