GWAS for genotype by lactation stage interaction in dairy cattle

H. Lu[1] & H. Bovenhuis[1]

[1]Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH, Wageningen, the Netherlands.

Summary

Most Genome Wide Association Studies (GWAS) identified quantitative traits locus (QTLs) based on their average effect over the lactation. However, it is known that additive genetic variance changes during lactation and genetic correlations among milk production traits during different lactation stages deviates from unity. We conducted a GWAS specifically for SNPs whose effects on milk protein content change during lactation, i.e. those that show significant SNP by lactation stage interaction. Significance thresholds for SNP by lactation stage interaction cannot be based on standard distributions but requires permutation of the data. We detected regions on BTA 10 and 14 with significant SNP effects as well as significant SNP by lactation stage effects. Regions on BTA 3 and 27 only showed significant SNP by lactation stage effects. Regions on BTA 6, 15, 16, 20 and 26 only showed significant effects in a traditional GWAS. These results show that a GWAS accounting for a SNP by lactation stage interaction effect can help elucidate the genetic background of milk protein content.

Keywords: GWAS, genotype by lactation stage interaction, permutation, milk protein content

Introduction

QTL detection benefits the traditional animal breeding programs that solely rely on phenotypes and pedigree information and increases the understanding of genetic background of complex economic traits in livestock. Some QTLs underlying milk production traits have been reported, for example, diacylglycerol O-acyltransferase 1 (*DGAT1*) (Grisart *et al.*, 2004), ATP binding cassette subfamily G member 2 (*ABCG2*) (Cohen-Zinder *et al.*, 2005) and growth hormone receptor (*GHR*) (Blott *et al.*, 2003). Using GWAS additional chromosomal regions affecting milk production traits were detected for example on BTA 6 where the casein gene cluster is located (Jiang *et al.*, 2010, Cole *et al.*, 2011, Schopen *et al.*, 2011). Most GWAS, studies are based on 305-day lactation records. Consequently, these studies do not account for possible changes of genetic effects throughout the lactation. However, QTL effects might change during the course of a lactation. When the QTL effect changes during lactation, longitudinal models might substantially increase QTL detection power (Lund *et al.*, 2008).

There is some evidence that QTL affecting milk production traits change during lactation. The additive genetic variance (Druet *et al.*, 2003) as well as the heritabilities are not stable throughout a lactation (Druet *et al.*, 2005, Bastin *et al.*, 2011). Moreover, genetic correlations of milk yield and milk protein yield between the beginning and later lactation stages differ from 1 (Jakobsen *et al.*, 2002). Expression of genes involved in milk protein synthesis also change during the course of lactation (Bionaz & Loor, 2011, Wickramasinghe *et al.*, 2012). It also has been shown that effects of the *DGAT1* polymorphism on milk production traits changes during lactation: a significant effect of *DGAT1* on the shape of the

lactation curve has been detected (Strucken *et al.*, 2011) and significant *DGAT1* by lactation stage interactions were detected for milk production traits (Bovenhuis *et al.*, 2015).

To our knowledge, there has not been a systematic screening of the genome for QTLs whose effects on milk production traits change throughout lactation. The objective of this study is to conduct genotype by lactation stage GWAS for milk protein content and compare results with those of traditional GWAS, i.e. assuming genotypic effects are constant throughout lactation. Such a study might result in the detection of novel chromosomal regions which cannot be detected based on traditional GWAS and may help to elucidate the genetic background of milk protein content.

Materials and Methods

Phenotypes and genotypes

For this study data on 1,829 first-parity cows were available. Details about the animals used are available in Schopen *et al.* (2009). Milk protein content was measured by infrared spectroscopy. For each cow on average 10.7 test-day records were available and the total number of test day records was 19,593.

DNA was isolated from the blood samples of all cows and genotyped by a 50k SNPs chip with the Infinium assay (Illumina). The SNP chip was designed by CRV (cooperative cattle improvement organization, Arnhem, the Netherlands) and obtained from Illumina (San Diego, CA). In total cows were genotyped for 50,856 SNPs. SNP were not included in the GWAS if observations of a genotype class contained less than 10 test day records at a specific lactation stage. After pruning, 30,348 SNPs were left for the GWAS.

Statistics model and significance

GWAS for milk protein content was performed using an animal model. First a GWAS was performed without a SNP by lactation stage interaction using the following model:

$$y_{jklmnop} = \mu + b_1 * afc_{jklmnop} + season_j + scode_k + lact_l + SNP_m + HYM_n + animal_o + pe_p + e_{jklmno}$$

where $y_{jklmnop}$ is milk protein content; μ is the overall mean; $afc_{jklmnop}$ is the covariate describing the effect of age at first calving; $season_j$ is the fixed effect of calving season; $scode_k$ is the fixed effect accounting for differences in genetic level between groups of proven bull daughters, young bull daughters, and other bull daughters; $lact_l$ was the fixed effect of the lactation stage (26 classes of 15 days each); SNP_m was the fixed effect of SNP genotype; HYM_n was the random effect of herd year month, assumed to be distributed as ; animal was the random additive genetic effect of the individual, assumed to be distributed as ; pe_p was the permanent environmental effects, assumed to be distrusted as ; and $e_{jklmnop}$ was the random residual, assumed to be distributed as . **A** is the additive genetic relationships matrix which was constructed based on 14,062 individuals and **I** is the identity matrix.

Subsequently, we specifically searched for SNPs who showed a SNP by lactation stage interaction. For this purpose the model was extended with a SNP by lactation stage interaction term $(SNP*lact)_{lm}$. All analyses were performed using in ASReml 4 (Gilmour *et al.*, 2006).

In the GWAS without accounting for SNP by lactation stage interaction, the genome-wide false discovery rate (FDR) was controlled based on the P values of SNP effects using the R package “qvalue”. A genome-wide FDR < 0.01 was considered significant. When the

interaction term was included in the model the false discovery rate could not be applied to test the significance of the interaction term. Instead we used permutation to estimate the significance threshold. In 100 permutations, we simultaneously assigned all 30,348 SNPs to another animal. After each permutation, we ran the GWAS and selected the lowest P value for the SNP by lactation stage interaction in order to form the empirical distribution under the null hypothesis. The 5% level of the empirical distribution was used as a threshold.

Results and discussions

The $-\log_{10}(p)$ values of all SNPs using a traditional GWAS, i.e. without a SNP by lactation stage interaction are shown in Figure 1A. Significant regions for milk protein content were detect on BTA 6, 14, 15, 16, 20 and 26. In addition, on BTA 4, 7, 8, 10, 11 one single SNP passed the significance threshold. The significant region on BTA 6 ranges from 56.71Mb to 97.25Mb. It contains the SNP rs43703016 which is 1 of the 2 SNPs that is causal for the protein variants A and B of κ-CN. This SNP is significant associated with κ-CN, β-LG, casein index and milk protein percentage (Schopen *et al.*, 2011). *ABCG2* are located on around 38 Mb and are reported associated with milk protein content (Cohen-Zinder *et al.*, 2005). On BTA 14, two significant regions were detected for milk protein content. The first region ranges from 0.08Mb to 11.96Mb and another region ranges from 49.13Mb to 49.33Mb. SNP ULGR_SNP_AJ318490_1c located in the first significant region is 1 of the 2 SNP responsible for the DGAT1 K232A polymorphism. This SNP is significant associated with α_{S1}-CN, α_{S2}-CN and milk protein percentage (Schopen *et al.*, 2011). A significant region on BTA 20 ranges from 31.18Mb to 36.79Mb and contains the *GHR* gene. This region has also been associated with milk protein percentage in other studies (Jiang *et al.*, 2010). Significant SNPs on BTA 10, 15 and 26 are in agreement with results reported in previous studies (Cole *et al.*, 2011, Schopen *et al.*, 2011).

The $-\log_{10}(p)$ values for SNP by lactation stage interaction are shown in Figure 1B. First we used the FDR to control false positives, but the test statistic for the interaction term is strongly inflated. Inflation of test statistics for interaction terms has been observed in other studies as well, e.g. GWAS for genotype by environment interaction in humans (Almli *et al.*, 2014, Marigorta & Gibson, 2014). Voorman *et al.* (2011) also noted that the FDR adjustment for multiple testing did not worked for interaction terms in a GWAS. One of the tactics to alleviate this problem is permutation which was applied in the current study. In the current study, the threshold was based on 100 permutations. The estimated 5% threshold was $-\log_{10}(p) = 15.74$. In Figure 1B, we took a conservative approach by taking the lowest P value out of the 100 permutations which is a $-\log_{10}(p) = 18.56$.

Significant regions whose effect changed during lactation stages were detected on BTA 3, 10, 14 and 27. Regions on BTA 10 and 14 were also significant in the traditional GWAS, i.e. without a SNP by lactation stage interaction (Figure 1A). The lead SNP for the interaction model on BTA 10 is rs41591350 located at 48.72Mb. The lead SNP for the model without interaction (rs43629218) is 3 MB away from rs41591350. SNP rs43629218 is in the intron region of myosin IE (*MYO1E*) and associated with milk protein percentage (Schopen *et al.*, 2011). The region on BTA 14 contains the *DGAT1* polymorphism. This region is in line with previous studies which showed that the effect of *DGAT1* on milk protein content changes during lactation (Strucken *et al.*, 2011, Bovenhuis *et al.*, 2015). Besides the significant regions on BTA 10 and 14, we also identified two new regions on BTA 3 and 27. These two regions were not significant in a traditional GWAS. The lead SNPs of these two significant regions were rs29011303 and rs109651365, and they were located on 93.22Mb of BTA 3 and 37.92Mb of BTA 27, respectively. These two regions do not contain obvious candidate genes.

In the genotype by lactation stage interaction GWAS, regions on BTA 6, 15, 16, 20 and 26 did not show significant effects. This indicates that the effects of these regions stay constant during lactation.

We conducted a GWAS to explore genes whose effect change through lactation and proposed a permutation strategy to estimate threshold in jointly GWAS exploring the SNP effect and SNP by lactation stages interaction effect together. These strategy resulted in the detection of new regions associated with milk protein content.

Reference

Almli, L.M., R. Duncan, H. Feng, D. Ghosh, E.B. Binder, B. Bradley, K.J. Ressler, K.N. Conneely & M.P. Epstein, 2014. Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. JAMA Psychiatry. 71(12):1392-1399.

Bastin, C., N. Gengler & H. Soyeurt, 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J Dairy Sci. 94(8):4152-4163.

Bionaz, M. and J.J. Loor, 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform Biol Insights. 5:83-98.

Blott, S., J.J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir & W. Coppieters, 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics. 163(1):253-266.

Bovenhuis, H., M.H. Visker, H.J. van Valenberg, A.J. Buitenhuis & J.A. van Arendonk, 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J Dairy Sci. 98(9):6572-6582.

Cohen-Zinder, M., E. Seroussi, D.M. Larkin, J.J. Loor, A. Everts-van der Wind, J.H. Lee, J.K. Drackley, M.R. Band, A.G. Hernandez, M. Shani, H.A. Lewin, J.I. Weller & M. Ron, 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15(7):936-944.

Cole, J.B., G.R. Wiggans, L. Ma, T.S. Sonstegard, T.J. Lawlor, Jr., B.A. Crooker, C.P. Van Tassell, J. Yang, S. Wang, L.K. Matukumalli & Y. Da, 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 12(408):408.

Druet, T., F. Jaffrezic, D. Boichard & V. Ducrocq, 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J Dairy Sci. 86(7):2480-2490.

Druet, T., F. Jaffrezic & V. Ducrocq, 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genet Sel Evol. 37(3):257-271.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham & R. Thompson, 2006. ASReml User Guide Release 1.0.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.J. Kim, A. Kvasz, M. Mni, P. Simon, J.M. Frere, W. Coppieters & M. Georges, 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc Natl Acad Sci U S A. 101(8):2398-2403.

Jakobsen, J.H., P. Madsen, J. Jensen, J. Pedersen, L.G. Christensen & D.A. Sorensen, 2002.

Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. Journal of Dairy Science. 85(6):1607-1616.
Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu & Q. Zhang, 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One. 5(10):e13661.
Lund, M.S., P. Sorensen, P. Madsen & F. Jaffrezic, 2008. Detection and modelling of time-dependent QTL in animal populations. Genet Sel Evol. 40(2):177-194.
Marigorta, U.M. and G. Gibson, 2014. A simulation study of gene-by-environment interactions in GWAS implies ample hidden effects. Front Genet. 5(225):225.
Schopen, G.C., J.M. Heck, H. Bovenhuis, M.H. Visker, H.J. van Valenberg & J.A. van Arendonk, 2009. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. J Dairy Sci. 92(3):1182-1191.
Schopen, G.C., M.H. Visker, P.D. Koks, E. Mullaart, J.A. van Arendonk & H. Bovenhuis, 2011. Whole-genome association study for milk protein composition in dairy cattle. J Dairy Sci. 94(6):3148-3158.
Strucken, E.M., D.J. de Koning, S.A. Rahmatalla & G.A. Brockmann, 2011. Lactation curve models for estimating gene effects over a timeline. J Dairy Sci. 94(1):442-449.
Voorman, A., T. Lumley, B. McKnight & K. Rice, 2011. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. PLoS One. 6(5):e19416.
Wickramasinghe, S., G. Rincon, A. Islas-Trejo & J.F. Medrano, 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics. 13:45.

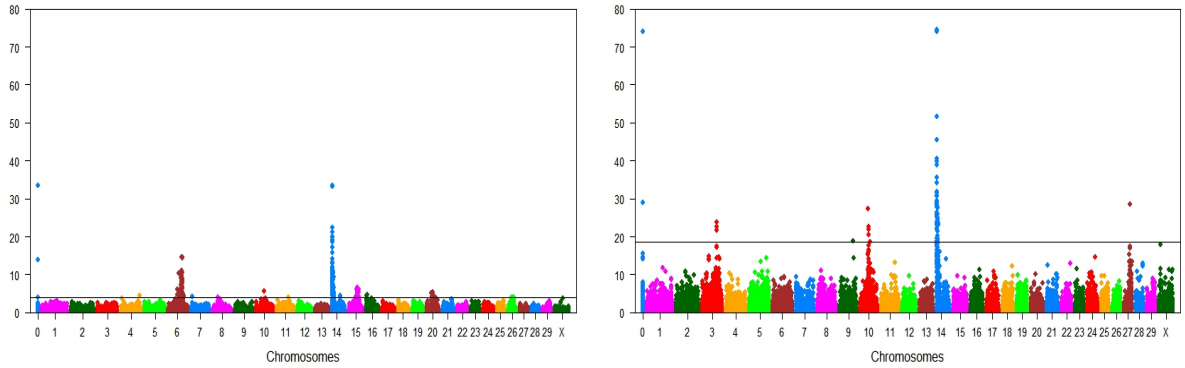

Figure 1, A:Manhattan plot of the SNP effects for milk protein content. The horizontal line indicates a FDR < 0.01 B: Manhattan plot for SNP by lactation stage interaction on milk protein content. The horizontal line indicates the genome wide significance threshold based on permutation.