GWAS for Robustness traits

H. Bovenhuis[1], D. Berry[2], A. Lunden[3], E. Wall[4], J. Bastiaansen[1], S. Wijga[1], M. Calus[5] and R. Veerkamp[5]

[1] *Animal Breeding and Genomics Centre, Wageningen University, The Netherlands*
[2]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Co. Cork, Ireland.*
[3]*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Sweden*
[4]*Sustainable Livestock Systems Group, Scottish Agricultural College, United Kingdom*
[5]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, The Netherlands*

Introduction

Genomic information is expected to make an important contribution to selection for traits which are difficult to improve by means of traditional selection. These might be traits which are difficult or expensive to measure and therefore routine recordings are not available, or traits which are recorded routinely but the phenotypes are not very accurate in a sense that the genetics is masked by disturbing environmental factors. However, before genomic information can make a contribution to genetic improvement of these traits either genes need to be identified that affect these traits or a reference population for setting up genomic prediction equations is needed (genomic selection). This requires animals that have both genotypes and phenotypes available. For several traits populations of sufficient size might not be readily available.

In the field of human genetics it has become increasingly common to form genome-wide association consortia (e.g. Bennett et al 2011). These consortia are an effective way to increase sample size and therefore to increase the statistical power to detect loci. In livestock science genome-wide association consortia are less common, however, combining data from several sources offers excellent opportunities for association studies and for the development of genomic selection tools.

Here we will present results of genome-wide association studies that have been conducted as part of the RobustMilk project, which is an initiative funded by the European Union in which data sets collected at experimental herds have been combined.

Materials and Methods

Genotypes

Animals from experimental farms at WUR Livestock Research (**LR**, The Netherlands), Teagasc Moorepark (**MPK**, Ireland), Swedish University of Agricultural Sciences (**SLU**, Sweden) and Scottish Agricultural College (**SAC**, Scotland (UK)) were available. Animals included in the genotyping were selected based on availability of sufficient amounts of good quality DNA and the availability of phenotypes. Further, only animals of the Holstein Friesian breed were included.

DNA was extracted from blood samples. Cows were subsequently genotyped for 54k SNP by a commercial genotyping company (ServiceXS, Leiden, the Netherlands) using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA). SNP genotypes were scored using Illumina BeadStudio software (v3.3.4). Quality control was performed on the genotypic data of the separate countries. SNP were included in the dataset when the following criteria were met: 1) the minor allele frequency (MAF) was > 1% in each country; 2) the percentage of missing genotypes for a SNP across all samples was <5%; 3) the Gen Train score (statistical score for accuracy of clustering) was > 0.55 and the Gen Call score (statistical score for genotyping accuracy) was > 0.20; and 4) the SNP did not deviate from Hardy Weinberg equilibrium (Hardy Weinberg $\chi2$ values < 600). A SNP that failed a criterion in at least one country was discarded from the GWAS. Furthermore, animals with SNP call-rates <95% were removed from the dataset (n=70). A total of 37,590 SNP were retained and used for analyses. Pedigree data and SNP information were compared to check for inconsistencies in pedigree of cows using the methodology outlined by Calus et al. (2011).

In total 696 animals from **LR**, 577 from **MPK**, 243 from **SLU** and 452 from **SAC** were genotyped as part of the RobustMilk project. In addition **SAC** made available genotypes of 318 animals that were genotyped as part of other projects.

Further, 58 bulls, which were sires of the cows in our data, were genotyped. **MPK** made available genotypes of an additional 154 bulls. In total

genotypes of 2286 cows and 212 bulls were used for analyses in the RobustMilk project.

Phenotypes

The present studies used phenotypic data of first lactation Holstein cows. Many phenotypes are available and were analysed. In the present paper we will describe the GWAS for a limited number of traits typical for the RobustMilk project. The three examples chosen are somatic cell count, start of luteal activity based on hormonal profiles and feed intake.

Somatic Cell Count. Wijga et al. (2011) analysed lactation-average SCC and test-day SCC standard deviation which were calculated for each cow based on test-day records. The standard deviation of test-day SCC aims to capture fluctuations in SCC associated with infection. The number of test-day records per cow on the experimental herds ranged between ten and 52 with an average of 31 test-days.

Fertility Berry et al. (2011) analysed fertility traits. The traditional fertility traits were days from calving to first observed heat, days from calving to first service, calving interval, number of services, and pregnancy rate to first service. Besides traditional fertility traits information on the interval from calving to first luteal activity (CLA) based on hormonal profiles was available. CLA was defined as the number of days from calving to the first occurrence of two consecutive test-day records with a milk progesterone concentration of ≥3 ng/ml.

Feed utilisation Veerkamp et al (2011) analyzed several traits related to feed utilization. Records were available for live weight, body condition score and dry matter intake. The averages for the predicted values for week 3 – 15 in lactation were used in the GWAS. Only animals with ten or more observations in this period were included in the analysis.

Results and discussion

We will briefly highlight the main results of the genome-wide association studies that were conducted. A detailed description of the results can be found in Wijga et al. (2011) on somatic cell count traits, Berry et al. (2011) on fertility traits and Veerkamp et al (2011) on feed utilization traits.

Somatic Cell Count Two SNP were found to be significantly associated with lactation-average SCC and one SNP with the standard deviation of SCC. One SNP on BTA18 was associated with both lactation-average SCC and the standard deviation of SCC. A SNP on BTA4 was uniquely associated with the lactation-average SCC. In total 98 cows in the data set had a case of clinical mastitis during their first lactation. The SNP with significant effects on the lactation-average SCC or the standard deviation of SCC were not significantly associated with clinical mastitis.

Relatively few significantly associated SNP were detected, suggesting that both somatic cell count traits are controlled by many loci, each with a relatively small effect. Different SNP contributing to lactation-average SCC and the standard deviation of SCC could reflect differences in genetic regulation of both traits where lactation-average SCC may refer to base-line somatic cell count during lactation and standard deviation of SCC may reflect immune reactivity to infection.

The number of test-day records per cow on the experimental herds ranged between ten and 52 with an average of 31 test-days. These numbers are much higher as compared to routinely collected data on commercial herds where SCC usually are recorded monthly. Monthly SCC recordings have lower probability of detecting infections of short duration than intensive recording as practiced on experimental herds.

Fertility For traditional fertility traits (days from calving to first observed heat, days from calving to first service, calving interval, number of services, and pregnancy rate to first service) only weakly significant SNP were detected. Much stronger evidence for the presence of QTL was detected for CLA: a SNP on BTA 2 explained 0.51% of the genetic variance in CLA while a SNP on BTA 21 explained 0.35% of the genetic variance in CLA.

Traditional fertility measures based on calving and insemination dates are heavily affected by decisions made by the farmer and therefore are not very accurate indicators for fertility. The interval from calving to first luteal activity (CLA) based on hormonal profiles has been suggested as a more accurate and therefore preferable measure for fertility. However, routine measures of CLA based on hormonal profiles are usually not available for reasons of cost and labour but were available on data collected on experimental herds. The GWAS results clearly

illustrate the benefits of having an accurately measured phenotype and the potential benefits of physiological traits whose genetic control may be less complex than high level phenotypes.

Feed utilisation The GWAS for feed utilization traits revealed several significant chromosomal regions. Highly significant effects were found on BTA4 for body condition scores, on BTA7, BTA13 and BTA18 for live weight and on BTA27 for dry matter intake. The percentage of the total genetic variance of most significant SNPs ranged from 0.5 to 0.6% indicating that the magnitudes of the effects were relatively small.

Costs of feed are a major component of the variable costs of milk production. Feed intake and feed efficiency are therefore traits which should be taken into account in breeding decisions. Recently, the importance of feed efficiency has received renewed interest in order to reduce the environmental impact of dairy production. However, in practical dairy cattle breeding there is no direct selection for feed intake or feed efficiency. This is a consequence of the practical difficulties of measuring individual feed intake in dairy cows. The detection of genes regulating feed intake or the development of genomic selection tools could bring selection for feed intake based on genomic information within reach.

Conclusion

The success of genome-wide association studies is a function of, amongst others, the heritability of the trait under investigation and the number of phenotypic records on that trait. For traits which are expensive and difficult to measure the number of available records is a limiting factor. The present study is one of the first studies to show the potential of combining detailed phenotypic and genotypic data from research herds located in different countries, thereby increasing the power of the study.

Acknowledgments

References

Calus, M. P. L., H. A. Mulder, and J. W. M. Bastiaansen. 2011 Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genetics Selection Evolution In press.

Berry, D.P., J.W.M. Bastiaansen, R.F. Veerkamp, S. Wijga, E. Wall, E. Strandberg and M.P.L. Calus. 2011 Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. Submitted to Animal

Bennett SN, Caporaso N, Fitzpatrick AL, Agrawal A, Barnes K, Boyd HA, Cornelis MC, Hansel NN, Heiss G, Heit JA, Kang JH, Kittner SJ, Kraft P, Lowe W, Marazita ML, Monroe KR, Pasquale LR, Ramos EM, van Dam RM, Udren J, Williams K; GENEVA Consortium. Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. 2011 Genet Epidemiol. 35:159-73.

Veerkamp, R.F., M. Coffey, D.P. Berry, Y. de Haas, E. Strandberg, H. Bovenhuis, M.P.L. Calus, and E. Wall. 2011. Genome-wide associations for feed utilisation complex in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. In preparation.

Wijga, S. J. W. M. Bastiaansen, E. Wall, E. Strandberg, Y. de Haas, L. Giblin and H. Bovenhuis. 2011 Genomic Associations with Somatic Cell Score in First Lactation Holstein Cows. Submitted to JDS