# SNP-Chip Analysis For Investigating Genetic Effects Over A Timeline

*E.M. Strucken*[*†], D.J. de Koning[†] and G.A. Brockmann[*]

## Introduction

High throughput methods have been recently developed to genotype individuals for 50,000 or more single nucleotide polymorphisms (SNPs) in a cost effective way. However, besides increasing the power of association studies by using dense marker maps the importance of the accurate selection of the phenotypic data should not be disregarded. The level of milk production differs between parities and shows a decrease for higher parities. Moreover, differences in milk yield can be observed between different lactations as well as within a single lactation. A lactation curve shows the peak production between day 35 and 50 after which production slowly decreases until the end of the lactation (Stanton et al. 1992, Dematawewa et al. 2007).
This study aims to locate genetic effects underlying variation in milk production over time. In order to estimate genetic effects between lactations the average production level for each cow was considered. To estimate genetic effects within a lactation, lactation curve parameters were estimated.

## Material and Methods

**Phenotypic Data.** 152 German Holstein Frisian cows, offspring of 90 sires, were selected according to their production level in milk yield. 78 cows were considered as high producing with an average of 31.53 kg milk per day and 72 cows were considered as low producing with an average of 21.51 kg milk per day. Cows were milked twice a day and average test-day records for lactation 1 to 8, sampled from June 2000 until June 2009 on a monthly basis, were provided by the VIT (Vereinigte Informationssysteme Tierhaltung w.V., Verden, Germany). Due to the heterogeneity of the number of records per cow and lactation restrictions were made and only lactations 1 to 4 with a minimum of 4 test-day records were taken into account. We ignored test-day records beyond day 340. Despite the selective genotyping, milk production was normally distributed.

**Genotype Data.** Using the BovineSNP50 BeadChip (Illumina), genotypes were produced for all 152 cows. The chip features 54,001 SNP with an average probe spacing of 51.5kb and an average minor allele frequency (MAF) of 0.22 across all loci for Holstein breeds (according to producer). SNPs are located on 30 chromosome pairs including the X-

[*] Institute for Breeding Biology and Molecular Animal Breeding, Faculty for Agriculture and Horticulture, Humboldt-Universität, 10099 Berlin, Germany

[†] Roslin Institute and R(D)SVS, University of Edinburgh, EH25 9PS Roslin, United Kingdom

chromosome and with a frequency of 0.06 on chromosome 1. 1,672 SNPs are not yet allocated to a chromosome.

**Statistical Analyses.**
Lactation curve parameters were estimated with the Wilmink curve (Wilmink 1978).

$$Y_{DIM} = a + b * 2.718^{-k * DIM} + c * DIM \qquad \text{model 1}$$

Where $Y_{DIM}$ is the test day record at a given days in milk (DIM), parameter *a* represents the basal level of production, parameter *b* the production increase towards the peak, and parameter *c* the production decrease after the peak. Scale parameter *k* was estimated on the whole data set and fixed for further analyses with *k*=0.12.
The estimated parameters were then used as phenotypes to conduct a GWA for each parameter (a, b, c, in model1) as well as the average production for each lactation individually or over lactation 2 to 4 and 1 to 4, respectively.

$$Y_{ijklm} = f_i + s_j + se_k + g_m + a_n + e_{ijklm} \qquad \text{model 2}$$

Where $Y_{ijklm}$ is either the average of lactation or parameter a, b, or c estimated by model 1. $f_i$ is the fixed effect of farm i; $s_j$ is the fixed effect of sire j; $se_k$ is the fixed effect of season k; $g_m$ is the fixed effect of group m in the stable; $a_n$ is the fixed effect of age n at first calving; and $e_{ijklm}$ is the random residual term.

## Results and Discussion

During quality control (QC), 7809 markers were excluded because of a low MAF (<1.64%). Furthermore, 1423 markers and one animal were excluded due to a low call rate (<95%). In total 44,962 SNPs and 151 animals passed the QC. Because animals are highly selected, no consideration was given to deviation from Hardy-Weinberg-Equilibrium in the QC.
Despite the small number of animals, the selection for high and low producing cows allowed detection of several highly significant results. The additive effects of the most significant loci on each chromosome are given in kg per day and written in brackets after the chromosome number. The average milk production over all lactation showed significant markers (P<0.001) on chromosome 2 (1.59), 6 (2.76), 11(1.28), and 18(1.42). Excluding lactation 1 from the average production performance showed significant markers on chromosome 4 (6.60), 5 (2.48), 12 (5.41), 13 (2.50), and 16 (2.24) suggesting a shift in genetic control between different lactations (Figure 1). Looking at the average production for each lactation separately showed significant markers on chromosome 6 (1.18) and 18 (1.08) for lactation 1, on chromosome 12 (5.07), 16 (2.25), and 13 (1.82) for lactation 2, on chromosome 26 (2.47), 9 (1.75), and 18 (1.59) for lactation 3 and on chromosome 1 (9.39) and 15 (for two regions 8.98 and 3.64) for lactation 4 (Figure 2a). 98 markers showed in more than one lactation *P* values <0.01 over all 4 lactations and 27 markers showed significant association over lactation 2 to 4. Increasing the threshold to P<0.001 showed only two markers with an effect across all lactations and 1 marker across lactation 2 to 4 with a significant association. From the comparison of the lactation it can be argued that only in multi-parity cows highly significant effects can be detected. Previous studies investigating quantitative trait loci

(QTLs) presented chromosome 3, 5, and 16 (Daetwyler et al. 2008) as well as chromosome 1, 3, 5, 8, 9, 11, 14, 18, 19, 21, and 23 associated with milk yield (Kolbendari et al. 2009). Several of our most significant markers are located close to already reported regions observed in more than one lactation. Nevertheless, the observations made in most of other studies were gained using estimated breeding values of bulls and thus, no separation between lactations was made.

Moreover, the genotypic contribution differs not only from lactation to lactation but also during a single lactation. Using the Wilmink model to estimate parameters influencing different stages of the production curve shows that different genome regions are associated with specific terms in the model. A GWA on the parameters from model 1 reveals that markers with significant effects ($P<0.01$; 57 markers over all lactations and 24 markers over lactation 2 to 4) in more than one lactation on the level of production (parameter *a*) are located on chromosome 9, 11, and 26. Markers with significant effects on the production increase towards the peak (parameter *b*) are found on chromosome 1, 9, and 16 (39 markers over all lactations and 7 markers over lactation 2 to 4) and markers with significant effects on the production decrease after the peak (parameter *c*) are located on chromosome 9, 1, and 5 (27 markers over all lactations and 21 markers over lactation 2 to 4; Figure 2).

Regarding the fact that the level of production increases with the number of parities it can be assumed that genomic regions affecting parameter *a* of model 1 also show larger effects in later lactations. This is illustrated by chromosomes 9 and 18, which harbour regions that are significantly associated with milk production in lactation 3 as well as with parameter *a*. No such correlation could be found between parameter *b* and *c* in any lactation number. Nevertheless, some of the genomic regions on chromosome 2, 5, and 9 that are associated with parameter *c* also show significant association with parameter *b*. Following this, it can be hypothesized that genomic regions that are influencing the incline towards the peak also influence the decline after the peak.

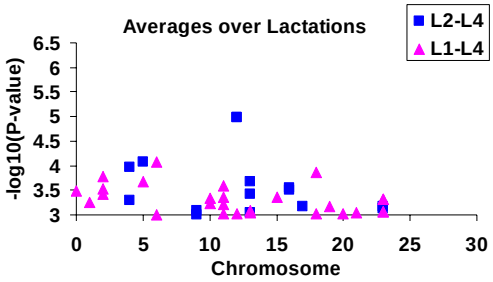


**Figure 1: Manhattan Plot for P-values <0.001 for the different lactation contributing to the average milk yield production; L2-L4= average performance per cow over lactation 2 to 4**

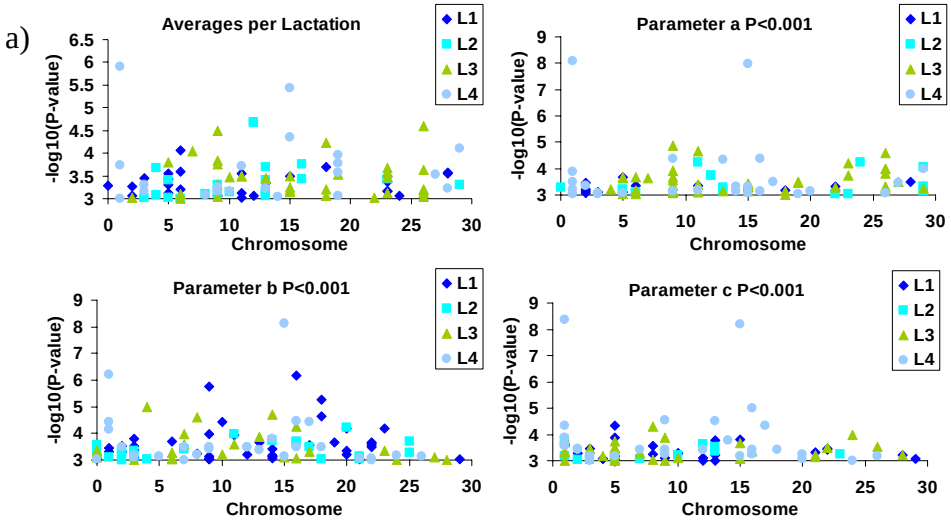


**Figure 2: Manhattan Plot for P-values <0.001 for averages per lactation and parameter a, b, and c estimated with model 1; L1= Lactation 1, L2= Lactation 2, L3=Lactation 3, L4=Lactation 4**

## Conclusion

Association analyses are not only a matter of the number of markers and individuals but also a matter of which phenotypic data is chosen. As this study shows, each lactation has its own genetic footprint. Different genomic regions show influence in different parities as well as on the different stages of lactation. We could show that parameter *a* influencing the peak production is affected by genomic regions that are also affecting the average production in later lactation. Not only the level of production but also the shape of the lactation curve differs between early and late parities. Whether the genomic regions affecting parameter *b* or *c*, and thus the shape of the production curve, also have different effects in different lactations has to be tested.

## References

Daetwyler, H.D., Schenkel, F.S., Sargolzaei, M. *et al.* (2008). *J. Dairy Sci,*. 91:3225-3236.

Dematawewa, C.M.B., Pearson, R.E., VanRaden, P.M. (2007). *J. Dairy Sci.*, 90:3924-3936.

Kolbendari, D., Wang, Z., Grant,J.R. *et al.* (2009). *J.Anim.Breed.Genet.*, 216-227.

Stanton, T.L., Jones, L.R., Everett, R.W. *et al.* (1992). *J. Dairy Sci.*, 75:1691-1700.

Wilmink, J.B.M. (1987). *Livest. Prod. Sci.* 16:335–348.