Genome Wide Association Studies Using Multiple-lactation Breeding Value in Holsteins

Kwang-Hyun Cho[a], Jae-Don Oh[1,a], Hee-bal Kim[2], Kyung-Do Park[1], and Joon-Ho Lee[1,*]

National Institute of Animal Science, RDA, Cheonan 330-801, Korea

ABSTRACT: A genome wide association study was conducted using estimated breeding value (EBV) for milk production traits from 1st to 4th lactation. Significant single nucleotide polymorphism (SNP) markers were selected for each trait and the differences were compared by lactation. DNA samples were taken from 456 animals with EBV which are Holstein proven bulls whose semen is being sold or the daughters of old proven bulls whose semen is no longer being sold in Korea. High density genome wide SNP genotype was investigated and the significance of markers associated with traits was tested using the breeding value estimated by a multiple lactation model as a dependent variant. As the result of significance comparisons by lactations, several differences were found between the first lactation and subsequent lactations (from second to 4th lactation). A similar trend was noted in mean deviation and correlation of the estimated effects by lactation. Since there was a difference in the genes associated with EBV for each trait between first and subsequent lactations, a multi-lactation model in which lactation is considered as a different trait is genetically useful. Also, significant markers in all lactations and common markers for different traits were detected, which can be used as markers for quantitative trait loci exploration and marker assisted selection in milk production traits. (**Key Words:** Genome Wide Association, Single Nucleotide Polymorphism, Milk Production Traits, Lactation, Holstein)

INTRODUCTION

Milk production ability is one of the most important traits in dairy cattle. Milk, milk fat, and milk protein yields are quantitative traits that are not affected by one chromosome or a small number of loci. Therefore, it is desirable to detect genetic variation that is effective for the expression of traits using high density genome-wide significant single nucleotide polymorphism (SNP) rather than to study minor genetic variation by candidate gene approach (Liu and Dekkers, 1998). Recently, with the advent of genome-wide panels of SNPs chips it became easier to explore quantitative trait loci (QTL) and SNPs associated QTL. Several genome wide association studies have been conducted using SNP chips in dairy cattle (Pryce et al., 2010; Jiang et al., 2010; Mai et al., 2010; Guo et al., 2012).

Because of the variations in milk yields by lactation, the genetic performance of dairy cattle is tested using a multiple lactation model utilizing different lactation records which are considered as different traits (Jamrozik et al., 1997). The first lactation in cows is affected by growth performance and nutrient distribution mechanism, and consequently genes associated with milk yield vary by lactation. The multiple lactation model can be indirect evidence since SNP association analysis varies at different lactations. Therefore, significant SNPs common to all lactations can be pure SNPs associated with milk production which are not affected by outside environment.

In this experiment a genome wide association study was conducted using estimated breeding value (EBV) for milk production traits from 1st to 4th lactation; significant SNPs

* Corresponding Author: J. H. Lee. Tel: +82-31-670-5656, Fax: +82-31-670-5657, E-mail: zoonolee@gmail.com
[1] Animal Genomics and Breeding Center, Hankyong National University, Anseong 456-749, Korea.
[2] Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea.
[a] These authors contributed equally to this work.

were selected for each trait and compared the differences by lactation.

MATERIALS AND METHODS

DNA sampling and data collection

To investigate high density genome-wide genotype, DNA samples were taken from 456 animals which are Holstein proven bulls whose semens are being sold or the daughters of old proven bulls whose semens are not being sold in Korea. The EBVs of all genotyped animals for milk, fat, and protein yields were also collected by lactation. Phenotypes from 1st to 4th lactation were considered as different traits and genetic (co)variances between lactations were considered to model an equation for breeding value estimation.

Estimated breeding value as the dependent variable of association test

The EBVs of all genotyped animals by lactation which were used as the dependent variable of marker association test were estimated using single trait-multiple lactation model.

$$y_{ijkl} = hy_{il}+ag_{jl}+a_{kl}+e_{ijkl}$$

Where,
y is 305-day adjusted milk, milk fat, milk protein
hy_i is ith herd-year effect
ag_j is jth age group effect of delivery age
a_k is animal genetic effect as a random effect
e_{ijk} is residual error.
l is lactation number (1 to 4)
Breeding value of 464,216 individuals were estimated and selected for 456 genotyped individuals.

Genotyping and quality control

Genomic DNA was extracted from samples and chip analysis was performed using Illumine Bovine SNP50 v2. Using Genome Studio program, genotypes of 54,609 SNPs were investigated. Then SNPs with more than 10% missing genotype rate, with less than 1% of minor allele frequency and with Hardy-Weinberg Disequilibrium (p<0.000001), SNPs on sex chromosome, and SNPs without position information were deleted and were not used in this experiment. The remaining missing genotypes were imputed using BEAGLE program (Browning and Browning, 2007), and imputation results were tested again with above standards. As the results of quality control of SNP information, 41,050 SNPs were used for the analysis.

Statistical analysis

Genome wide association test was performed using single marker regression.

$$y = 1\mu+bx+e$$

Where,
y is the EBV for milk yields, milk fat and protein yields of animals with all genetic information,
μ is mean effect of a SNP
b is the regression coefficient of EBV on SNP genotype,
x is allele code substituted minor homozygote, heterozygote and major homozygote with 0, 1 and 2, respectively,
e is residual error.

Significance for SNP association with traits was tested using F-test and significance level was corrected using Bonferroni correction. For all SNPs correlation coefficients for each estimated effect by lactation were calculated, and mean deviation and standard deviation by lactation were calculated and compared using the standardized SNP effects.

RESULTS AND DISCUSSION

Significant single nucleotide polymorphisms for milk production traits

Results of association analysis by trait using F-test are shown in Figure 1 (Figure 1 is for milk yield and Supplementary Figure S1, S2 are for fat yield and protein yield, respectively). It was observed that many SNPs had genome-wide effects and SNP frequency decreased sharply as significance level increased (Table 1). At the same significance level (1.22×10^{-6} same as Bonferroni corrected p-value<0.05) more SNPs were detected in the order of milk protein yields, milk fat yields, and milk yields. There were similar number of significant SNPs in milk yields and milk protein yields by lactation, however, more SNPs were detected in first lactation than in second to fourth lactation for milk fat yields. For milk yields, 10 significant SNPs located on chromosomes 10, 17, 21, and 24 were detected at least in one lactation. Thirteen SNPs located on chromosomes 2, 16, 19, and 21 for milk fat while milk protein had 28 SNPs located on chromosomes 1, 2, 3, 6, 8, 9, 12, 14, 17, 21, 24, and 28 which showed genome-wide distribution.

Some of the significant SNPs in one trait were also detected in different traits. Three SNPs BTB-01440888, ARS-BFGL-NGS-22135, and ARS-BFGL-NGS-101670 were significantly associated in both milk yield and milk protein yield trait (Table 2 and Supplementary Table S2). And another three SNPs, BTB-01536920, ARS-BFGL-NGS-35056 and ARS-BFGL-NGS-21956, were significantly associated in both milk fat yield and milk protein yield trait (Supplementary Tables S1 and S2). For milk protein yields, significant SNPs common to all other
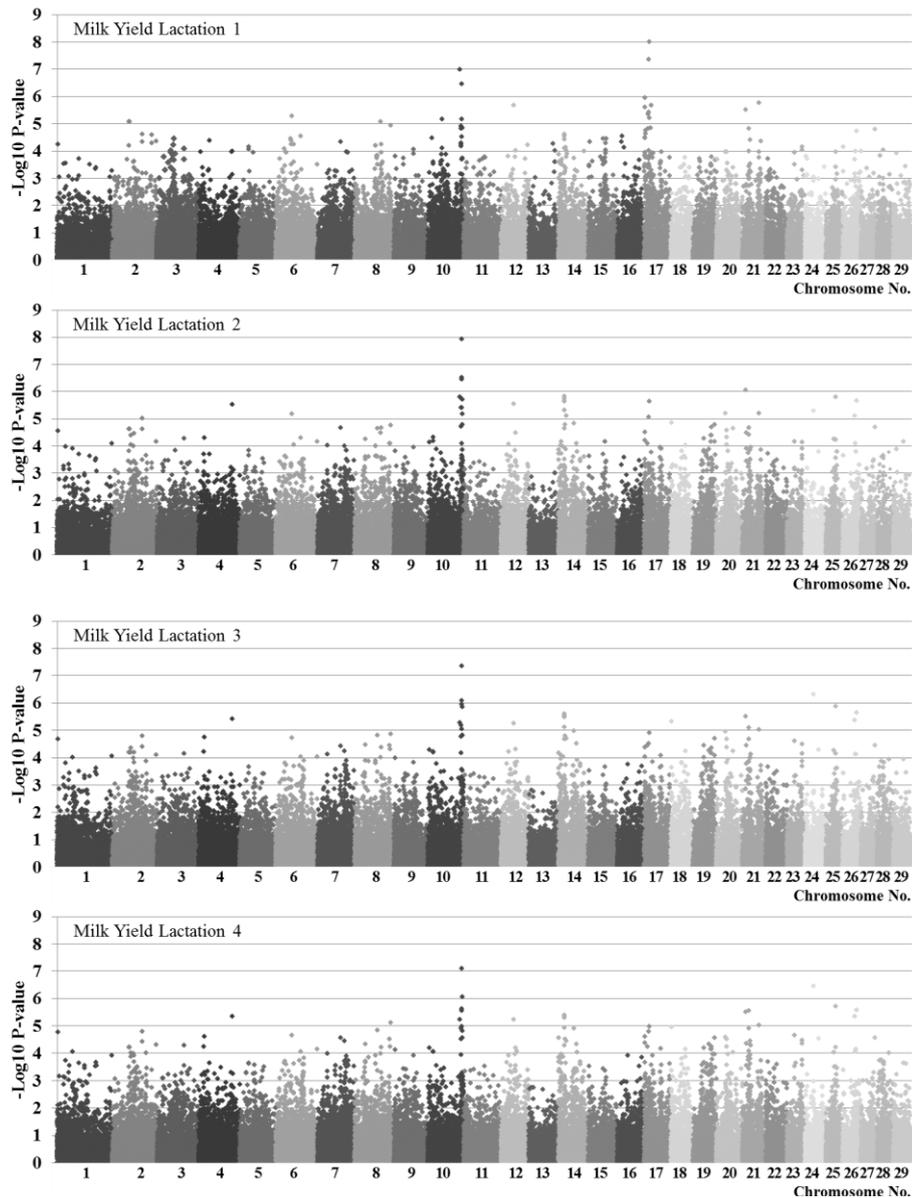
Figure 1. Manhattan plots of association significance for milk yields by lactation.

Table 1. Number of SNPs by −Log$_{10}$ p-value range for milk production traits

−Log$_{10}$ p-value <	Milk yield				Fat yield				Protein yield			
	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4
1	32,640	32,542	32,557	32,592	33,104	33,047	32,973	33,021	31,465	31,354	31,314	31,220
2	6,557	6,611	6,602	6,569	6,183	6,292	6,323	6,281	6,991	7,074	7,107	7,177
3	1,435	1,455	1,433	1,449	1,359	1,309	1,354	1,360	1,967	1,966	1,967	1,963
4	329	348	361	347	297	299	308	301	462	488	479	504
5	70	67	76	75	68	80	70	67	117	123	138	137
6	15	23	18	15	29	19	19	18	32	31	31	32
7	2	3	2	2	7	4	3	2	15	11	11	13
8	1	1	1	1	3	0	0	0	1	3	2	2
9	1	0	0	0	0	0	0	0	0	0	1	2
Significant SNPs	5	4	4	3	10	5	4	4	18	17	16	17

L1 to L4, lactation 1 to 4.
Significance is p-value <1.22E-06 (0.05/41050, Bonferroni correction).

Table 2. Genome-wide significant SNPs with milk yields

SNP names	Chromosome	Position	p-value for milk yields			
			Lactation 1	Lactation 2	Lactation 3	Lactation 4
BTB-00445660	10	93897741	1.00E-07*	1.56E-06	5.16E-06	5.83E-06
BTB-01646116	10	99103087	1.34E-05	3.00E-07*	8.10E-07*	2.41E-06
ARS-BFGL-NGS-117433	10	99132260	3.07E-05	3.60E-07*	1.13E-06*	2.76E-06
BTB-01901596	10	99234390	3.40E-07*	1.00E-08*	5.00E-08*	8.00E-08*
ARS-BFGL-NGS-34509	10	101474077	1.41E-05	1.89E-06	1.37E-06	8.60E-07*
BTA-90683-no-rs	17	2728598	1.12E-06*	6.02E-05	9.35E-05	4.05E-05
BTB-01440888[1]	17	11049283	4.00E-08*	8.44E-06	3.08E-05	1.64E-05
ARS-BFGL-NGS-22135[1]	17	13800376	1.00E-08*	2.25E-06	1.26E-05	1.06E-05
BTA-12959-no-rs	21	12095049	3.07E-06	8.80E-07*	3.13E-06	3.07E-06
ARS-BFGL-NGS-101670[1]	24	24873093	8.85E-04	5.12E-06	4.80E-07*	3.50E-07*

* Significant SNPs (p-value <1.22E-06).
[1] SNPs are significant for different traits as well (Protein yield).

traits were detected, while there were no significant SNPs common to milk yields and milk fat yields.

Some of the researchers reported highly significant SNPs on Chromosome 14 which differs from our results (Jiang, 2010; Pryce et al., 2010; Meredith et al., 2012; Minozzi et al., 2013), while other researchers reported that there was no significant SNPs on Chromosome 14 for other breeds (Mai, 2010; Guo et al., 2012). Also there is the tendency to find significant SNPs from a causal variant in Chromosome 14 (Hayes et al., 2010).

Difference of significant single nucleotide polymorphisms by lactation

There were significant SNPs that may have effects on all lactation. That is, for milk yields, BTB-01901596 in BTA10, for milk fat ARS-BFGL-NGS-95316 in BTA2, and for milk protein ARS-BFGL-NGS-53141 and ARS-BFGL-NGS-35056 in BTA2, ARS-BFGL-NGS-103603, ARS-BFGL-BAC-33343, ARS-BFGL-NGS-11578, ARS-BFGL-NGS-56762 and BTA-51937-no-rs in BTA21, and BTA-20879-no-rs in BTA28. However, most of the significant SNPs had different effects in first and subsequent lactation

(Table 2, Supplementary Tables S1 and S2). As their effect varies especially between first lactation and posterior lactation, it indicates the different genetic mechanism for milk production yields in first lactation and posterior different lactation. The same trends were noted in mean deviation of standardized effects and correlation (Table 3). The difference of mean deviation between first and subsequent lactation was the largest and correlation was relatively low. This result shows similar pattern to the result that genetic correlation is higher between second and third lactation than between first and second lactation using multiple lactation model (Liu et al., 2002). For standard deviation of estimated SNP effects by lactation, since the SNPs with large values of standard deviation by lactation showed genome-wide distribution (Figure 2), many genes differently express effects by lactation. And these may be candidate genes associated with genes affecting the deviation of milk production yields by lactation.

CONCLUSION

In conclusion, since the genes associated with each

Table 3. Mean deviation and correlation of the estimated marker effect among lactations

Trait		Lactation 1	Lactation 2	Lactation 3	Lactation 4
Milk yield	Lactation 1		0.307	0.370	0.361
	Lactation 2	0.918		0.092	0.125
	Lactation 3	0.882	0.993		0.060
	Lactation 4	0.888	0.987	0.997	
Fat yield	Lactation 1		0.284	0.364	0.415
	Lactation 2	0.930		0.106	0.156
	Lactation 3	0.886	0.991		0.057
	Lactation 4	0.852	0.979	0.997	
Protein yield	Lactation 1		0.301	0.375	0.348
	Lactation 2	0.922		0.107	0.106
	Lactation 3	0.879	0.990		0.040
	Lactation 4	0.896	0.990	0.999	

Upper diagonal, mean deviation; Lower diagonal, correlation coefficient.
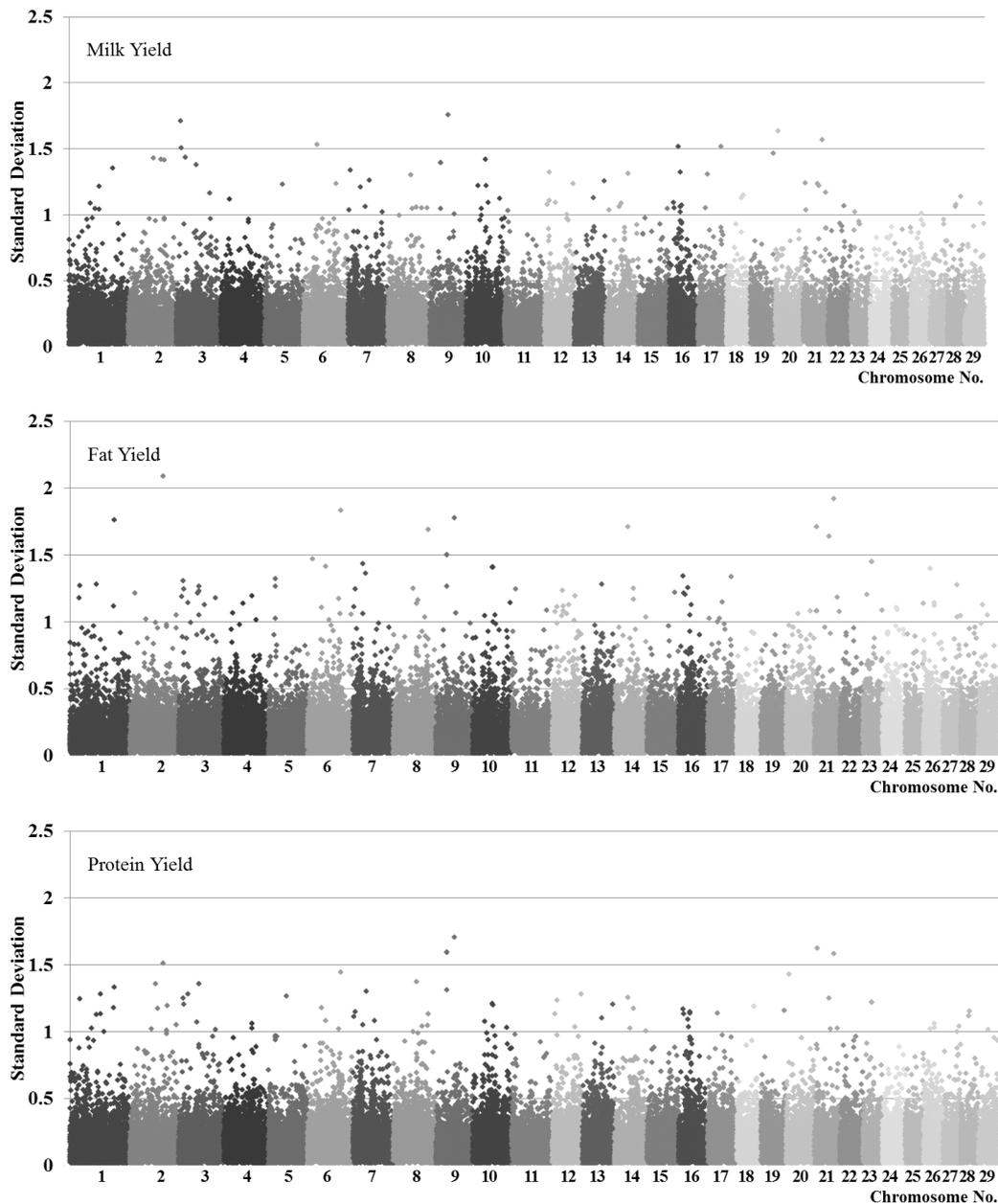
Figure 2. Manhattan plots for standard deviation of estimated single nucleotide polymorphism effects by lactation.

breeding value had different effects at different lactation period, a multi-lactation model in which lactation is considered as different traits can be genetically ideal for the analysis of lactation traits. Also, significant genetic markers for all lactations and markers which are common to all milk production traits at different traits were detected. These can be utilized for QTL exploration and marker assisted selection in milk production traits.

ACKNOWLEDGMENTS

REFERENCES

Browning, S. R. and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084-1097.

Guo, J., H. Jorjani, and Ö. Carlborg. 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genetics 13:82.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and M. E. Goddard. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat

percentage, and type in holstein cattle as contrasting model traits. PloS Genet. 6(9):e1001139.

Jamrozik, J., L. R. Schaeffer, Z. Liu, and G. Jansen. 1997. Multiple trait random regression test day model for production traits. Interbull Bulletin 16:43-47

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS ONE 5(10):e13661.

Liu, Z. and J. C. M. Dekkers. 1998. Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics 148:495-505.

Liu, Z., F. Reinhardt, and R. Reents. 2002. Genetics correlation estimates of a multiple lactation multiple country model for milk production traits based on performance records. Interbull Bulletin 29:12-17.

Mai, M. D., G. Sahana, F. B. Christiansen, and B. Guldbrandtsen. 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J. Anim. Sci. 88:3522-3528

Meredith, B. K., F. J. Kearney, E. K. Finlay, D. G. Bradley, A. G. Fahey, D. P. Berry, and D. J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genetics 13:21.

Minozzi1, G., E. L. Nicolazzi1, A. Stella, B. Biffani, R. Negrini, B. Lazzari, P. Ajmone-Marsan, and J. L. Williams. 2013. Genome wide analysis of fertility and production traits in Italian Holstein cattle. PLoS ONE 8(11):e80219.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331-3345.