Asian Australas. J. Anim. Sci.
Vol. 26, No. 6 : 766-771 June 2013
http://dx.doi.org/10.5713/ajas.2012.12715

www.ajas.info
pISSN 1011-2367 eISSN 1976-5517

A SNP Harvester Analysis to Better Detect SNPs of CCDC158 Gene That Are Associated with Carcass Quality Traits in Hanwoo

Jea-Young Lee[1], Jong-Hyeong Lee[1], Jung-Sou Yeo and Jong-Joo Kim*
School of Biotechnology, Yeungnam University, Gyeongsan, 712-749, Korea

ABSTRACT: The purpose of this study was to investigate interaction effects of genes using a Harvester method. A sample of Korean cattle, Hanwoo (n = 476) was chosen from the National Livestock Research Institute of Korea that were sired by 50 Korean proven bulls. The steers were born between the spring of 1998 and the autumn of 2002 and reared under a progeny-testing program at the Daekwanryeong and Namwon branches of NLRI. The steers were slaughtered at approximately 24 months of age and carcass quality traits were measured. A SNP Harvester method was applied with a support vector machine (SVM) to detect significant SNPs in the CCDC158 gene and interaction effects between the SNPs that were associated with average daily gains, cold carcass weight, *longissimus dorsi* muscle area, and marbling scores. The statistical significance of the major SNP combinations was evaluated with x^2-statistics. The genotype combinations of three SNPs, g.34425+102 A>T(AA), g.4102636T>G(GT), and g.11614+19G>T(GG) had a greater effect than the rest of SNP combinations, e.g. 0.82 vs. 0.75 kg, 343 vs. 314 kg, 80.4 vs 74.7 cm^2, and 7.35 vs. 5.01, for the four respective traits ($p<0.001$). Also, the estimates were greater compared with single SNPs analyzed (the greatest estimates were 0.76 kg, 320 kg, 75.5 cm^2, and 5.31, respectively). This result suggests that the SNP Harvester method is a good option when multiple SNPs and interaction effects are tested. The significant SNPs could be applied to improve meat quality of Hanwoo via marker-assisted selection. (**Key Words:** Single Nucleotide Polymorphisms, Harvester Method, CCDC158 Gene, Hanwoo)

INTRODUCTION

Detection of genes or single nucleotide polymorphism (SNP) for economically important traits has been extensively performed in farm animals, and so far 5,920 quantitative trait loci (QTL) in cattle were reported from 315 publications (www.animalgenome.org). Most important traits in farm animals are multi-factorial, *i.e.* influenced by interaction of multiple genes and environmental factors. Recently, an advanced SNP genotyping technology such as high throughput SNP chips are available, e.g. the bovine Illumina 770k or Affymetrix 640k SNP arrays. To evaluate whether any SNP is associated with a trait of interest, a large amount of SNPs need to be considered simultaneously, e.g. by fitting the SNPs into the conventional Animal model, which may yield over-parameterization problems.

To handle high-order dimensional data, a multifactor dimensionality reduction method was proposed to efficiently detect multiple genes and interactions effects between the genes (Ritchie et al., 2001; Cho et al., 2004; Su et al., 2012). The method was designed to address high-dimensional data and to uncover complex relationships without relying on the models that fit multiple gene interactions in a parametric fashion (Bastone et al., 2004). Yang et al. (2009) developed a new genetic interaction approach, a SNPHarvester method, to reveal gene interactions and interaction-interaction relationships between a large pool of genes. However, the method was applied only to binary data in a case-control study.

Previously, association studies between CCDC158 gene and growth and carcass traits in Korean cattle, Hanwoo, were performed under linear models, in which a single SNP or haplotype (additive) effects were fitted (Lee et al., 2008; Lee and Lee, 2009; Lee et al., 2010). In this study, a SNPHarvester method with a support vector machine was applied to detect significant SNPs in the CCDC158 gene and interaction effects between the SNPs that were associated with growth and carcass quality traits in Hanwoo.

MATERIAL AND METHODS

Animals and phenotypes

A sample of Hanwoo steers (n = 476) was chosen from

* Corresponding Author: Jong-Joo Kim. Tel: +82-53-810-3027, Fax: +82-53-801-3027, E-mail: kimjj@ynu.ac.kr
[1] Department of Statistics, Yeungnam University, Gyeongsan, 712-749, Korea.
Submitted Dec. 26, 2012; Accepted Feb. 16, 2013; Revised Feb. 22, 2013

the National Livestock Research Institute (NLRI) of Korea. The steers that were sired by 50 Korean proven bulls were born between the spring of 1998 and the autumn of 2002 and reared under a progeny-testing program. All steers were fed under a tightly controlled feeding program at the Daekwanryeong and Namwon branches of NLRI. The steers were castrated at six months of age and each set of four individuals were raised in a pen (4 m×8 m). After six months of age, they were fed with concentrates consisting of 15% crude protein (CP)/71% totally digestible nutrients (TDN) for a period of 60 to 90 d; 15% CP/71% TDN for a period of 180 days; and 13% CP/72% TDN for a period of 90 to 120 days of self-feeding. Roughage was offered *ad libitum*, and steers had free access to fresh water throughout the entire period. After two years, the steers were slaughtered. Average daily gain (ADG) was measured between 6 and 24 months of age. After slaughter, the carcass was chilled for 24 h and cold carcass weight (CWT) was measured. Also, *longissimus dorsi* muscle area (LMA) and marbling score (MS) were measured according to the standards of the Korean Animal Product Grading Service. The means and standard deviations of ADG, CWT, LMA, and MS were 0.752±0.089 kg, 316.8±34.5 kg, 75.3±8.1 cm^2, and 5.61±4.18, respectively.

SNP genotyping

Genomic DNA was extracted from white blood cells using the phenol-chloroform method (Sambrook and Russell, 2001). A total of 19 polymorphic SNPs of the coiled-coil domain containing 158 (CCDC158: Gene ID 534614) were obtained according to Lee et al. (2010). For the SNP genotyping, primers for the amplification and extension were designed for the single-base extension (Vreeland et al., 2002). Primer extension reactions were conducted using the SNaPshot ddNTP Primer Extension Kit (Applied Biosystems, Foster City, CA, USA). For the cleanup of the primer extension reaction, one unit of SAP (shrimp alkaline phosphatase) was added to the reaction mixture, and this mixture was incubated for 1 h at 37°C, followed by 15 min at 72°C for enzyme inactivation. DNA samples containing extension products and the Genescan 120 LIZ size standard solution were added to HiDi formamide (Applied Biosystems, Foster City, CA, USA) in accordance with the manufacturer's recommendations. The mixture was incubated for 5 min at 95°C, followed by 5 min on ice, after which electrophoresis was conducted using the ABI PRISM 3130XL Genetic Analyzer. The results were analyzed using GeneMapper v4.0 (Applied Biosystems, Foster City, CA, USA).

SNP Harvester method with a support vector machine

A support vector machine (SVM), a statistical algorithm, has an advantage of solving the problem of nonlinear regression by restructuring high-dimensional spatial data into linear regression functions (Vapnik, 1998). A Soft-margin technique adopting slack variables was applied (Figure 1), which allowed for hyper-plane with minimal misclassification and soft margins (Tan et al., 2006). In the SVM model, a Kernel function of the RBF (radial basis function) was used, for which RBF Gamma 0.1 was set as a default parameter from Modeler 14 (IBM-SPSS, ex-Clementine) and ten was set as a regularization parameter.

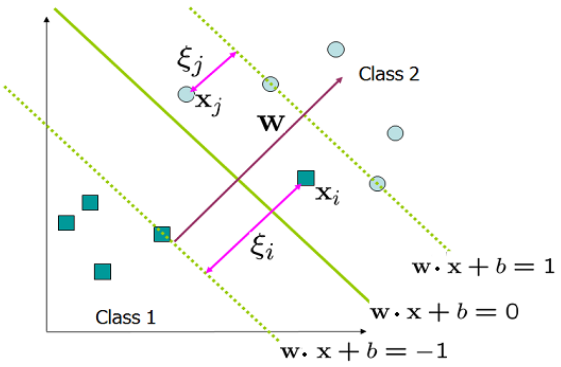

Figure 1. Soft-margin technique with slack variables. (*) ξ_i: slack variable, w, b: parameters.

In the SNP Harvester method that enables to sort out major genotype combinations between genes, several SNPs were selected among a number of SNPs by grouping and exchanging SNPs within a group (Yang et al., 2009). The process was repeated to increase test statistics values. The x^2 statistic, classification accuracy, and *B*-statistic values were used as score functions. The x^2 statistic value was determined with degree of freedom 3^k-1, in which k indicates number of SNP groups, e.g. two or three in this study. To identify statistically significant groups, α was set at 0.001 level. The SNPHarvester procedure is summarized as follows (Figure 2):

Step 1. Randomly select *k* number of groups in the entire SNP groups and assign group name, e.g. group A. Set the rest of SNPs as SNP_i.

Step 2. Exchange SNP_i that do not belong to group A with group A elements on a one-by-one basis to calculate scores.

$A_1 = (SNP_{s1}, SNP_{s2}, \ldots, SNP_i) \Rightarrow$ Score A_1
$A_2 = (SNP_{s1}, SNP_i, \ldots, SNP_k) \Rightarrow$ Score A_2
…
$A_k = (SNP_i, SNP_{s2}, \ldots, SNP_k) \Rightarrow$ Score A_k

Step 3. Set the greatest value from Step 2 as A*.

Step 4. If A* has a greater score than A, then replace A with A*.

Step 5. If the score of the A* is greater than a threshold

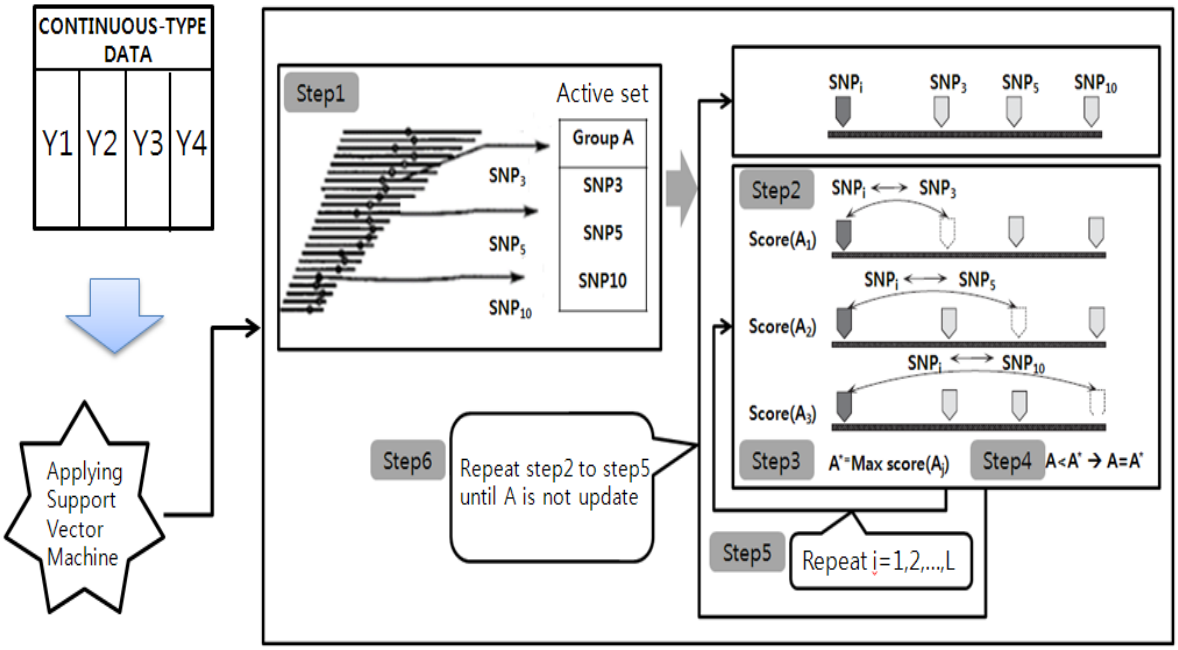

Figure 2. SNP Harvester procedure with a support vector machine.

value, then A* is classified as a significant group.

Step 6. For SNP_{i+1} that do not belong to group A, repeat Steps 2-5.

Step 7. If A* is not replaced with any other SNP_{i+1}, then stop the process and A* is determined as the final SNP combination set.

By repeating the above steps, SNP combinations influencing the test traits were selected. Because the SNPHarvester method was designed to analyze interaction effects for binary traits, the measures of the four traits in this study were converted into binary values under a multi-trait model. The SVM technique was employed by taking the four continuous variables as input variables and the binary value as a dependent variable, and two- or three-way interaction models were applied to determine ten significant SNP combinations.

RESULTS

Table 1 shows the most significant SNP combinations that were related to the four economic traits in Hanwoo. Among the two or three SNP combinations, the set of g.34425+102A>T, g.4102+36T>G, and g.11614+19G>T SNPs yielded the lowest p-value. However, the subsets of the genotypes for the three SNP combinations could not be identified using the SNP Harvester method. Instead, the genotype within the g.34425+102A>T, g.4102+36T>G, and g.11614+19G>T combination was investigated in detail by using the CART algorithm (Table 2). Table 2 shows the best SNP combinations for the four economic traits between superior genotypes and others (not presented here). The AAGTGG genotype combination for the three respective SNPs had the best performance, i.e. the greatest t-values and the lowest p-values (<0.001) for the four economic traits. Mean and standard deviations for the AAGTGG genotype group were 0.82±0.09 kg for ADG, 342.6±23.6 kg for CWT, 80.4±5.9 cm^2 for LMA, and 7.35±4.75 for MS, respectively. These estimates were significantly greater than for the rest of the genotype groups, i.e. 0.75±0.09 kg, 314.2±34.0 kg, 74.7±7.8 cm^2, and 5.01±3.97 for ADG, CWT, LMA, and MS, respectively (Table 2).

For the genotypes of the three SNPs that had the best combination with the greatest performance for the economic traits, i.e. AA, GG, and GT for g.34425+102A>T, g.11614+19G>T, and g.4102+36T>G, respectively, least-squares means were obtained for the genotype and the rest of genotypes when each SNP was analyzed for each trait (Table 3). The results show that when the three SNPs were combined, the estimates were greater than when each SNP was considered (Tables 2 and Table 3). For example, the individuals with AAGTGG combination had an average value of 0.82 kg for ADG, while those with GG genotype for g.11614+19G>T had 0.76 kg, which was the greatest value when single SNPs were analyzed. Also, for CWT, LMA and MS, the estimates of the genotype combination of the three SNPs were 342.6 kg, 80.4 cm^2 and 7.35, while the greatest estimates from single SNP analyses were 319.9 kg, 75.5 cm^2 and 5.31, respectively (Tables 2 and 3).

DISCUSSION

In this study, the bootstrap sampling method (Efron and

Table 1. The most significant sets of SNP combinations among the 19 SNPs of CCDC158 by the SNP Harvester analysis for the four economic traits in Hanwoo

Number of SNPs	SNPs of significant groups			x^2	$-\log_{10}P^a$
2	g.8420-173T>C	g.34425+102A>T		420.3	85.1
	g.11614+19G>T	g.34425+102A>T		401.8	81.1
	g.11614+19G>T	g.4102+36T>G		190.6	36.2
	g.11614+19G>T	g.66995-169insdelC		182.6	34.5
	g.3885-18C>G	g.66995-169insdelC		176.1	33.2
	g.-74-34G>T	g.66995-169insdelC		167.4	31.3
	g.32330-48A>G	g.32488+95		154.8	28.7
	g.3885-18C>G	g.32330-48A>G		141.8	26.0
	g.3885-18C>G	g.4102+36T>G		132.1	24.0
	g.70+20C>T	g.3885-18C>G		85.2	14.4
3	g.34425+102A>T	g.4102+36T>G	g.11614+19G>T	560.2	100.9
	g.34425+102A>T	g.-74-34G>T	g.11614+19G>T	451.8	78.5
	g.8420-173T>C	g.32488+95	g.66995-169insdelC	355.0	58.7
	g.3885-18C>G	g.32488+95	g.66995-169insdelC	342.7	56.3
	g.3885-18C>G	g.32488+95	g.32330-48A>G	229.3	33.7
	g.3885-18C>G	g.8420-173T>C	g.32330-48A>G	170.8	22.5
	g.3885-18C>G	g.8643-21T>C	g.32330-48A>G	150.6	18.8
	g.3885-18C>G	g.8529+19G>A	g.32330-48A>G	143.4	17.5
	g.70+20C>T	g.8529+19G>A	g.32330-48A>G	111.2	11.8
	g.3885-18C>G	g.8643-21T>C	g.8529+19G>A	93.7	8.9

[a] p values were obtained from x^2 statistics with (3^k-1, k = 2 or 3) 8 or 26 degree of freedom.

Tibshirani, 1993) was used to generate 3,830 samples that were based on the 476 steers in Lee et al. (2010), and the top ten SNP combinations of two- and three-way SNP interaction for four economic traits of Hanwoo were selected using the SNPHarvester with SVM method (Table 1). Although multifactor dimensionality reduction (MDR) to detect gene-gene interactions worked well when the number of genes were moderate, in genome-wide association (GWA) studies, direct application of thousands of SNPs is computationally limited (Yang et al., 2009). Further, MDR is computationally intensive, especially when more than 10 polymorphisms are evaluated (Ritchie et al., 2001).

Lee et al. (2010) reported that the single SNPs of g.34425+102 A>T(AA), g.11614+19G>T(GG), and g.4102+36T>G(GT) within CCDC158 gene were associated with body weight and cold carcass weight in Hanwoo. However, they did not report interaction effects between the SNPs. In this study, by applying the SNPHarvester method, the three SNP combinations, i.e. g.34425+102 A>T(AA), g.11614+19G>T(GG), and g.4102+36TM>G(GT), had the greatest test statistics, x^2 value as 560 (Table 1). Also, the estimates of the best genotype combinations for the three SNPs were much greater than the estimates from single SNP analyses, and the differences between the single and combination effects of

Table 2. *t*-test statistics of the four economic traits between the best genotype and other genotypes of the most significant SNP combination

SNP combination	Trait[a]	Genotype	Mean±SD[b]	t-value	p-value
g.34425+102A>T g.4102+36T>G g.11614+19G>T	ADG	AAGTGG	0.82±0.09	8.4	<0.001
		Others	0.75±0.09		
	CWT	AAGTGG	342.6±23.6	13.0	<0.001
		Others	314.2±34.0		
	LMA	AAGTGG	80.4±5.9	10.5	<0.001
		Others	74.7±7.8		
	MS	AAGTGG	7.35±4.75	5.4	<0.001
		Others	5.01±3.97		

[a] ADG = Average daily gain (kg), CWT = Cold carcass weight (kg) , LMA = *longissimus dorsi* muscle area (cm^2), MS = Marbling score.
[b] Standard deviation.

Table 3. Least squares means and standard errors between the best genotype and other genotypes for each of the three SNPs, with which the most significant SNP combination was obtained by the SNPHarvester analysis for the four economic traits in Hanwoo

Traits[a]	SNP	Genotype	Mean±SE	t-value	p-value
ADG	g.34425+102A>T	AA	0.74±0.03	2.42	0.016
		Others	0.75±0.02		
	g.11614+19G>T	GG	0.76±0.02	4.74	<0.001
		Others	0.74±0.02		
	g.4102+36T>G	GT	0.75±0.02	0.55	0.579
		Others	0.75±0.02		
CWT	g.34425+102A>T	AA	315.1±0.9	0.05	0.958
		Others	315.2±0.7		
	g.11614+19G>T	GG	319.9±0.8	8.11	<0.001
		Others	311.1±0.7		
	g.4102+36T>G	GT	313.0±0.8	3.81	<0.001
		Others	317.2±0.7		
LMA	g.34425+102A>T	AA	75.5±0.2	2.83	0.005
		Others	74.7±0.1		
	g.11614+19G>T	GG	75.3±0.2	2.44	0.015
		Others	74.6±0.2		
	g.4102+36T>G	GT	74.6±0.2	2.47	0.014
		Others	75.2±0.2		
MS	g.34425+102A>T	AA	5.31±0.12	2.19	0.029
		Others	5.00±0.08		
	g.11614+19G>T	GG	5.00±0.10	1.31	0.190
		Others	5.17±0.09		
	g.4102+36T>G	GT	5.21±0.10	1.78	0.075
		Others	4.98±0.09		

[a] ADG = Average daily gain (kg), CWT = Cold carcass weight (kg) , LMA = *longissimus dorsi* muscle area (cm^2), MS = Marbling score.

the three SNPs were 0.06 kg, 22.7 kg, 4.9 cm^2 and 2.04 for ADG, CWT, LMA and MS, respectively (Table 2 and Table 3). This result suggests that interaction effects need to be taken into account when multiple SNPs are tested simultaneously to detect significant SNPs for economically important traits in Hanwoo.

In conclusion, the application of SNPHarvester with SVM method could be a good option for multiple SNP analyses, especially to characterize interaction effects between SNPs, and the significant SNPs may be applied via marker-assisted selection to the Hanwoo industry for genetic improvement of the economically important traits.

ACKNOWLEDGEMENTS

Jea-Young Lee's work was supported by the Yeungnam University Research Grant 2010.

REFERENCES

Bastone, L., M. Reilly, D. J. Rader and A. S. Foulkes. 2004. MDR and PRP: A comparison of methods for high-order genotype-phenotype associations. Hum. Hered. 58:82-92.

Cho, Y. M., M. D. Ritchie, J. H. Moore, J. Y. Park, K. U. Lee, H. D. Shin, H. K. Lee and K. S. Park. 2004. Multifactor-dimensionality reduction shows a two-locus interaction associated with type 2 diabetes mellitus. Diabetologia 47:549-554.

Efron, B. and R. Tibshirani. 1993. An introduction to the bootstrap. Champman & Hall/CRC, Florida, USA.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in predicting of breeding values. Biometrics 32:69-83.

Lee, J. Y., J. C. Kwon and J. J. Kim. 2008. Multifactor Dimensionality Reduction (MDR) analysis to detect single nucleotide polymorphisms associated with a carcass trait in a Hanwoo population. Asian Australas. J. Anim. Sci. 21:784-788.

Lee, J. Y. and H. G. Lee. 2009. A study on the comparison between E-MDR and D-MDR in continuous data. Comm. Korean Stat. Soc. 16:579-586.

Lee, Y. S., D. Y. Oh, J. J. Kim, J. H. Lee, H. S. Park and J. S. Yeo. 2010. A single nucleotide polymorphism in LOC534614 as an unknown gene associated with body weight and cold carcass weight in Hanwoo (Korean cattle). Asian Australas. J. Anim. Sci. 23:1543-1551.

Ritchie, M. D., L. W. Hahn, N. Roodi, W. D. Dupont, F. F. Parl and J. H. Moore. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen- metabolism genes in sporadic breast cancer. Am. J. Hum. Genet. 69:138-147.

Sambrook, J. and D. W. Russell. 2001. Molecular cloning: A Laboratory Manual. 3rd Ed., Cold Spring Harbor Laboratory Press, New York, ISBN-13: 9780879695774, Page 99.

Su, M. W., K. Y. Tung, P. H. Liang , C. H, Tsai, N. W. Kuo and Y. L. Lee. 2012. Gene-gene and gene-environmental interactions of childhood asthma: A multifactor dimensionality reduction approach. PloS One 7:1-9.

Tan, P. N., M. Steinbach and V. Kumar. 2006. Introduction to data mining. Addison Wesley. New York, USA.

Vapnik, V. 1998. Statistical learning theory. 1st Ed. Wiley, Boston, USA.

Vreeland, W. N., R. J. Meagher and A. E. Barron. 2002. Multiplexed, high through put genotyping by single base extension and end labeled free solution electrophoresis. Anal. Chem. 74:4328-4333.

Yang, C., Z. He, X. Wan, Q. Yang, H. Xue and W. Yu. 2009. SNPHarvester: A filtering-based approach for detecting epistatic interactions in genome-wide association studies. Bioinformatics 25:504-511.