A Restricted Partition Method to Detect Single Nucleotide Polymorphisms for a Carcass Trait in Hanwoo

Ji-Hong Lee[1], Dong Chul Kim, Jong-Joo Kim[2] and Jea-Young Lee*

Department of Statistics, Yeungnam University, Gyeongsan, Korea

ABSTRACT : The purpose of this study was to detect SNPs that were responsible for a carcass trait in Hanwoo populations. A non-parametric model applying a restricted partition method (RPM) was used, which exploited a partitioning algorithm considering statistical criteria for multiple comparison testing. Phenotypic and genotypic data were obtained from the Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea, in which the pedigree structure comprised 229 steers from 16 paternal half-sib proven sires that were born in Namwon or Daegwanryong livestock testing station between spring of 2002 and fall of 2003. A carcass trait, *longissimus dorsi* muscle area for each steer was measured after slaughter at approximately 722 days. Three SNPs (19_1, 18_4 and 28_2) near the microsatellite marker ILSTS035 on BTA6, around which the quantitative trait loci (QTL) for meat quality were previously detected, were used in this study. The RPM analyses resulted in two significant interaction effects between SNPs (19_1 and 18_4) and (19_1 and 28_2) at $\alpha = 0.05$ level. However, under a general linear (parametric) model no interaction effect between any pair of the three SNPs was detected, while only one main effect for SNP19_1 was found for the trait. Also, under another non-parametric model using a multifactor dimensionality reduction (MDR) method, only one interaction effect of the two SNPs (19_1 and 28_2) explained the trait significantly better than the parametric model with the main effect of SNP19_1. Our results suggest that RPM is a good alternative to model choices that can find associations of the interaction effects of multiple SNPs for quantitative traits in livestock species. (**Key Words :** SNP, Hanwoo, Carcass Trait, Restricted Partition Method)

INTRODUCTION

Most traits of economic importance in livestock species such as growth and carcass quality are multi-factorial, *i.e.*, influenced by multiple genes and their interactions with environmental factors. Many studies to detect genes responsible for economic traits in farm animals have been undertaken and numerous quantitative trait loci (QTL) are reported (www.animalgenome.org). Parametric models are used to test association of SNPs in the QTL region with the traits of interest, *e.g.* general linear models or the Animal model (Henderson, 1976), in which the main effects to be fitted are additive and dominance mode of gene action.

The rapid development of sequencing technologies such as next generation sequencing and high throughput

genotyping platforms has enabled to genotype large volumes of SNPs using high density chips (Goddard and Hayes, 2009). However, over-parameterization problems can arise, especially when multiple SNPs and interactions between the SNPs are taken into account. As an option for better detection of multiple genes and their interaction effects, Ritchie et al. (2001) proposed a multifactor dimensionality reduction (MDR) method to handle high-order dimensional data and to uncover complex relationships without relying on specified models that fit interaction effects of the multiple genes (Bastone et al., 2004). The MDR method was applied to detect SNPs for meat quality in Hanwoo and the results were better at identifying interaction effects between multiple SNPs (Lee et al., 2008a). Culverhouse et al. (2004) also reported a restricted partition method (RPM) which was based on a partitioning algorithm and statistical criteria from a multiple comparison tests.

Herein, we report results of association tests applying a modified RPM to test main and interaction effects of multiple SNPs on meat quality around the QTL on BTA6 in Hanwoo. The QTL on BTA6 was located near the ILST035

* Corresponding Author : Jea-Young Lee. Tel: +82-53-810-2326, Fax: +82-53-810-4615, E-mail: jlee@yu.ac.kr
[1] Gyeongbuk Livestock Research Institute, Yeongju, Gyeongbuk 750-781, Korea.
[2] School of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, Korea.

microsatellite region in the cattle (Yeo et al., 2004).

MATERIALS AND METHODS

Animals, phenotypes and genetic markers

Collection of data was described in detail in Lee et al. (2008b). Data were obtained from the Hanwoo Improvement Center, National Agricultural Cooperation Federation, Korea, for the 229 steers that were born at Namwon or Daegwanryong livestock testing station from 16 paternal half-sib proven sires between spring 2002 and fall 2003. Each steer was slaughtered at approximately 722 days of age. The *longissimus dorsi* muscle area (LMA) was measured according to the standards of the Korean Animal Products Grading Service. QTL for meat quality were detected around ILSTS035 on BTA6 in previous studies (Yeo et al., 2004). Three SNPs were chosen near the QTL, *i.e.* AH1_5 (28_2), 31465_446 (18_4) and 12273_165 (19_1), which was based on EST-based linkage map (Lee et al., 2008b).

Statistical analysis using general linear model

The parametric model to be applied was described in Lee et al. (2008a). The general linear model was:

$$Y_{ijklmn} = \mu + C_i + S_j + \beta X_{age} + M1_k + M2_l + M3_m + M1M2_{kl} + M1M3_{km} + M2M3_{lm} + M1M2M3_{klm} + \varepsilon_{ijklmn}$$

where, Y_{ijklmn} is an observed phenotype, μ is the overall mean, C_i is the ith contemporary group ($i = 1$ to 8), S_j is the jth sire's random effect ($j = 1$ to 16), β is a linear effect of the steer's age, Mi_k is the kth genotype effect of a marker ($k = 1,2,3$) for SNP i, $MiMj$ is an interaction term between markers i and j, $M1M2M3$ is a three-way interaction term for the three markers and ε_{ijklmn} is random error. The contemporary group was defined as a group of individuals fed in the same pen and slaughtered on the same date. The analyses were performed using MIXED procedure of SAS v.9.1.

Restricted partition method (RPM) analysis

The RPM algorithm was based on Culverhouse et al. (2004), which was an iterative search procedure to find the best partition of the genotypes. Genotypes were sequentially merged based on the similarity of the mean values of their quantitative trait. Initially, each multi-locus
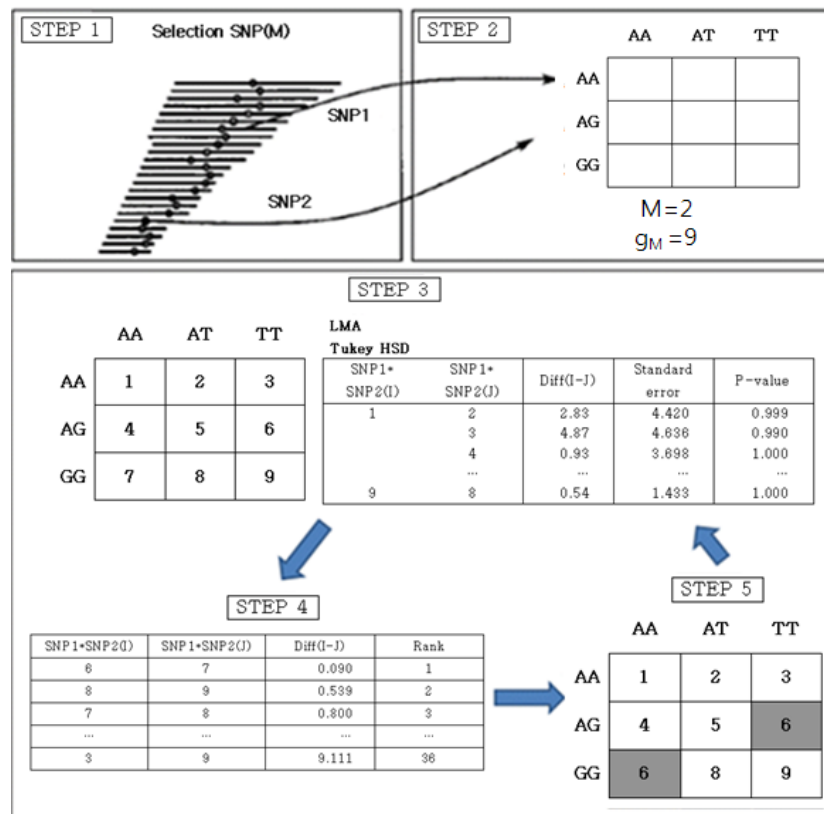

Figure 1. Restricted partition method (RPM) procedures: Steps 1-5 are an iterative search procedure to find the best partition of the genotypes in explaining the variation of the *longissimus dorsi* muscle area in Hanwoo. Genotypes are sequentially merged based on the similarity of the mean values of the trait. Selection of the genotypes to be merged at each step was based on statistical criteria from multiple comparison tests.

genotype formed its own group, and the algorithms were processed as following (Figure 1):

Step 1. A set of *M* genetic factors is selected from the pool of all factors.

Step 2. The *M* factors and their possible multifactor classes or cells are represented in *M*-dimensional space.

Step 3. A multiple comparison test is performed to identify which (if any) groups have significantly different mean quantitative trait values. The procedure halts if all groups have different means.

Step 4. Pairs of genotype groups with means that are not significantly different from each other are ranked according to the difference in means between the two groups.

Step 5. The pair with the smallest difference (i.e., most similar mean values) is merged to form a new group.

Step 6. The algorithm returns to step 3.

r^2 value was estimated for the model with the quantitative trait values, which was regressed on the final genotype group. If the genotypes were merged into a single group, the r^2 value was assigned as 0, indicating the lack of evidence for differences of the testing trait between the genotypes. The r^2 value was also used to derive a measure of statistical significance for the results.

In step 3, Tukey's HSD (honest significant difference) was applied at $\alpha = 0.05$ level for multiple comparison testing (Games and Howell, 1976). Each time of iteration from step 5 to step 3, the number of defined groups of genotypes would be reduced, *i.e.* the genotypes that had no significant difference on the phenotype were merged into one genotype group.

Experimental-wise p values for r^2 value were obtained using permutation tests (Good, 1994). Data were generated by permuting the trait values against genotypes and fixed effects. Ten thousand permuted samples were generated and the RPM analysis was performed on each permuted sample. Significance was estimated by the frequency with which the r^2 value from the analysis of the original data exceeded the permuted r^2 values.

RESULTS AND DISCUSSION

The results under the parametric mixed model by using SAS were reported in Lee et al. (2008a). There was no statistical significance for the interaction effects between the three SNPs (p>0.05), and only the additive effect of the SNP (19_1) was significant for LMA (p<0.01).

The RPM analyses, however, showed that the model with SNP19_1 had the r^2 value of 0.027, which did not have a statistical significance at the $\alpha = 0.05$ level (Table 1). When considering interaction effects between any two SNPs, the model fitting SNP19_1 and SNP18_4 had the value of 0.042 with a statistical significance (p = 0.052),

Table 1. r^2 values and p value to test main and interaction effects of three SNPs on *longssimus dorsi* muscle area in Hanwoo using RPM analysis*

Number of factors	Marker	r^2	p value
1	19_1 (SNP1)	0.027	0.0702
	18_4 (SNP2)	0	
	28_2 (SNP3)	0	
2	19_1*18_4	0.042	0.0520
	19_1*28_2	0.050	0.0020
	18_4*28_2	0.028	0.0865

* The RPM was a non-parametric, model-free method to reduce multifactor-dimensionalities considering interaction effects between SNPS. r^2 and p values were empirically obtained with 10,000 permutated samples.

and the interaction model with SNP19_1 and SNP28_2 had the largest r^2 value with significant statistical support (p = 0.002).

A novel nonparametric method, RPM, was applied to reduce the dimensionality caused by simultaneously fitting multiple SNPs and their interactions. In the RPM method, the genotypes that did not have a significant difference on LMA were merged in a sequential fashion, which was based on the similarity of the mean values of the trait. Our results showed that in the two interaction models in which the SNP interaction term was fitted, the combination of SNPs (19_1 and 18_4) and (19_1 and 28_2) had significant effects on LMA at the $\alpha = 0.05$ level, while there was no statistical significance when each of the three SNPs was fitted individually in the model (Table 1). Previously, the MDR was applied using the same sample as in this study. Only the model with the interaction effect of the two SNPs (19_1 and 28_2) explained the LMA significantly better than the model with the main effect of SNP19_1 (Lee et al., 2008a). However, the RPM analysis allowed detection of an additional SNP interaction effect (SNPs 19_1 and 18_4) with greater statistical support than the additive model with SNP19_1 (Table 1).

The linear mixed model did not detect any interaction effects between the SNPs, but detected only the main effect of one SNP (19_1). As described in Lee et al. (2008a), parametric models are based on parameter orders such that quadratic *e.g.* interactions between two SNPs or higher-orderly effects are tested conditionally on the main effects, which may affect power to detect interaction effects between SNPs. The RPM as well as the MDR method is, however, free of factor-dimension orders, allowing detection of the interaction effect with significant contribution to phenotype variation without any restriction of parameter dimensionality of the model.

Two significant interaction effects were found in the RPM, while no main SNP effect was detected (Table 1). However, only the one (SNPs 19_1 and 28_2) or no interaction effect was significant in the MDR and the

parametric model, respectively (Lee et al., 2008a). This result suggests that the RPM may provide more preferable conditions to detect interaction effects between SNPs, suggesting that RPM is a good alternative to model choices that can find associations of the interaction effects of multiple SNPs for quantitative traits in livestock. However, the number of SNPs and phenotypes in this study was not enough to get reliability in evaluating interaction effect of multiple SNPs. Thus, further studies are needed to validate the efficiency of the RPM to better characterize SNP effect on carcass traits in Hanwoo.

One disadvantage in implementation of the RPM is that computation time is very demanding when high-throughput SNP arrays are used such as the Illumina 50K or Affymetrix bovine 640k SNP arrays (Barendse et al., 2007), which renders it infeasible in terms of current computation capabilities. However, application of the method with moderate numbers of genes or SNPs, *e.g.*, tens or hundreds of significant SNPs that have been detected after whole genome association analysis, may be a good strategy to better identify interaction effects between the SNPs. We already detected several tens of SNPs with additive mode of gene action that were significantly associated with carcass quality and body conformation traits in Hanwoo (Alam et al., 2010; Lee et al., 2011). Implementation of the RPM using the much greater number of multiple SNPs would allow better characterization of SNP effects on carcass traits in Hanwoo.

ACKNOWLEDGEMENTS

REFERENCES

Alam, M., Y. M. Lee, B.-L. Park, J.-H. Kim, S.-S. Lee, H.-D. Shin, K.-S. Kim, N.-S. Kim and J.-J. Kim. 2011. A whole genome association study to detect single nucleotide polymorphisms for body conformation traits in a Hanwoo population. Asian-Aust. J. Anim. Sci. 24(3):322-329.

Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris and M. B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics 176:1893-1905.

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.

Culverhouse, R., T. Klein and W. Shannon. 2004. Detecting epistatic interactions contributing to quantitative traits. Genet. Epidemiol. 27:141-152.

Games, P. A. and J. F. Howell. 1976. Pairwise multiple comparison procedures with unequal n's and/or variances. J. Educ. Behav. Stat. 1:113-125.

Goddard, M. E. and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. Nat. Rev. Genet. 10:381-391.

Good, P. 1994. Permutation test: a practical guide to resampling for testing hypotheses. Springer, New York.

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. 2008a. Multifactor dimensionality reduction (MDR) analysis to detect single nucleotide polymorphism associated with a carcass trait in a Hanwoo population. Asian-Aust. J. Anim. Sci. 21(6):784-788.

Lee, Y. S., J. H. Lee, J. Y. Lee, J. J. Kim, H. S. Park and J. S. Yeo. 2008b. Identification of candidate SNP for growth and carcass traits related to QTL on chromosome 6 in Hanwoo (Korean cattle). Asian-Aust. J. Anim. Sci. 21(12):1703-1709.

Lee, Y.-M., C.-M. Han, Y. Li, J.-J. Lee, L. H. Kim, J.-H. Kim, D.-I. Kim, S.-S. Lee, B.-L. Park, H.-D. Shin, K.-S. Kim, N.-S. Kim and J.-J. Kim. 2010. A whole genome association study to detect single nucleotide polymorphism for carcass traits in Hanwoo population. Asian-Aust. J. Anim. Sci. 23(4):417-424.

Ritchie, M. D., L. Hahn, W. Roodi, L. R. Bailey, 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(1):138-147.

Yeo, J. S., J. Y. Lee and J. W. Kim. 2004. DNN marker mining of ILST035 microsatellite locus on chromosome 6 of Hanwoo cattle. J. Genet. 83:245-250.