Proc. Assoc. Advmt. Anim. Breed. Genet. 20:266-269

A STUDY ON EFFECTS OF FAMILY AND HAPLOTYPE BLOCKS ON CONSERVATION OF GENE EXPRESSION TRAITS IN HALF SIB SHEEP FAMILIES

H.A. Al-Mamun[1], P. Kwan[1], R.L. Tellam[2], J.W. Kijas[2] and C. Gondro[3]

[1]School of Science and Technology, University of New England, Armidale NSW 2351, Australia. [2]CSIRO Animal, Food and Health Sciences, Queensland Bioscience Precinct, St. Lucia, Brisbane, QLD 4067, Australia. [3]The Centre for Genetic Analysis and Applications, University of New England, Armidale NSW 2351, Australia

SUMMARY

The objective of this study was to explore the relationship between SNP and haplotype variation on gene expression traits. The data used included expression levels from 24,128 probe sets of *logissimus lumborum* muscle from 38 half-sib Poll Dorset sheep from six families and genotypes from 49,034 SNPs collected from the same animals. The analytical approaches used sought to analyse the effects of family and haplotype blocks on conservation of gene expression traits in this sheep population. Our study indicated that there is a genetic component in gene expression traits and hence gene expression is heritable to non-negligible extent. On average, our estimated heritability for gene expression obtained from skeletal muscle samples of sheep is 0.27 and 0.29 based on two different approaches. These preliminary results are consistent with previous heritability estimates.

INTRODUCTION

A primary goal in molecular biology is to understand how patterns of genetic variation affect the gene expression levels and higher level phenotypes. In recent years, studies of the relationship between genotype and gene expression, or other quantitative traits, have gained considerable attention due to the availability of high throughput technologies in profiling single nucleotide polymorphisms (SNPs) data and global gene expression. Several studies have suggested that the variation in gene expression traits is associated with genetic variation such as SNPs and copy number variants (CNVs) (Spielman *et al.* 2007; Stranger *et al.* 2007), and have demonstrated that a significant proportion of gene expression is heritable both in human (Cheung *et al.* 2003; Price, *et al.* 2011) and in other organisms (Nätt *et al.* 2012; Schadt *et al.* 2003). Most of these association studies comprised a large numbers of SNP from multiple individuals, and made use of the allele frequencies to search for associations with variation in trait data. One potential drawback of this approach is the large number of SNP–wise testing required and the potential for false positive outcomes. Moreover, these methods did not consider the information present in associations between neighbouring SNPs. Neighbouring SNPs tend to be inherited as blocks (Daly *et al.* 2001). These haplotype blocks can be used to find associations with quantitative traits such as gene expression traits. This strategy decreases the impact of multiple testing corrections as fewer hypotheses are tested.

In this study, SNPs and gene expression data obtained from 38 half-sib sheep were used to (i) quantify the heritability of gene expression in a sheep population and (ii) determine the degree of conservation of the gene expression between haplotype blocks within different families of the sheep population.

MATERIAL AND METHODS

Animals. 38 progeny (18 months old ewes) from six Poll Dorset sires (4-8 progeny/sire) were used for genotyping and microarray analysis of skeletal muscle samples. The six sires were

grouped into a high muscling and low muscling sire group (Table 1) based on their yearling trait, Eye Muscle Depth (EMD). Details of these sires have been described in (Kogelman *et al.* 2011).

Table 1 Number of progeny in each family

Family	2	5	7	11	16	17
No. of Progeny	7	8	4	8	8	5
Muscling Group	High	Low	High	High	Low	Low

Data pre-processing and normalization. The Affymetrix GeneChip® Bovine Genome Array (Affymetrix Santa Clara, CA) was used to measure the gene expression of the 40 animals. The Affymetrix GeneChip contains 24,128 probe sets, representing ~19,000 UniGene clusters. The microarray data were initially processed using the statistical software R (http://www.r-project.org) and additional Bioconductor packages (http://www.bioconductor.org). Normalization was performed using the RMA (Robust Multi-chip Average) method. After normalization and removing the control probe sets, 24,016 probe sets remained for further analysis. Linear Models for Microarray Data (limma) package from Bioconductor were used for differential gene expression analysis. Genotyping was undertaken using the Illumina 50K Ovine SNP chip containing 49,034 SNPs and 38 animals genotyped. The SNP data were pre-processed using the software PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) and 47,680 SNPs remained for further analysis. These 47,680 SNPs were subjected to phasing and haplotype block construction using the method described in (Ferdosi *et al.* 2013).

GRM and IBD estimates. The Genetic Relationship Matrix (GRM) was calculated according to the VanRaden's method (VanRaden, 2007) and the Identity by Decent (IBD) values for each pair of animals were computed using the method described in Price *et al.* 2011. The whole genome was partitioned into 2Mbp blocks and for each block 2x2=4 comparisons were performed between haplotypes from each pair of animals. For each chromosome, 0.5 Mbp from each end were excluded as these data might be noisy and could affect subsequent analysis. We did not consider sex chromosomes in our IBD calculation. Two haplotypes were considered IBD if they matched at > 95% of alleles in the block. Local IBD was defined as the total number of comparisons that produced a match. Genome-wide IBD was computed as the average of the local IBD estimates across all 2Mbp blocks.

Heritability estimates using IBD and GRM. Narrow sense heritability (Visscher *et al.* 2008) was calculated using variance-components analysis (Amos, 1994). We followed the method described in (Price, *et al.* 2011) and used their source code to calculate a heritability estimate for each gene. Let e_{gs} denote normalized gene expression of gene g for each individual animal s and θ_{st} denotes the genome-wide IBD or GRM between the individuals s and t ($0 \leq \theta_{st} \leq 1$). $\Theta = (\theta_{st})$ was assigned the NxN matrix of genome-wide IBD or GRM, where N is the number of animals. Vg was the covariance matrix of normalized gene expression for gene g. We fitted h_g^2, the heritability of gene g, using the model $Vg = h_g^2 \Theta + (1 - h_g^2)I$ to the observed normalization gene expression values e_{gs} by maximizing the likelihood $L(e_g|V_g) \propto \frac{1}{\sqrt{\det(V_g)}} \exp\left(-\frac{1}{2} e_g{}^T V_g^{-1} e_g\right)$, where $e_g = (e_{gs})$.

RESULTS

Variation of gene expression is higher between individuals than within families. After pre-processing and normalization, differential gene expression analysis was performed using the 24,016 probes. Between families, 473 genes were significantly differentially expressed (DE). The lower number of DE genes might be an effect of the small sample size (4-8 animals/sire). For each

DE gene, we calculated the variance among all 38 animals (i.e. total variance) and the variance within each family. As a measure of variability, we then calculated the ratio between the total variance and the variance within each family. For most genes, this ratio had a value greater than one, suggesting higher variation in gene expression among the population than within family. Figure 1 shows a scatter plot of variance in gene expression level among the population and between individuals from Family 11 for the 473 DE genes. As all the progeny were raised in the same places and in the same condition to minimize the environmental variation, the results suggest that a significant portion of the variation in gene expression is genetically determined and thus there exists a heritable component in gene expression.

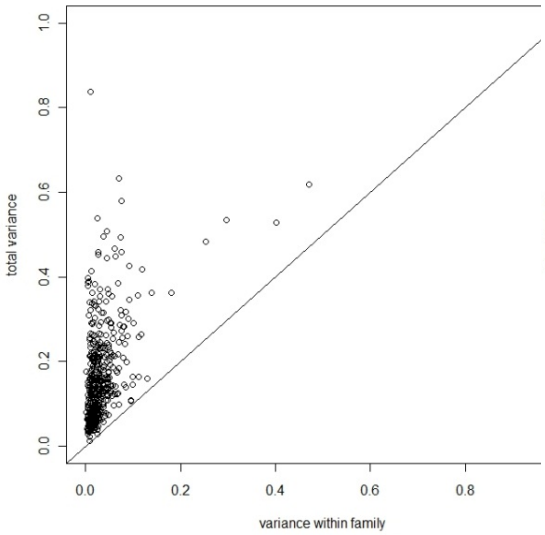

Figure 1. Scatter plot of total variance vs. variance within Family 11 for 473 differentially expressed genes.

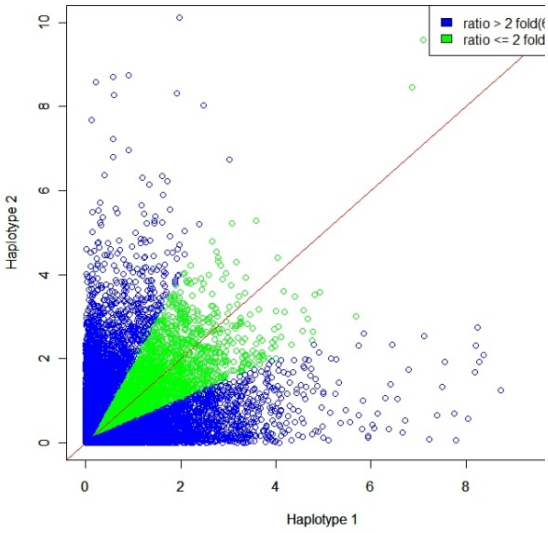

Figure 2. Scatter plot of gene expression level of haplotype 1 group vs. haplotype 2 group in Family 11.

Heritability of gene expression. For the analysis of the gene expression, the normalized intensity values for 22,246 probe sets (probe sets on the X chromosome were removed) were co-analysed along with SNP data from the 38 animals. Two animals were discarded that did not have SNP data. Using the genome-wide IBD and GRM, the overall heritability h_g^2 was estimated for each gene *g* using the variance-component method described in the Methods section. We then computed the overall heritability of gene expression h^2, by averaging all h_g^2 values. The result showed h^2 = 0.25 (standard error ±0.0023) when using the IBD matrix and h^2 = 0.24 (standard error ±0.0027) using the GRM matrix. Some negative values for h_g^2 were observed which do not have any biological interpretation and in most cases these values are very close to zero. These might be attributed to statistical noise. If we ignore negative values and assign each to zero, we obtained h^2 = 0.27 (standard error ±0.0021) and h^2 = 0.29 (standard error ±0.0024) when we used the IBD matrix and the GRM matrix respectively. Both estimates are consistent with previous results which reported that a significant portion of gene expression is heritable at the level of h^2 = 0.3 or higher (McRae *et al.* 2007; Price *et al.* 2011).

Gene expression varies between haplotype groups within families. For each gene within a family, the animals were grouped into two groups (i.e. haplotype 1, if the gene comes

from sire's parental strand and haplotype 2, if the gene comes from sire's maternal strand). Then the variance of gene expression within each haplotype group for each gene was calculated. As a measure of variability of gene expression between two haplotype groups, the variance ratio for each gene was calculated by dividing the variance of the expression levels from the haplotype 1 group by the variance of the expression levels from the haplotype 2 group. This revealed 65% (family 16) to 78% (family 7) of genes showed at least two-fold difference between the variances of the gene expressions in the haplotype 1 group and the haplotype 2 group. These percentages are much greater than expected from random ($P < 10^{-10}$) for every family tested. The results achieved suggested that there are differences in gene expression if the gene is coming from sire's parental or maternal side. Figure 2 shows a plot for the variance of gene expression level of the haplotype 1 group against the variance of gene expression level of haplotype 2 group for Family 11. This demonstrated that a significant number of values deviated from the straight line indicating equal variance for the two groups.

Family effect and haplotype effect on gene expression traits. We wanted to ascertain (1) if family and haplotype affect gene expression levels, and (2) if there is any variation in gene expression between the families. To test the hypothesis that there are family and haplotype effects on gene expression traits, a linear model was fitted in R (expression ~ family + haplotype + family * haplotype). Then, we conducted analysis of variance (ANOVA) test using this linear model. The result was a highly significant effect of family on the gene expression traits ($F = 18.6161$, $P < 2.2e-16$). Further, the effect of the interactions between family and haplotype were also highly significant ($F = 3.6527$, $P < 0.002$), although the haplotypes themselves did not have any significant impact on the gene expression traits.

ACKNOWLEDGEMENTS

This study was funded by SheepGenomics and the Australian Research Council project DP130100542.

REFERENCES

Amos, C. I. (1994) *Am J Hum Genet* **54**: 535.

Cheung, V. G., Conlin, L. K., Weber, T. M., Arcaro, M., Jen, K. Y., Morley, M., and Spielman, R. S. (2003) *Nature genetic.* **33(3)**: 422.

Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J., and Lander, E. S. (2001) *Nature Genetics.* **29**: 229.

Ferdosi, M. H., Kinghorn, B. P., Werf, J. H. v. d., and Gondro, C. (2013) *Bioinformatics, Under review*.

Kogelman, L. J., Byrne, K., Vuocolo, T., Watson-Haigh, N. S., Kadarmideen, H. N., Kijas, J. W., Tellam, R. L. (2011) *BMC genomic.* **12:** 607.

McRae, A. F., Matigian, N. A., Vadlamudi, L., Mulley, J. C., Mowry, B., Martin, N. G., Visscher, P. M. (2007) *Human molecular genetics.* **16(4)**: 364.

Nätt, D., Rubin, C.-J., Wright, D., Johnsson, M., Beltéky, J., Andersson, L., and Jensen, P. (2012) *BMC genomics.* **13(1)**: 59.

Price, A. L., Helgason, A., Thorleifsson, G., McCarroll, S. A., Kong, A., and Stefansson, K. (2011) *PLoS genetics.* **7(2)**: e1001317.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., *et al.* (2003) *Nature.* **422**: 297.

Spielman, R. S., Bastone, L. A., Burdick, J. T., Morley, M., Ewens, W. J., and Cheung, V. G. (2007) *Nature genetics.* **39(2)**: 226.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Dermitzakis, E. T. (2007) *Science.* **315(5813)**: 848.

VanRaden, P. M. (2007) *Interbull Annual Meeting Proceedings*. **37**:33.

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008) *Nature reviews. Genetics.* **9(4)**: 255.