Proc. Assoc. Advmt. Anim. Breed. Genet. 19:387-390

ASSOCIATION BETWEEN SNP INTENSITY DATA AND GROWTH AND MEAT YIELD TRAITS IN SHEEP

G.M. Payne[1,2], M.E. Goddard[2,3], J.C. McEwan[1], M.A. Black[4] and K. G. Dodds[1]

[1]AgResearch, Invermay Agricultural Centre, Puddle Alley, Mosgiel, New Zealand
[2]Department of Agriculture and Food Systems, University of Melbourne, Australia
[3]Department of Primary Industries, 1 Park Drive, Bundoora, Victoria 3083, Australia
[4]Department of Biochemistry, University of Otago, New Zealand

SUMMARY

SNPs from the Illumina OvineSNP50 BeadChip are used for, among various other analyses, genome wide selection (GWS) and association studies. There are over 50,000 SNPs on this chip. However, these SNPs are filtered prior to downstream analysis, to include only those with reliable genotype information. This study uses information from previously discarded SNPs in an association analysis with growth and meat yield traits to determine if they contain currently untapped genetic variation which is associated with these traits. Initial results have identified four of these SNPs that are significantly associated with phenotypes after adjustment for multiple testing. These SNPs were found to have moderate heritability.

INTRODUCTION

The Illumina OvineSNP50 BeadChip is a commonly used platform for generating genotype information, with these data then used to investigate genotypic relationships with sheep traits of interest. The SNP assay yields two pieces of data for each animal and SNP – the genotype and the intensity. Usually only the genotype is tested for association with phenotype but it is also possible that the intensity contains useful information about the DNA sequence in the vicinity of the probe. For instance, the intensity might indicate the number of copies of a copy number variant (CNV) or the presence of another polymorphism nearby. These causes of variation in intensity often make it impossible to call the SNP genotype and so these SNPs have been discarded for the analysis of genotypic associations with phenotype. However, the intensity information at these SNPs might still be useful information that is associated with phenotype and useful for GWS. For instance, CNVs may be associated with phenotype (Stranger *et al* 2007) but they are likely to make it impossible to cluster genotypes at a SNP within the CNV into three clear clusters or genotypes. However, if the intensity of the SNP indicates the number of copies of the CNV it may be correlated with phenotype. This paper is an exploration into using SNP intensity values of currently discarded SNPs as potential sources of genetic variation. We performed an association analysis between log intensity values of SNPs, at which genotype was not called, and weaning weight, liveweight at 8 months, liveweight at 12 months, carcass weight, ultrasonic eye muscle depth and width and ultrasonic fat depth.

METHODS

Animal resource. Sheep used in this analysis come from multiple flocks, multiple years and are a commercial resource. They consist primarily of four breeds – Romney, Perendale, Coopworth and Texel. Romney, Perendale and Coopworth composite animals are also included in the analysis. Rams with recorded offspring were primarily used in the analysis.

SNP data. SNP data were obtained from the Illumina 50K ovine SNP chip. This technology uses two dyes to genotype SNPs – one dye for each allele at a given SNP. SNPs that were classified

according to Illumina criteria as intensity only, zeroed or with a nearby polymorphism or deletion were selected. These SNPs will be referred to as "intensity only" SNPs for the remainder of this paper. An overall intensity (I) value is derived from the two dye (x and y) intensities as,

$$I_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2}$$

where, i=SNP and j=animal, this notation is used for the remainder of this paper unless stated otherwise. The average intensity over all autosomal SNPs (over and above selected SNPs, n=47,318) was calculated per animal. To account for differences between animals, an adjusted intensity value (logr) for selected SNPs (i=1..1081) for each animal was derived as,

$$logr_{ij} = log_2\left(\frac{I_{ij}}{\sum_{i'=1}^{47318} I_{i'j}}\right)$$

Principal component analysis (PCA) was performed on the animals using a filtered sub-set of SNPs (n=47,656). Filters applied were 1) SNP must be autosomal, 2) minor allele frequency (MAF) must not be equal to zero, 3) weighted Illumina gencal10 score of < 0.422 and 4) SNP must not deviate greatly from Hardy-Weinberg equilibrium.

Trait data. Data from seven traits was used – weaning weight (wwt), liveweight at 8months (lw8), liveweight at 12 months (lw12), carcass weight (cw), ultrasonic eye muscle depth (umd), eye muscle width (umw) and ultrasonic fat depth (ufd - measured above the eye muscle). For each animal and each trait a 'phenotypic' record was calculated from the data normally used to estimate breeding values (BVs). Phenotypic records were corrected for non-direct genetic effects, such as maternal genetics. Correlated traits were used to help estimate BVs. Reliabilities of these phenotypic records were calculated using the standard errors of the BV predictions, adjusted for contemporary group, where contemporary group consists of birth year, flock, mob information and sex. To ensure the BV variation was representative of the true phenotypic variation, BVs were deregressed by their reliability. Own and progeny deregressed BVs were combined using the method of Mrode and Swanson (2003). Progeny BVs were not combined with own values when an animal was seen as both a progeny for a sire with genotype information and had its own genotype and trait information. This was done to prevent double counting of trait information.

SNP intensity heritability. Sixty eight intensity only SNPs were tested to determine if they were heritable or not, using Asreml (Gilmour *et al.* 2008), with model,

$$logr_{ij} = platform_j + b_i\ theta_{ij} + animal_j$$

where platform is the genotyping laboratory; theta is the value that indicates what genotypes are likely present for a given SNP, as it is a function of relative dye fluorescence, and thus provides information about the alleles present at the SNP locus; and animal is a random effect with variance matrix equal to the numerator relationship (A) matrix times the genetic variance. For some SNPs, the level of intensity differs between genotypes. Theta was fitted to remove any dye effects and hence get more accurate heritability estimates. Heritability estimates of SNPs on the X chromosome were calculated using the !XLINK function of Asreml to calculate a relationship matrix for X-linked inheritance (Fernando and Grossman 1990). The number of animals tested per SNP ranged from 2,349 to 2,691.

Association testing. The total number of animals with both SNP and trait data ranged from 1,546 for umd to 2,389 for lw8 and lw12. Asreml was used to fit a linear mixed (animal) model to test for an association between the adjusted intensity values (for each of the 1081 intensity only SNPs)

and each of the 7 traits. The model used was,

$$trait_j = pc1_j + pc2_j + pc3_j + pc4_j + pc5_j + pc6_j + breed_j + sex_j + platform_j + logr_i + animal_j$$

where, *animal* is the animal's unique identifier (relationships were included by fitting the A matrix). Trait information was weighted by *weight*, where *weight = reliability of the trait record/(1-reliability)*. *Platform* was fitted as differences in intensity values were identified between the two locations where genotyping was performed. The first 6 principal components of the PCA (pc1-pc6) were fitted to remove any underlying population sub-structure within the dataset (Price *et al* 2006). The heritability (h^2) of each trait was determined in previous analyses and fixed in these analyses to reduce computational load. The Wald F statistic for $logr_i$ was obtained from the asreml output files and the probability of observing the F value by chance calculated. Resultant probabilities were corrected for multiple testing using the Bonferroni method.

RESULTS AND DISCUSSION

Heritability estimates for the intensity only SNPs ranged from 0 to 0.83, with a mean and median of 0.36 and 0.31 respectively. This suggests that at least some of these intensity only SNPs, while non-"normal" in a genotyping sense, may be heritable genetic units. It was also observed that some of these heritable intensity only SNPs have higher logr variability; consistent with being in a CNV region.

We found four of the 1081 intensity only SNPs significantly associated with one or more traits (Table 1). Bonferroni corrected p-values ranged from 4.32×10^{-4} to 3.65×10^{-2}. No SNPs were significantly associated with wwt, lw8 or lw12. Heritability estimates for significant SNPs were moderate (Table 2). Using the Illumina GenomeStudio software to view these SNPs revealed that one of the significant SNPs potentially had more than the three possible clusters, while the remaining significant SNPs had less (Table 2). This explains why these SNPs could not be clustered by GenomeStudio and suggests that copy numbers different to the expected diploid copy number may be present at these genomic loci. However, it is unclear why these would be associated with traits. One possibility is that they are in LD with loci affecting trait variation.

Table 1. Significant associations between intensity only SNPs and traits

SNP	Significantly associated traits	Bonferroni corrected p-value
SNP 1	umd	3.65×10^{-2}
SNP 2	umd	1.82×10^{-2}
SNP 2	ufd	3.03×10^{-2}
SNP 3	cw	1.71×10^{-2}
SNP 3	umd	4.81×10^{-4}
SNP 3	umw	1.34×10^{-3}
SNP 3	ufd	4.32×10^{-4}
SNP 4*	cw	8.12×10^{-3}
SNP 4*	umd	2.56×10^{-3}
SNP 4*	umw	2.75×10^{-3}
SNP 4*	ufd	1.53×10^{-3}

*sex chromosome SNP

Table 2. Heritability estimates for SNPs found to be significantly associated with tested traits. Information on the number of clusters observed in the Illumina GenomeStudio intensity versus theta plots is included

SNP	Heritability estimate (standard error)	Number of clusters observed in Illumina GenomeStudio
SNP 1	0.42 (0.05)	1
SNP 2	0.49 (0.05)	1
SNP 3	0.30 (0.05)	1, possibly 2
SNP 4*	0.35 (0.07)	3, possibly 4

*sex chromosome SNP

Further work includes fitting theta into the model when determining if there is an association between SNP and trait. This will remove any dye effects which could be confounding the data. In addition to this, an independent dataset will be used to validate these results.

CONCLUSIONS

This study has provided a novel means of utilising currently discarded Illumina SNP chip data to identify regions that may be significantly associated with traits of interest. This method could also be applied to all Illumina SNP chip data, including "normal" SNPs, to bypass the clustering and genotyping step and potentially tap into currently uncovered genetic variance. It also has application further afield for use in polyploid species.

ACKNOWLEDGEMENTS

This work was supported by Ovita Ltd, Sheep Improvement Limited and Beef+LambNZ.

REFERENCES

Fernando R.L. and Grossman M. (1990) *Theor. Appl. Genet.* **80**:75.

Gilmour A.R., Gogel B.J., Cullis B.R., and Thompson R. (2008) ASReml User Guide 3.0.

Mrode R.A. and Swanson G.J.T. (2003) *Livest. Prod. Sci.* **86**:253.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. and Reich D. (2006) *Nature Genetics* **38**:904.

Stranger B.E., Forrest M.S., Dunning M., Ingle C.E., Beazley C., Thorne N., Redon R., Bird C.P., de Grassi A., Lee C., Tyler-Smith C., Carter N., Scherer S.W., Tavaré S., Deloukas P., Hurles M.E. and Dermitzakis E.T. (2007) *Science* **315**:848.