The accuracy of genomic selection in predicting carcass traits in meat sheep

A. Slack-Smith[†], B.P. Kinghorn[†] J.M. Hickey[†] and J.H.J. van der Werf[†]

Introduction

Genome-wide association studies (GWAS) for quantitative traits in livestock are primarily focused on genomic selection and the prediction of genomic breeding values (GEBV). Genomic selection is a form of marker-assisted selection in which genetic markers covering the whole genome are used so that most QTL are in linkage disequilibrium with at least one marker (Meuwissen *et al.* 2001). As an alternative to gene discovery, which is genome research focused on mapping and characterising quantitative trait loci (QTL) (Gao *et al.* 2007) genome-wide association analysis allows prediction of breeding value or phenotype for traits of economic importance using all SNP across the whole genome simultaneously (Lee *et al.* 2008). Prediction of breeding vale is relevant for the stud sector, for genome assisted selection of breeding animals, whereas prediction of phenotype can be relevant in production systems, for early allocation of animals into specific cohorts.

Phenotypic prediction of growth and composition aims to sort animals into homogeneous groups to increase uniformity and profitability (Tedeschi *et al.* 2004) and therefore helps to satisfy downstream consumers. Meat quality grade, yield grade and growth performance factors explain much of the variation in profit under grid pricing (Greer and Trapp 2000) and the cost penalties when specifications are missed can be large. Trials conducted by (Cox *et al.* 2006) showed 42% of product did not meet specification in the Australian food service industry. Using phenotypic prediction in the management of animals to increase the proportion that meet specification could increase profitability in sheep meat production by increasing the consistency in meeting consumer requirements. Better allocation of animals to cohorts allows better animal management and can be used to create a relationship between consumers and a product (Walker and Olson 1991) thereby increasing customer retention and satisfaction (Eriksson and Vaghult 2000).

In this study we explore the accuracy of predicting phenotype for carcass and growth traits. We use data collected on animal phenotypes and genotypes based on a 50k SNP chip, using a subset of a data to estimate SNP effects and a remaining test set to evaluate the accuracy of genome based prediction of phenotype. Training and test sets were created either randomly; within sire families or across breeds.

Materials and methods

Phenotypes on growth (post weaning weight, kg, PWWT) and carcass (hot carcass weight, kg, HCWT; cold eye muscle area, $cm^{2,}$ CEMA; intramuscular fat, %, IMF and cold carcass C site fat depth at the 13^{th} rib, mm, CCFAT), fixed effects (gender, birth type, rear type, age, flock, year, breed and season, management group) and genetic data (50K SNP genotypes); were obtained from the Cooperative Research Centre for Sheep Industry Innovations information nucleus.

[†] Cooperative Research Centre for Sheep Research Technologies, Armidale, Australia 2351
[†] School of Environmental and Rural Science, University of New England, Armidale, Australia 2351

1810 animals within a pedigree of 7481 animals contained measured phenotype data and included animals raised as singles (27%), twins (69%) and triplets (4%). Gender included castrate males (65%) and whole females (35%). Post weaning weight age was 95±9 days, average slaughter age was 238±56 days. Breeds included Border Leicester x Merino crosses (20%), Booroola x Border Leicester, purebred Merino, Merino cross (20%), Terminal x Border Leicester x Merino cross (28%), and Terminal breeds (Poll Dorset, Suffolk, White Suffolk and Texel) x Merino crosses (41%;). These animals were raised on Sheep CRC research associated farms in New South Wales, Victoria, South Australia and Western Australia.

Initially a model without SNPs but with polygenic effects was used for variance components and fixed effect analysis in ASReml (Gilmour *et al.* 2002) to determine heritabilities (Table 1) . Significant terms were then modeled jointly with SNP data using GibbsMendelGenome (Hickey and Tier 2009). We used BayesBFast (Fernando 2009 with the priors for proportion of SNPs excluded from the model set at 0.99 for prediction of SNP effects. Cross validation (Kohavi 1995) was performed using training data (subsets which included 75% of the data) and test data (remaining 25%) for 8 replicates where the training sets were chosen from within sire family (75% of the offspring from each sire family) or across breed (Terminal Border Leicester Merino were used to predict Merino x Merino and Border Leicester x Merino.

We calculated correlations between predicted phenotype and actual phenotype in the test data. Predicted phenotype was based either on fixed effects alone (model F), or on fixed effects and SNP effects (model FG). Phenotypes corrected for fixed effects (residuals, model G) were predicted from a model using SNP effects alone.

Results and discussion

Data are summarised in Table 1 for the entire dataset only, however the subsets (Random 25%, within sire 25% and Breed 25%) were very similar with 4 exceptions; average PWWT for the Breed test set was 1.3kg higher (33.6kg), maximum HCWT was lower for the 8 Random reps by 1.7kg (34.3kg), maximum HCWT was lower for the Breed test set 0.8kg (35.2kg) and maximum IMF was lower for the Breed test set by 0.8 (7.6).

Table 1 Mean, standard deviation, minimum and maximum of the dependent variables post weaning weight and carcass traits, n = 1810 animals. Heritability was calculated using ASReml (Gilmour *et al.* 2002).

Trait and heritability	Heritability	Mean	Std Dev	Min	Max
Post weaning weight kg	(0.48±0.05)	32.3	6.8	11.2	54.5
Hot carcass weight (kg)	(0.65±0.06)	22.5	3.3	12.8	36.0
Cold eye muscle area cm	(0.46±0.07)	14.9	2.3	8.3	23.0
Intramuscular fat	(0.47±0.06)	4.3	1.0	1.5	8.4
Cold C site fat (mm)	(0.46±0.08)	3.7	2.0	0.0	10.0

Results in Table 2 show that there can be a substantial increase in the correlation between true phenotype and predicted phenotype when genetic data is incorporated into prediction

models. Across the test sets it was clear that having information on relatives (Sire) increased the correlations and predicting phenotypes across breeds (Breed) reduced phenotypic correlations in weight traits but to a lesser extent in CEMA, IMF and CCFAT. CEMA had the largest increase in correlation between predicted and true phenotype (Table 2), increasing from a correlation of 0.51, 0.51 and 0.52 when BLUP only models (Henderson 1975) are used to a correlation of 0.78 across random data sets, 0.79 within sire families and 0.76 across breeds. Using genomic data in predictions for IMF increased correlations by ca. 0.10 (Rand = 0.51 to 0.61 and Breed = 0.61 to 0.65), CCFAT doubled in predictive capacity however this was still very low, the combined correlation was 0.10.

The GEBVs are expected to explain a higher proportion for the genetic variance than that of the phenotypic variance where sampling error is relatively low, and this can be estimated by dividing the correlations for Model G in Table 2 by the trait heritabilities in Table 1. The estimated adjusted phenotypic variance explained would be 0.5 and 0.5 (PWWT Rand and Sire), 0.30 and 0.50 (HCWT Rand and Sire), 1.5 (CEMA Rand and Sire), 0.5 and 0.4 (IMF Rand and Sire) and 0.3 and 0.5 (CCFAT Rand and Sire). Across breed the residual phenotype explained was generally lower for all traits (HCWT 0.0, IMF 0.2), however CEMA gave similar results to Rand and Sire.

Table 2 Correlation of the predicted phenotype (Phen F (fixed)) and true phenotype; predicted phenotype (Phen FG (fixed and genetic)) and dense 50K SNP data; and the correlation between the residuals of Phen F and predicted phenotype from dense 50K SNP.

Trait	Dependent variable	Model	Random	Sire	Breed
PWWT	Phen	F	0.80	0.81	0.80
	Phen	FG	0.82	0.84	0.81
	Resid	G	0.23	0.26	0.13
HCWT	Phen	F	0.69	0.69	0.55
	Phen	FG	0.71	0.71	0.44
	Resid	G	0.19	0.30	0.03
CEMA	Phen	F	0.51	0.51	0.52
	Phen	FG	0.78	0.79	0.76
	Resid	G	0.64	0.69	0.72
IMF	Phen	F	0.51	0.47	0.39
	Phen	FG	0.61	0.54	0.39
	Resid	G	0.23	0.18	0.08
CCFAT	Phen	F	0.61	0.61	0.05
	Phen	FG	0.61	0.65	0.10
	Resid	G	0.15	0.21	0.20

These results show economically valuable carcass and growth traits on the live animal can be predicted with higher accuracy using dense 50k SNP particularly within a random population and within sire family subsets. Due to the large amount of phenotypic variance explained by

genotypic (SNP data the allocation of animals to market end points could be done with greater accuracy and give potentially substantial economic benefits to the Australian Sheep industry.

References

Cox, R. J., Johnson, S. and Cunial, C. M. (2006). *Australasian Agribusiness Review* Charles Sturt University, 14, 17.

Eriksson, K. and Vaghult, A. L. (2000). *Industrial Marketing Management* 29: 363-372.

Gao, Y., Zhang, R., Hu, X. and Li, N. (2007). *Meat Science* 77: 36-45.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. and Thompson, R. (2002). ASReml User Guide Release 1.0. . VSN International Ltd., Hemel Hempstead, UK.

Greer, H. C. and Trapp, J. N. (2000). Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Chicago, Illinois.

Henderson, C. R. (1975). *Biometrics* 31(2).

Hickey, J. M. and Tier, B. (2009). GibbsMendelGenome (Beta): University of New England, Armidale, Australia.

Kohavi, R. (1995). International Joint Conference on Artificial Intelligence Montreal, Quebec, Canada. 2: 1-7.

Lee, S. H., van Der Werf, J. H. J., Hayes, B. J., M.E., G. and Visscher, P. M. (2008). *PLOS Genetics* 4(10): 1-10.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001). *Genetics* 157: 1819-1829.

Tedeschi, L. O., Fox, D. G. and Guiroy, P. J. (2004). *Agricultural Systems* 79(2): 171-204.

Walker, B. A. and Olson, J. C. (1991). *Journal of Business Research* 22: 111-118.