MULTI-TRAIT GENOME WIDE ASSOCIATION META-ANALYSIS OF BODY WEIGHT, CARCASE COMPOSITION AND EATING QUALITY TRAITS IN AUSTRALIAN SHEEP

N. Moghaddar[1], A.A. Swan[2] and J.H.J. van der Werf[1]

[1] School of Environmental & Rural Science, University of New England, Armidale, NSW 2351, Australia
[2] Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

The objective of this study was to perform a multi-trait meta-analysis of summary statistics of a single-trait genome-wide association study (GWAS) on 10 body weight, carcase composition and eating quality traits in Australian sheep. Meta-analysis was performed based on an approximate chi-squared test according to estimated SNP effects and their associated standard errors obtained in a single-trait GWAS. Single-trait association testing was based on single-marker regression analysis in linear mixed models using imputed whole genome sequence data and between 2,707 and 135,022 adjusted phenotypes across the traits studied. Meta-analysis showed higher power of QTL detection compared to single trait GWAS, it confirmed the highly significant QTL regions in single-trait GWAS and revealed numerous pleiotropic QTLs on chromosomes 1, 3, 6, 8, 11, 16 and 18, affecting two or more traits. In total meta-analysis showed 4,823 SNPs in strong association with at least one trait (*-Log P≥6.0*) but did not show any new highly significant QTL regions across the traits.

INTRODUCTION

Identification and estimation of the genetic parameters of Quantitative Trait Loci (QTLs) is valuable for understanding the biology of traits and is useful to accelerate the rate of genetic gain of economically important traits in plant and livestock breeding programs. Literature shows higher genomic evaluation accuracy and hence faster genetic progress by prioritizing and weighting genetic variants with larger effect in genomic prediction statistical models (e.g. MacLeod *et al*. 2017). Moghaddar *et al*. (2019) showed improvement in prediction accuracy of weight and eating quality traits in a combined dataset from multiple research and commercial sheep populations using information about polymorphisms affecting the genetic variation of the traits.

Identification of QTLs in polygenic traits has been broadly based on single-trait GWAS. However, when multiple correlated phenotypes are available, a joint analysis of multiple traits enabled via meta-analysis, could increase the statistical power of detecting genetic associations. This could be more important for traits with smaller numbers of observations which show weaker associations with the genetic variants in single-trait GWAS (Fang & Pausch, 2019). The objective of this study was to perform a multi-trait GWAS meta-analysis using summary statistics of single-trait GWAS which was performed on two growth trait, four carcase trait and four eating quality traits using imputed whole genome sequence data.

MATERIALS AND METHODS

Studied population and phenotypes. Phenotypic records of two body weight, four carcase composition and four eating quality traits were derived from the Australian national Sheep Genetics database (https://www.sheepgenetics.org.au/). Table 1 shows the names of the traits studied, phenotypic summary statistics and heritability estimates derived from phenotypes and pedigree information. The phenotypes belonged to a multi-breed/admixed sheep population from both

* A joint venture of NSW Department of Primary Industries and the University of New England

research (Information Nucleus Flocks and MLA resource flock) and industry flocks (Sheep Genetics). More than 30 breeds were represented in the data set which was constructed by merging data from three separate Sheep Genetics evaluations, for maternal breeds, terminal sire breeds and Merinos. The dominant breeds represented in the data either as purebreds or crossbred/admixed with other breeds were Border Leicester, Coopworth and maternal composites, White Suffolk, Poll Dorset, Suffolk, Texel, and Merino. The definition and measurement of the traits between research and industry flocks and within maternal breeds, terminal sire breeds and Merinos were based on the same standard.

Table 1. Trait abbreviation and definition, phenotypic summary statistics and heritability estimates for studied traits

Trait	N of Records*	Average	sd	h^2(se)**
PWT (post weaning weight, kg)	92,586	45.44	12.65	0.30 (0.001)
CWT (carcass weight, kg)	20,831	21.889	3.29	0.31 (0.001)
CCFAT (carcass scanned fat, mm)	20,281	3.84	1.96	0.27 (0.001)
CEMD (carcass scanned eye muscle depth, mm)	20,393	31.32	4.23	0.31 (0.001)
DRESS (dressing percentage,%)	15,977	44.34	2.86	0.30 (0.001)
IMF (intra muscular fat, %)	20,320	4.47	1.15	0.55 (0.001)
LMY (lean meat yield, kg)	2,707	56.58	9.69	0.48 (0.02)
PCF (post weaning scanned fat, mm)	51,319	2.84	0.79	0.26 (0.001)
PEMD (post weaning EMD, mm)	51,597	27.15	3.85	0.31 (0.001)
SF5 (shear force at day5 aging, Newton)	20,474	33.45	13.57	0.32 (0.001)

*: number of records with both phenotypes and genotypes, sd: standard deviation, **: heritability (standard error)

Genotypes. The whole genome sequence data on 26 Ovine autosomes which were imputed from a mixture of different low, medium and high-density SNP genotypes were used in this study. In the imputation pipeline, research and industry data with low and medium density SNP genotypes (12k, 15k and 50k) were imputed to a common 60k genotype based on a large reference set. In the next step the 60k genotypes were imputed to high-density genotypes (500k) using a multi-breed reference set of 2,266 animals. Animals with high-density genotypes were then imputed to whole genome sequence using 726 multi-breed animals as a reference set. Genotype phasing and imputation was performed in Beagle 5.3 (Browning *et al*. 2021). The final set of sequence data was comprised of 31,154,249 genetic variants after applying quality control on genotypes and removing genetic variants with low imputation accuracy based on a significant threshold level suggested by software ($r < 0.63$).

Statistical analysis. Phenotypes used in single-trait association studies were obtained as outputs of the multi-trait industry evaluation analyses for Merino, maternal breeds, and terminal breeds respectively run by AGBU (Animal Genetics Breeding Unit). These analyses use phenotypes corrected for known environmental effects such as age and birth/rearing status, then fit a multi-trait mixed model with contemporary group fitted as a fixed effect, and genetic groups, direct and maternal genetic effects, maternal permanent environment, and sire by flock-year fitted as random. The three analyses were run with genetic effects fitted with a pedigree relationship matrix, and pre-corrected phenotypes derived as the sum of the estimated breeding values for direct genetic effects and residual values ($\boldsymbol{y}^* = \hat{\boldsymbol{a}} + \hat{\boldsymbol{e}}$).

Single-trait GWAS was performed by single SNP regression analysis in a linear mixed model using Gemma V0.96 software (Zhou *et al*. 2014) and based on the following equation: $\boldsymbol{y^*=Xb + Zu + e}$. In this equation y^* refers to the pre-corrected phenotypes explained above, $\boldsymbol{b}$ includes a fixed effect modelling the mean of each of the three analyses described above and the SNP effect at each

marker, ***u*** refers to the random additive genetic effect of the animal fitted by genomic relationship matrix (G), and ***e*** is the residual effect. G was calculated according to Yang *et al*. 2011 using 50k genotypes and *X* and *Z* are incidence matrix which relate fixed and random effects to phenotypes.

Meta-analysis was performed based on multi-trait approximate chi-squared test for each SNP, distributed according to chi-square distribution and estimated as $\boldsymbol{\chi^2_{df,n}= t'V^{-1}t}$ (Bolormaa *et al*. 2014). In this equation ***n*** is number of traits, ***t*** is the signed *t*-value derived from single trait GWAS SNP effect and its standard error across all 10 traits (*t=b/se(b)*), and $\boldsymbol{V^{-1}}$ is the inverse of correlation matrix derived from SNP effect (signed t-values between traits). An arbitrary p-value of equal or less than 1.0×10^{-6} was considered as the SNP significance threshold level. Significant SNPs were pruned for high LD in Plink (Purcell *et al.*2007) according to window size of 5000 SNPs, sliding window of 200 SNPs and LD≥0.95).

RESULTS AND DISCUSSION

Figure 1 shows the results of the multi-trait meta-analysis as a plot of SNP p-values versus chromosomal position. The significant regions in Figure 1 are those which are significant for at least one trait. Compared to p-values derived in a single-trait GWAS and meta-analysis, meta-analysis showed higher power of QTL detection which was in line with results of comparison of single-trait GWAS and a meta-analysis in cattle (Bolormaa et al. 2014). Meta-analysis confirmed the highly significant regions in single trait GWAS, and furthermore showed stronger association for some regions that had weaker association with phenotype. However, there were regions around significant thresholds with weak association with phenotypes in single-trait GWAS which were not significant in the meta-analysis, particularly for traits with smaller number of phenotypes. Numerous highly significant pleiotropic QTL region were found across the studied traits, including regions on chromosome 1 (CCFAT, PCF, IMF and PEMD), chromosome 2 (CWT, CWT, CCFAT and IMF), chromosome 3 (affecting both weight traits), chromosome 6 (weights and EQ traits except SF5 and CEMD), chromosome 8 (PWT, PCF and PEMD), chromosome 11 (PWT, CWT, PCF and PEMD), chromosome 16 (weights and PCF traits) and chromosome 18 (PWT, IMF, CEMD and SF5). Meta-analysis did not reveal new significant QTL regions across these traits compared to single-trait GWAS on sequence data. In total, meta-analysis discovered 4,823 SNPs that were in significant association with at least one trait *(-log p≥6.0*) after pruning for high LD.

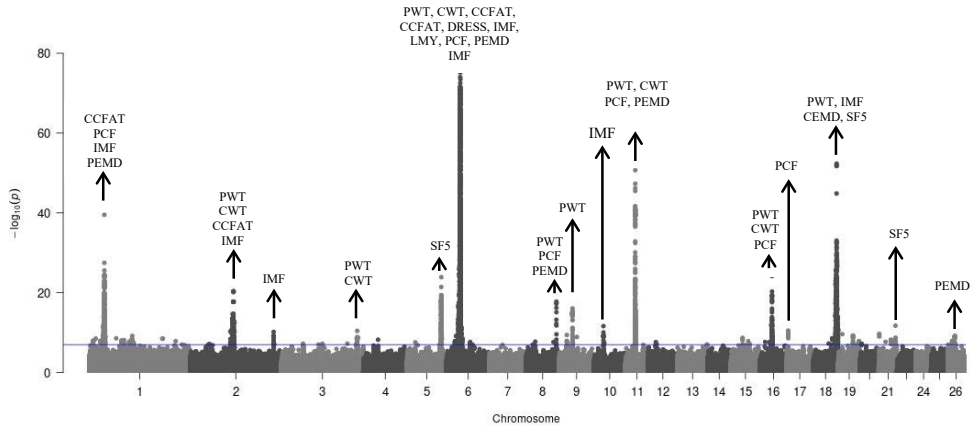

Figure 1. Manhattan plot of results of multi-trait GWAS meta-analysis of weight, carcase composition and eating quality traits

Comparison of genetic correlations between traits based on signed t-values derived from p-values, and the genetic correlation obtained based on pedigree information showed a similar correlation direction. However, the strength of correlation coefficients was different in some cases, most particularly for traits with smaller number of records. LMY had the smallest number of phenotypes and genotypes in the meta-analysis which showed the most different correlation estimated in meta-analysis in comparison to the genetic correlation estimated from pedigree.

Meta-analysis uses information in summary statistics of the results of single-trait GWAS on genetically related traits to improve the power of QTL detection, and together with single-trait GWAS is useful to confirm pleiotropic QTL effects. In this study meta-analysis showed notably stronger evidence of QTL affecting the traits with moderate to high genetic correlations. Meta-analysis was also useful to flag those regions which were close to the significance threshold in single-trait GWAS. Meta-analysis did not show new highly significant regions. This could be related to a high resolution in the single-trait GWAS due to strong linkage disequilibrium provided by using whole genome sequence data.

CONCLUSIONS

Multi-trait meta-analysis of weight and eating quality traits using SNP effects derived from single-trait GWAS on imputed whole genome sequence data showed higher power of detecting genetic variants in significant association with phenotypes compared to single-trait GWAS. Meta-analysis was also useful to flag those genetic markers which were close to the significance threshold in single-trait. GWAS Results of meta multi-trait analysis and single-trait GWAS revealed numerous pleiotropic QTL regions affecting two or more traits in this study. In total, meta-analysis showed 4,206 genetic variants in significant association with at least one trait (-log p>7.0), however, it did not show new highly significant QTL compared to results of single-trait GWAS on whole genome sequence data.

ACKNOWLEDGEMENTS

This research was partly funded by Meat Livestock Australia project L.GEN.1815. The authors thank Sheep Genetics and contributing breeders for access to the data used in this study.

REFERENCES

Bolormaa S., Pryce J.E., Reverter A., Zhang Y., *et al*. (2014) *PLoS Genet*.

Browning B.L., Tian X., Zhou Y. and Browning S.R. (2021) *Am. J. Hum. Genet.* **108:** 1880.

Fang Zih-Hua. and Pausch H. (2019) *BMC Genomics* **20:** 695.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S., *et al.* (2015) *Hemel Hempstead: VSN International Ltd*.

MacLeod I.M., Bowman P.J., Vander Jagt C.J., Haile-Mariam M., *et al* (2017) *BMC Genomics* **17:** 144.

Moghaddar N., Khansefid M., Van der Werf J.H.J., Bolormaa S., *et al*. (2019) *Genet. Sel. Evol*. **51:** 72.

Purcell S., Neale B., Todd-Brown K., Thomas L., *et al.* (2007) *Am. J. Hum. Genet* **81**.

Zhou X and Stephens M., (2012) *Nature Genetics* **44:** 821,

Yang J., Benyamin B., McEvoy B.P., *et al*. (2010) *Nature Genetics* **42:** 565.