PLEIOTROPIC MULTI-TRAIT GENOME-WIDE ASSOCIATION REVEALS PUTATIVE CANDIDATE GENES FOR FATTY ACID COMPOSITION IN AUSTRALIAN SHEEP

S. Bolormaa[1,2], J.H.J. van der Werf[2,3], B.J. Hayes[1,2,4], M.E. Goddard[1,5], and H.D. Daetwyler[1,2,4]

[1]Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia
[2]CRC for Sheep Industry Innovation, Armidale, Australia
[3]University of New England, Armidale, Australia
[4]La Trobe University, Bundoora, Australia
[5]University of Melbourne, Parkville, Australia

SUMMARY

The aims of this paper were to detect SNP and genes that affect the fatty acid composition and other carcass and meat quality traits in sheep. We performed genome-wide association studies (GWAS) for 56 traits including carcass weights, muscling, fatness, tenderness, meat color, mineral content, and fatty acid composition on 10,613 animals genotyped for 510,174 SNPs. The use of a meta-analysis to combine information from the 56 traits increased the power to detect QTL compared with the single trait analyses. We found pleiotropic QTL, which appear to cluster into 5 functional groups based on their trait effects. Candidate genes were identified in the groups that have functions consistent with the biology of the traits.

INTRODUCTION

The fatty acid (FA) composition (FAC) of meat products has received considerable attention for its significance to human health. The FAC of meat is influenced by various environmental effects such as diet (e.g. Suzuki *et al*. 2007), the level of fatness, and genetic factors. Changing FAC through selection could decrease the saturated FA (SFA) content of meat and perhaps improve human health. Breed or genotype differences in the FAC have been reported, even after correction for fat level (De Smet *et al*. 2004). FAC can be described by a combination of traits, but GWAS are usually performed one trait at a time, which may reduce the power to detect mutations that affect the trait complex. In fact, mutations that affect FAC may also affect other quality and carcass traits, therefore a multi-trait GWAS analysis is expected to be more powerful to assess the effect of a mutation on multiple related traits. Multi-trait analysis of linkage experiments has been reported to increase the power to detect QTL (e.g. Knot and Haley. 2000). Bolormaa *et al*. (2014) found increased power to detect QTL by a meta-analysis that combined the results from GWAS for 32 individual traits in beef cattle. The objectives of this study were to identify genes that affect FAC and other meat quality and carcass traits using a multi-trait GWAS in Australian sheep.

MATERIALS AND METHODS

SNP and phenotype data. In total, 510,174 single nucleotide polymorphisms (SNP) were genotyped or imputed from lower density SNPs. The SNP genotype and phenotype data was collected by the CRC for Sheep Industry Innovation (Sheep CRC) and SheepGENOMICS projects. The SNP were obtained from two different SNP arrays: the Illumina 600k (HD) and 50k Ovine SNP chips (Illumina Inc., San Diego, CA, USA). All SNP were mapped to the OAR 3.1 build of the ovine genome sequence assembled by the SNPchiMp v.3 (Nicolazzi *et al*. 2015). Details on genotyping, editing and imputation of the Sheep CRC 50k data set has been described by Daetwyler *et al*. (2012). The imputation from 50k to HD was done using Fimpute (Sargolzaei *et al*. 2014). Out of 22,684 animals with 50k and 1,735 HD genotypes, the accuracy of imputation between 50k and HD genotypes for non-50k SNPs was 0.9871. A total of 10,613 sheep (from multiple breeds and crosses) HD genotyped were measured for up to 56 meat related traits (carcass

weight, fatness, muscling, tenderness, meat colour, pH level, and FAC). A complete description of the design, methods and analyses of carcass and meat quality assessments is given by Mortimer *et al*. (2014).

Statistical analysis. A GWAS was carried out for each trait using ASReml software (Gilmour. 2009) to fit a mixed model: trait ~ mean + fixed effects + SNP_i + animal + dam + sire by flock interaction + error; with animal, dam (permanent environment), and sire by flock interaction, and error fitted as random effects including relationships between animals. All models included dataset, management group, flock, date of observation, drop year, sex, birth type, and rear type as fixed effects. The FA traits were corrected for intramuscular fat content. The individual trait results were combined using the meta-analysis described by Bolormaa *et al*. (2014). To avoid identifying a large number of closely linked SNPs, whose association with traits is due to the same QTL, only the most significant SNP ($P<10^{-5}$) from each 1Mbp interval was retained for validation. In a further restriction, a maximum of two or three SNPs on the same chromosome were selected, and only if they clearly represented different QTL based on the multi-trait test. In this way, 23 "Lead SNP" representing different QTL across the genome were selected. For each lead SNP, a linear index of the 56 traits, that was maximally correlated with the Lead SNP genotypes in the reference population (4/5 of total animals) was constructed and used to validate SNP effect in the other 1/5 of the data, as described by Bolormaa *et al*. (2014). Clustering analysis based on a correlation matrix among the 23 Lead SNP genotypes was used to assign the lead SNPs into 5 groups such that the SNPs within a group had a similar pattern of effects across the 56 traits. To find additional SNPs with a similar pattern of effects to each of the 23 lead SNPs, the linear index for the corresponding Lead SNP was used as a new "phenotype" in GWAS (which fitted the 23 Lead SNP themselves as fixed effects). Genes that occurred within 30 kb of the significant SNPs were annotated using UCSC Genome Bioinformatics (genome.ucsc.edu) and Ensembl (www.ensembl.org/biomart/). The statistical enrichment of the biological function or pathways in groups of genes were checked using the STRING program (Franceschini *et al*. 2013).

RESULTS AND DISCUSSION

Using the multi-trait analysis, 586 SNPs were significant ($P < 5\times10^{-7}$), corresponding to a false discovery rate of 0.04%, and this was better than for any individual trait. When traits were analyzed individually, for only 10 out 56 traits the FDR was less than 2.5%. Many highly significant SNPs from the multi-trait analyses were found within narrow regions on Ovine autosomal chromosomes (OAR) 2, 3, 5, 6, 11, 12, 14, 18, 20, and 26. Many of the significant SNPs in both single trait and multi-trait analyses were closely linked and could be associated with the same QTL. When only the most significant SNPs in each Mb interval were retained from the multi-trait analysis of the discovery dataset for validation purpose, 98 SNPs were significant at $P < 10^{-5}$. In the validation population, all 98 SNPs had an effect in the same direction as in the discovery population and 35 were significant ($P < 0.05$). One of the best single-traits was DRESSING% where 9 of 31 significant ($P < 10^{-5}$) SNPs found in reference population were significant at $P < 0.05$ in the validation population. Therefore the multi-trait analysis detected associations with higher reliability than single-trait analysis.

The multi-trait analysis was particularly successful in detecting pleiotropic QTL. Genes that operate in the same pathway might be expected to show the same pattern of pleiotropic effects. The cluster analysis of the 23 Lead SNP revealed 5 loosely-defined groups of SNPs where the SNPs within a group had a similar pattern of effects across traits: 1) changing skeletal or carcass size; 2) changing fat deposition (SFA profile and fatness); 3) influencing meat eating quality; 4) affecting meat colour; 5) a poorly defined group affecting only specific trait(s) (e.g. changing only glycogen level, omega-3 or omega 6 FA). The GWAS using Lead SNP linear indices identified 687 significant SNPs ($P < 5\times10^{-7}$) assigned to each of the 5 groups as follows: 1) 518; 2) 30; 3) 99,

4) 25; and 5) 15.
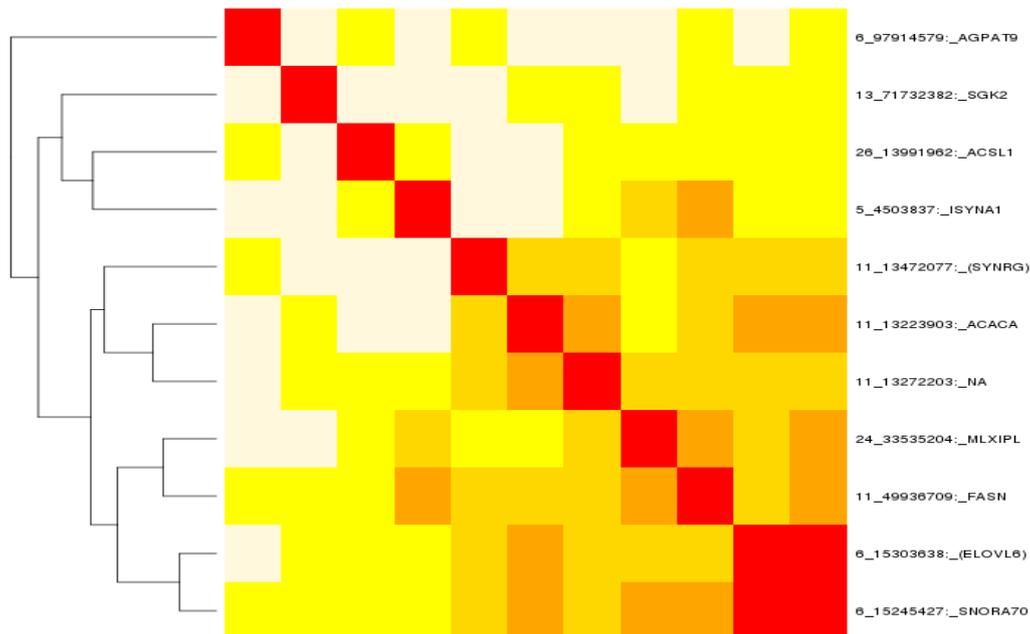
Figure 1. Correlation matrix between the effects of lead SNPs and their linear index SNPs (chromosome_position_annotated gene name) within Group 2.

Group 2 initially consisted of four strongly correlated Lead SNP on OAR 6 (OAR6_15.2Mb), 11 (OAR11_13.3Mb and OAR11_49.9Mb), and 26 (OAR26_13.9Mb). These 4 SNPs have alleles that increase the concentration of SFAs with carbon chain of C16, C14, C12, and C10 (palmitic, myristic, lauric, and capric acids, respectively) and may decrease stearic acid profile (C18:0) (Table 1). There was also a tendency for SNP alleles that increased saturated FAC to increase fatness. If each QTL group represented a particular physiological pathway, one might expect the genes that map near the QTL of a group to show some similarity of function. The 4 Lead SNPs in Group 2 were expanded with 7 additional SNP from the linear index GWAS and the clustering was repeated within group (Figure 1). These additional 7 SNPs were chosen to represent different QTL or as multiple candidates for the same QTL where more than one gene was a suitable candidate. The 11 SNPs were clustered into two subgroups (Figure 1) and they map near to or within genes affecting FA traits. Sub-group 1 consists of SNPs whose effects were moderately to highly correlated ($r > 0.4$) and tagged genes such as *FASN*, *MLXIPL*, *EVOLV6*, *ACACA*, and *SYNRG*. The other more loosely correlated sub-group included the *ACSL1*, *ISYNA1*, *SGK2*, and *AGPAT9* genes (Figure 1). The former group includes genes which play a role in the synthesis of fatty acid formation and the latter includes genes with a role in (glycero)lipid biosynthesis. The *MLXIPL* protein activates carbohydrate response element motifs in the promoters of triglyceride synthesis genes. Studies in gene expression (e.g. De Jager *et al*. 2013) have reported that *ACACA*, *FASN,* and *DGAT2* may directly influence IMF percentage in sheep and cattle. Dervishi *et al*. (2011) noted that changes in FA profile due to feeding systems implicate changes in the mRNA expression level of genes related with fat metabolism. Thus, group 2 does represent a set of genes with known function in fatty acid and lipid synthesis. This was confirmed by KEGG and GO enrichment analysis. There were some enriched interactions at medium confidence level ($P = 6.9 \times 10^{-6}$) between some Group 2 genes. GO and KEGG analysis found some evidence for

statistically significantly enriched in the biological function or pathways in our FA genes. For example, according to KEGG and GO terms, two proteins (FASN and ACACA) were involved together in FA biosynthesis (Bonferroni $P = 3.8\times10^{-4}$), 5 proteins (FASN, ACACA, ACSL1, AGPAT9, and ISYNA1) in metabolic pathways (Bonferroni $P = 7.6\times10^{-3}$), and 4 proteins (FASN, ELOVL6, ACSL1, and ISYNA1) in lipid biosynthesis process (Bonferroni $P = 3.8\times10^{-2}$).

Although the SNPs within a group share some features they may also be associated with seemingly quite different traits. Two SNPs from Group 3 (near to genes *MRPS25* and *CYP27A1*), one SNP from Group 4 (near *GLTPD1*) and two from Group 5 (within genes *PNPLA3* and *FADS2*) affected the concentration of polyunsaturated FA (Table 1).

Table 1. Effect of a subset of significant multi-trait SNPs in the individual fatty acid traits (signed t-values >1 are shown)

			Fatty acid with chain of										
#[1]	OAR[2]	POS[3]	C22: 6n-3	C22: 5n-3	C20: 5n-3	C20: 4n-6	C20: 3n-6	C18: 2n-6	C18: 0	C16: 0	C14: 0	C12: 0	C10: 0
2	26	13.9	2.8			1.5	-1.2	2.7				-4.5	-3.3
2	11	13.2	1.0		1.1	1.3	2.6		-2.5	3.6	4.9	2.2	7.2
2	11	49.9							-1.6	4.9	8.4	3.3	3.8
2	6	15.2	-1.8			-1.4	-1.2		-1.8	6.0	3.7	3.0	3.0
3	19	57.1	2.4	2.8	2.4	1.9	2.2		-1.0				-2.1
3	2	219	3.6	1.0	4.1	5.4	2.8	3.7		-1.6	-2.9	-3.3	-3.0
4	12	49.6	-6.8	-5.8	-8.2	-4.6	-5.7	-6.1	1.4				
5	3	21.8		3.5	3.3	-1.6	2.4			-2.0		2.9	1.2
5	21	39.7	1.5			3.1	8.2	-1.8					

[1]Group; [2]Chromosome; [3]Ovine chromosome position in megabases.

The pattern of effects of each QTL studied here indicated that some may be more useful for selection than others depending on the breeding goal. Some QTL have an allele with desirable effects on more than one trait and appear to be good targets for selection. For instance, the QTL on OAR 2 had allele that increases tenderness, improves meat colour, increases myoglobin, glycogen, and omega-FAs and decreases long chain saturated FAs, which is a highly valuable pattern. Selection for this allele would be beneficial in sheep intended for most markets because lamb prices reflect colour, tenderness, palatability, and juiciness.

ACKNOWLEDGEMENTS

We thank Klint Gore and Ken Geenty for managing the CRC information nucleus database and the many staff involved at the CRC and SG sites across Australia.

REFERENCES
Bolormaa S., Pryce J.E., Reverter A., Zhang Y., et al. (2014) *PLoS Genet.* **10**:e1004198
Daetwyler H.D., Swan A., van der Werf J.H.J., Hayes, B.J. (2012) *Gen. Sel. Evol.* **44**:33.
De Jager N., Hudson N.J., Reverter A., Barnard R., *et al.* (2013) *J. Anim Sci.* **91**:1112–1128.
De Smet S., Raes K., Demeyer D. (2004) *Anim. Res.* **53**:81.
Franceschini A., Szklarczyk D., Frankild S., *et al.* (2013) *Nucl. Acids Res.* **41**:D808-D815.
Knott S. Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. (2009) HP1 1ES, UK.
Knot A. and Haley C. S. (2000) *Genetics.* **156**:899.
Mortimera S.I., van der Werf J.H.J., Jacobd R.H., *et al.* (2014) *Meat Sci.* **96**:1016.
Nicolazzi N.L., Caprera A., Nazzicari N., Cozzi P., *et al.* (2015) *BMC Genomics* **16**:283.
Sargolzaei M., Chesnais J., Schenkel F. (2014) *BMC Genomics.* **15**:478.
Suzuki S., Ishikawa S., Arihara K., Itoh M. (2007) *J. Anim. Sci.* **78**:293.