Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep

Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep

D. Garza Hernandez [1,a], S. Mucha[1,4], G. Banos[1,2], K. Kaseja[1], K. Moore[1], N. Lambe[1], J. Yates[3], L. Bunger[1]

[1]*Animal and Veterinary Sciences, Scotland's Rural College, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom*

[2]*Roslin Institute, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom*

[3]*British Texel Sheep Society, National Agricultural Centre, Stoneleigh Park, Kenilworth, Warwickshire, CV8 2LG*

[4]*Poznan University of Life Sciences, Department of Genetics and Animal Breeding, Wolynska 33, 60-637 Poznan, Poland*

[a]*Present address: Universidad Autónoma de Nuevo León (UANL), Pedro de Alba S/N, Ciudad Universitaria, San Nicolás de los Garza, N.L. México 66451.*

Corresponding author: Lutz Bunger Email: Lutz.Bunger@sruc.ac.uk

Abstract

Sheep are an important part of the global agricultural economy. Growth and meat production traits are significant economic traits in sheep. The Texel breed is the most popular terminal sire breed in the UK, mainly selected for muscle growth and lean carcasses. This is a study based on a genome-wide association approach that investigates the links between some economically important traits, including Computed Tomography (CT) measurements, and molecular polymorphisms in UK Texel sheep. Our main aim was to identify Single Nucleotide Polymorphisms (SNP) associated with growth, carcass, health and welfare traits of the Texel sheep breed. This study used data from 384 Texel rams. Data comprised 10 traits, including 2 CT measured traits. The phenotypic data were placed in four categories: growth traits, carcass traits, health traits and welfare traits. De-regressed estimated breeding values (EBV) for these traits together with sire genotypes derived with the Ovine 50K SNP array of Illumina were jointly analysed in a genome wide association analysis. Eight novel chromosome-wise significant associations were found for carcass, growth, health and welfare traits. Three significant markers were intronic variants and the remainder intergenic variants. This study is a first step to search for genomic regions controlling CT based productivity traits related to body and carcass composition in a terminal sire sheep breed using a 50K SNP genome-wide array. Results are important for the further development of strategies to identify causal variants associated with CT measures and other commercial traits in sheep. Independent studies are needed to confirm these results and identify candidate genes for the studied traits.

Keywords: Sheep, Texel, CT, Associated, GWAS.

Implications

Sheep are an important part of the global agricultural economy. To the best of our knowledge GWAS for CT based productivity traits, for a UK terminal sire breed, has not been widely researched. The main aim of this work was to exploit improved genotypic tools, specifically the Illumina OvineSNP50 chip, allowing a simultaneous genotyping for up to 54,241 SNPs to identify those SNPs associated with growth, carcass composition, health and welfare traits of Texel sheep using de-regressed estimated breeding values of rams.

Introduction

Sheep are an important part of the global agricultural economy. They are particularly well adapted to convert short herbage to meat, milk and wool and they are very important to meet global needs for food security for an increasing population around the world (Hopkins and Lobley, 2009).

Currently the Texel breed is the most popular terminal sire breed in the UK accounting for 30% of all purebred rams used for crosses to maternal sheep breeds (Pollott, 2014) and is mainly selected for muscle growth and lean carcasses (Hopkins and Lobley, 2009).

There are only a few methods to predict body composition in live sheep. Over the last few decades mainly ultrasound technologies had been used on farm animals for evaluation of carcass composition (Silva, 2016). However, computed tomography (CT), a non-invasive imaging technology, can accurately measure carcass traits *in vivo* such as muscle and fat (Bünger *et al.*, 2011), muscularity (Jones *et al.*, 2002) and tissue weights (Macfarlane *et al.*, 2006). Additionally, it has been evidenced the potential of CT scanning to improve eating quality and tissue distribution of sheep meats (Macfarlane *et al.*, 2009). As CT scanning is however more expensive than ultrasound, a two-step-procedure is recommended. Only the best 15-20% of selection candidate ram lambs measured by ultrasound would be subsequently CT scanned (Lewis, 2004).

Sheep genetics studies

Breeders focus sheep selection on production traits, including carcass composition and growth traits but also integrate other traits such as meat quality, disease resistance, lambing ease and survival (Bünger *et al.*, 2011). According to the animal QTL database

there are currently (06/2017) 1,515 sheep QTLs curated in the animal QTL database (Hu *et al.*, 2013) representing 222 different sheep traits, reported in 126 publications. However, one of the main limitations of unscrambling the genetic architecture underlying production traits in sheep has been the relative lack of information on the sheep genome in addition to the lack of accurate phenotypic data obtained (Zhang *et al.*, 2013).

Currently, knowledge of the major genes or QTL associated with carcass composition and growth traits in sheep is limited (Zhang *et al.*, 2013). Walling *et al.* (2004) pioneered the first accounts of QTL studies for growth and carcass conformation traits in domesticated sheep covering several genomic regions, which led to characterization of the Texel muscling QTL (TM-QTL).

With the advent of genome-wide panels of single nucleotide polymorphisms (SNPs) and using the approach of a genome-wide association study (GWAS), it has become possible to identify and localize QTLs for complex traits in many livestock species (Georges, 2007). However, to date, only a small number of GWASs in sheep have been conducted because of either limited information available for the sheep genome and funding. These studies have been mainly focused on sheep growth, ultrasound-measured meat traits and body composition traits (Cavanagh *et al.*, 2010, Zhang *et al.*, 2013, Bolormaa *et al.*, 2016, Matika *et al.*, 2016)

Moreover, GWAS with high accuracy CT measured body composition traits are still very rare in the literature. Donaldson *et a*l. (2014) used spine characteristics measured from X-ray computed tomography (CT) scans in order to investigate if there were any subsequent associations between TM-QTL inheritance and underlying spine characteristics (Donaldson *et al.*, 2014). Also, Cavanagh *et al.* (2010) performed a QTL

mapping study in sheep based on in vivo obtained CT data providing predictions for 13 traits describing major fat depots, lean muscle, bone, body proportions and body weight; they identified 3 highly significant, 15 significant, and 11 suggestive QTL on eleven chromosomes. But, no tissue-specific QTL were identified. Furthermore, Matika *et al.* (2016) conducted recently a genome-wide association study (GWAS) for carcass composition phenotypes, including bone, fat and muscle components, which were captured using CT. The GWAS analyses revealed multiple SNPs and quantitative trait loci (QTL) that were associated with effects on carcass composition traits and were significant at the genome-wide level.

In this study we performed a genome wide association study to identify those SNPs associated with growth, carcass composition, health and welfare traits, including 2 CT measured phenotypes, of Texel sheep using de-regressed EBVs of rams.

Material and Methods

Traits and phenotypes

A total of 384 Texel rams descended from 252 sires and 351 dams were analysed for 10 productivity traits including 2 CT measured traits. These rams represent a group of well-monitored animals as only a proportion (10-20%) of the initial selection candidates will be put forward to CT scanning based on ultrasound results.

The phenotypic data were provided by the Signet Sheep breeder Service and comprised EBVs progeny test derived for: birth weight (BW), eight week body weight (EWW) and scan weight (SW), which is the live weight at US scanning at about 21 weeks of age. These were considered as growth traits. As carcass traits were used US measured fat depth (FD) and muscle depth (MD) which are obtained by US-scanning at the at the third lumbar vertebra at 90 degrees to the backbone. The CT measured carcass traits: fat weight (FW), CT lean weight (LW) and the muscularity score (MU), a measure of carcass shape (Bünger *et al.*, 2011), were also included. Details on the CT measured traits have been reported earlier (Bünger *et al.*, 2011). Faecal egg count (FEC) as a measure of worm egg count in sample from lambs at 21 weeks of age, and, Lambing ease (LE) as a direct assessment of the ease with which ram progeny will be born.

GWAS accuracy can also be affected by systematic environmental effects. De-regressed EBVs are an alternative to raw phenotypic measurements, because they represent aggregate phenotypes adjusted for systematic environmental effect. The phenotypic data used therefore consisted of de-regressed estimated breeding values (EBVs) of standard commercial traits.

Statistical model for de-regressed breeding values

The official Texel EBVs were used, those breeding values were derived from the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e},$$

where **y** is the vector of phenotypic observations for one of the analysed traits, **b** is the vector of fixed effects with design matrix X (relating observations to fixed effects), which varied depending on the trait, **a** is the vector of random animal effects, with design matrix Z (relating observations to random effects) and **e** is the vector of random residuals. The list of effects is summarized in Supplementary Table S1.

Random effects are assumed to be normally distributed with zero means and the following covariance structure:

$$Var\begin{bmatrix}\mathbf{a}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_a^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **A** is the pedigree-based relationship matrix, σ_a^2 is the genetic variance, and σ_e^2 is the residual variance.

The software package MIX99 was used for de-regression (Lidauer M, 2011), using a full animal pedigree with effective offspring contributions (EOC) as weighting factors. The de-regression procedure was based on the method published by Jairath *et al.* (1998), involving solving the mixed model equations with a full pedigree to obtain the right-hand side or de-regressed EBVs. Thus DRPs represent daughters averages adjusted for fixed effects and contributions from parents and relatives in the pedigree (Jairath *et al.*, 1998).

153

154 EOC were calculated as:

155
$$EOC_i = \frac{rel_i \cdot kdau}{1 - rel_i}$$
$$kdau = \frac{4 - h^2}{h^2}$$

156 where rel_i is the reliability of EBV for animal *i* and h^2 is the heritability of one of the
157 analysed traits.

158 The use of effective daughter or progeny contribution as a weighting factor is used to
159 avoid biases in sire variances (Fikse and Banos, 2001). The EOC provides a measure
160 of the precision of the daughter information used to compute the de-regressed EBV of
161 the animal as the estimates of reliability used in the computation accounts for factors
162 such as contemporary group (CG) structure for the ram's daughters, the correlation
163 between observations on the same daughter and the reliability of the performance of
164 the daughters' dams.

165 A Shapiro and Wilk's W-statistic test, conducted using the R-package (R Core Team,
166 2013) was used to test data distribution for normality (Royston, 1995). Traits not
167 normally distributed were rank transformed to a normal distribution for their use in
168 subsequent analysis. This rank-transformation method has been reported to give a
169 consistent performance in identifying causal polymorphisms with a slight increase in
170 false positive rate (Goh *et al.*, 2009). This method was used because according to Goh
171 *et al.* (2009) for small sample size or genetic effects, the improvement in sensitivity for
172 rank transformation outweighs the slight increase in false positive rate.

173 *Genotyping*

All rams were genotyped with the ovine 50k SNP chip (54,241 SNPs across the genome with an average of 20.4 SNPs per Mb) by AgResearch. The order of the SNPs was based on the Ovis_aries_4.0 assembly released by the International Sheep Genomics Consortium (Jiang et al., 2014).

Quality control (QC) was performed with the GenABEL R package by considering genotypes of all rams (Aulchenko *et al.*, 2007). The QC excluded 1,564 SNPs with call rates lower than 95%, 3,891 SNPs with minor allele frequencies less than 1%, 98 X-linked SNPs that were likely to be autosomal (cut off odds > 1000) and 777 SNPs not in Hardy-Weinberg equilibrium (p-value $<1x10e^{-5}$). The call rate per individual was always higher than 90% so no animal was removed from the analysis. After applying these quality control criteria 48,433 SNPs (89%) located on 26 autosomes and on the X chromosome were used in the subsequent analyses.

Statistical Model for GWAS

A Multidimensional Scaling Analysis (MDS) was performed first to evaluate the genetic structure of the population. For each trait, SNP effects were then tested, by a single marker regression, with a mixed animal model including the genomic kinship matrix (identity by state) between the genotyped animals, adjusted for allele frequencies. Kinship was computed based on the method proposed by Astle and Balding (2009), using GenABEL, to control for population structure or polygenic effect (Astle and Balding, 2009). The following model was used:

$$\mathbf{y}=\mathbf{X\beta}+\mathbf{Zu}+\mathbf{e}$$

where **y** is the vector of de-regressed EBV of rams, **β** is a vector of coefficients for the SNP effects, **u** is the vector of random animal effects, **e** is the vector of random residual effects, and **X** and **Z** are incidence matrices relating observations to fixed and random

198 animal effects, respectively. Random animal effect followed a normal distribution
199 MVN(0, $\mathbf{G}\sigma^2_u$) where **G** is the genomic kinship matrix and σ^2_u is the polygenic variance;
200 and the random residual effects of the model was assumed to be MVN(0, $\mathbf{I}\sigma_e^2$), where
201 σ_e^2 is the residual variance and I is an identity matrix. Each trait was analysed
202 separately and all analyses were run with GenABEL.
203 This procedure consisted of two steps: firstly it estimated the polygenic and residual
204 variance, not accounting for marker effects and fitting the genomic kinship matrix in the
205 model. Secondly, these estimated variance components were used to estimate all the
206 marker effects (fitting in the model the genotypes and the previously estimated
207 residuals). The j-th marker was fitted in the single-marker-based linear mixed model
208 without removing the j-th marker from the G matrix. Evidence has shown analytically
209 that, if variance components are kept constant, the estimation of the regression of
210 phenotype on *m* markers is invariant with respect to whether or not the marker(s) tested
211 for association is(are) included when constructing the **G matrix** (Gianola *et al.*, 2016).
212 Significance of the results was tested at genome-wise and chromosome-wise levels,
213 including a strict Bonferroni correction for multiple-testing, corresponding to 1x10−6 and
214 3.5x10-5, respectively.
215 In order to address possible population stratification problems, the inflation in the test
216 statistic was monitored with factor lambda, which does not depend on allele
217 frequencies (Aulchenko *et al.*, 2007). The allele effects estimated by GenABEL refer to
218 the least frequent allele in the population and are expressed in trait phenotypic
219 standard deviation (STD) units. Genes located on or around the identified SNPs were
220 examined using the ENSEMBL database and the Ovis_aries_3.1 and 4.0 assembly
221 released by the International Sheep Genomics Consortium (Jiang *et al.*, 2014). And

finally JBrowse was used to identify previously associated QTLs in the tagged regions (Skinner *et al.*, 2009).

Results

Descriptive statistics

For the 10 analysed traits (de-regressed EBVs) the means and standard deviations are shown in Table 1. The normal distributions of the 10 traits were tested with the Shapiro-Wilk's test (Table 1). For EWW, FD, FW, FEC and LE traits the null hypothesis of following a normal distribution was rejected according to a p value ≤ 0.1, which has been previously suggested as an acceptable threshold for this type of analysis (Royston, 1995). These records were rank-transformed to a normal distribution for their use in the subsequent analyses.

Genome Wide Association Analysis

A multidimensional scaling analysis using the GenABEL package showed that no genetic stratification was present in this population. Also, the average inflation factor (λ) was 1.008 ± 0.007, with a maximum value of 1.021 for FEC and a minimum of 1 for FD, FW and MU. Therefore, the population structure is not expected to affect the results of GWAS in the present study.

No genome-wise significant associations were found between any SNP and trait. However, 8 chromosome-wise significant SNPs were found for EWW, FD, MD, LW, FEC, and LE (Figure 1). These SNPs were located on chromosomes 3, 4, 6, 11, 16 and 17, respectively (Table 2). None of the associated SNPs found had been previously associated with any trait in sheep.

The proportion of total variance explained by each SNP was obtained by first scanning using the score test and then revaluating best hits, individually, using Maximum

Likelihood with significant SNP allelic effect fitted as covariate. The variance explained for chromosome wise significant SNP associated with EWW, FD, LW, MD and FEC were 0.029, 061, 0.062, 0.060 and 0.051, respectively. And for LE, each significant marker explained a variance of 0.006, 0.038 and 0.013.

Discussion

Until very recently, limited information on the sheep genome and lack of phenotypic data for many important traits have resulted in only a few studies on SNPs associated with production and welfare traits in sheep (Zhang *et al.*, 2013). It has been suggested that the use of more precise phenotypes derived from CT measures will lead to more accurate phenotypes for genetic analyses (Cavanagh *et al.*, 2010).

To date, only a small number of GWAS in sheep have been conducted, those have been mainly focused on sheep growth, ultrasound-measured meat traits and body composition traits (Cavanagh *et al.*, 2010, Zhang *et al.*, 2013, Bolormaa *et al.*, 2016, Matika *et al.*, 2016). Moreover, genetic analyses with high accuracy CT-measured body composition traits are still very rare in the literature (Walling *et al.*, 2004, Donaldson *et al.*, 2014, Bolormaa *et al.*, 2016, Matika *et al.*, 2016).

The main aim of the present study was to identify SNPs associated with traits currently in the selection index for a UK Terminal sire breed (Texel Sheep), including CT based productivity traits. In the UK, CT scanning has been used in sheep breeding programs since 2000. However, as CT scanning is more expensive than ultrasound, a two-step-procedure is recommended. Only the best 15-20% of selection candidate ram lambs measured by ultrasound are usually subsequently CT scanned (Lewis, 2004, Bünger et al., 2011).

A total of 384 Texel rams were analysed for 10 productivity traits including 2 CT measured traits. It should be noted that the dataset used in the present study was limited in its size, largely due to the restricted availability of CT-measured rams, due to CT costs. However, because this study analysed a small group of preselected animals

we acknowledged that the power to detect genome wide significant associations was diminished.

Genome Wide Association Analysis

In the current study no genome-wise significant association for any of the analysed traits was found. However, 8 chromosome-wise significant SNPs were found for: EWW, FD, MD, LW, FEC and LE. These SNPs were located on chromosomes 3, 4, 6, 11, 16 and 17, and were found to be either intronic or intergenic variants. None of the significant SNPs had been previously associated with any trait in sheep. However, chromosomes 11 and 16 have been previously tagged by SNPs associated with muscle, body and carcass weight (Cavanagh *et al.*, 2010).

We identified as candidate genes, those which were either directly tagged by a significant SNP (intronic variant) or those located within genomic regions of 30 kb up and downstream of an associated marker (Bolormaa *et al.*, 2016). However, due to the current relatively poor status of the ovine genome annotation, little information regarding the function of the tagged genes was obtained.

Regions tagged for EWW and LE have not been previously associated with any significant growth or welfare traits. However, two identified markers for LE, on chromosomes 6 and 17 (OAR6_108683365.1 and OAR17_11963200.1), belong to suggestive QTLs previously associated with parasite resistance (Beh *et al.*, 2002, Marshall *et al.*, 2009). Former studies have reported a low to moderate genetic correlation between lambing ease and birth weight (Brown, 2007), while a moderate genetic correlation between birth weight and parasite resistance has been suggested (Verbeek *et al.*, 2011). However, more information would be needed to estimate the genetic correlation between parasite resistance and welfare traits such as LE.

The region tagged by OAR16_20147789.1, significantly associated with FD, is an intronic variant of the NDUFAF2 gene, which encodes a NADH dehydrogenase (ubiquinone) complex I, assembly factor 2, a molecular chaperone for mitochondrial complex I assembly. OAR16_20147789.1 is located in a QTL region, which has been previously associated with final body weight, percent lean and subcutaneous fat area (Cavanagh *et al.*, 2010).

SNP s26074.1 was found to be significantly associated with LW. This SNP, is an intergenic variant, which is located in a QTL region formerly associated with body and carcass weight (Cavanagh *et al.*, 2010).

The region identified by SNP OAR11_12972551.1, was significantly associated with MD. This SNP is an intronic variant of the ACACA gene. ACACA encodes an acetyl-CoA carboxylase alpha, which is considered as a key enzyme of fatty acid synthesis in the mammary gland by catalysing the first step of fatty acid synthesis in mammalian cytosol. This gene has been described as a candidate gene for fat content in sheep, due to an observed significant association with variation in milk fat content, and change of fat composition in several sheep breeds (Bolormaa *et al.*, 2016). Moreover, OAR11_12972551.1 is located in QTL regions associated with body weight (Raadsma *et al.*, 2009), fat synthesis (Bolormaa *et al.*, 2016), internal fat amount and hot carcass weight (Cavanagh *et al.*, 2010).

Thus, results of significant associations with carcass traits provide evidence of a possible effect on FD, LW and MD by QTLs previously reported.by Raadsma *et al.* (2009), Cavanagh *et al. (*2010) and Bolormaa *et al.* (2016).

Finally, SNP s30868.1 associated with FEC, is an intronic variant of the ZNF227 gene, which encodes a zinc finger protein 227, probably involved in transcriptional regulation.

This gene is a paralogue of the ZNF229 gene, which has been previously associated with tuberculosis susceptibility in African human populations (Thye *et al.*, 2010). Also, s30868.1 tags a QTL region formerly reported to be associated with Immunoglobulin A level, an antibody that plays a crucial role in the immune function (Atlija *et al.*, 2016). This suggests that there might be a worm resistance QTL on chromosome 4.

A large number of QTLs have been identified for traits related to parasite resistance in sheep (Beh *et al.*, 2002, Marshall *et al.*, 2009, Atlija *et al.*, 2016) suggesting that those traits are not determined by individual genes acting alone but rather by complex multigene interactions. Thus, further identification of SNPs in strong LD with the casual variants, could contribute to the implementation of these results in breeding schemes for the Texel breed population.

The proportion of total variance explained by the significant SNPs was low, which is in agreement with Hayes and Goddard (2010), who explained that a small number of markers with validated associations would explain a small portion of the genetic variance in complex traits (Hayes and Goddard, 2010). This suggests that if alleles of large effect were present in our data, those would be in such a low frequency that they individually could only explain a small proportion of the variance.

Further improvement in sheep GWAS could be achieved by increasing the sample size and using the new ovine 700K HD chip, which has a much denser distribution of SNPs across the genome and thus should have higher LD with the potential QTLs controlling the traits of interest.

The present study found 8 chromosome-wise significant SNPs for 6 traits among them a CT measured trait (LW). Tagged regions on chromosomes 4, for worm resistance (FEC), 11 and 16, for carcass traits (MD, LW and FD), are consistent with other

studies, where QTL regions have been found for Immunoglobulin A level and meat and carcass traits, respectively. Whereas regions tagged on chromosomes 3, 6 and 17 for LE and EWW can be considered novel.

Among the tagged genes ZNF227, ACACA and NDUFAF2 were found. Hence, these genes could be considered as candidate genes for future research to further dissect the genomic architecture of the traits.

Conclusions

This study is one of very few studies using CT-derived carcass traits and other productivity traits already integrated in the selection index for terminal sire sheep breeds. It revealed some significant associations between genomic markers and important traits in sheep production. Further fine mapping the regions around these markers could lead to the identification of causative genes and better molecular predictors of CT based carcass composition, which might help to decrease phenotyping costs in the longer term. Results may also be integrated and inform genomic selection approaches and future SNP chip designs. The result may also guide similar studies in the other important Terminal Sire Breeds in the UK and beyond.

Acknowledgements

The research leading to this publication was supported by funds of the Scottish government (RERAD). DGH was supported by The Mexican National Council for Science and Technology. The opinions expressed and arguments employed in this paper are the sole responsibility of the authors and do not necessarily reflect those of RERAD.

This work acknowledges the support of AHDB Beef and Lamb, HCC and QMS in funding the use of the CT scanner and their annual contribution towards the provision of performance recording services delivered by Signet Breeding Services.

Many thanks to the British Texel Sheep society for their long-term support of the CT-scanning activities, for providing the tissue samples for subsequent DNA extraction and permitting the use of the genotypic information. Thanks go also to John McEwan and AgResearch in New Zealand for the genotyping.

References

Astle W and Balding D 2009. Population Structure and Cryptic Relatedness in Genetic Association Studies. Statistical Science 24, 451-471.

Atlija M, Arranz JJ, Martinez-Valladares M and Gutierrez-Gil B 2016. Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. Genet Sel Evol 48, 4.

Aulchenko YS, Ripke S, Isaacs A and van Duijn CM 2007. GenABEL: An R library for genome-wide association analysis. Bioinformatics 23, 1294-1296.

Beh KJ, Hulme DJ, Callaghan MJ, Leish Z, Lenane I, Windon RG and Maddox JF 2002. A genome scan for quantitative trait loci affecting resistance to Trichostrongylus colubriformis in sheep. Animal Genetics 33, 97-106.

Bolormaa S, Hayes BJ, van der Werf JH, Pethick D, Goddard ME and Daetwyler HD 2016. Detailed phenotyping identifies genes with pleiotropic effects on body composition. BMC Genomics 17, 224.

391 Brown DJ 2007. Variance Components for Lambing Ease and Gestation Length in Sheep. In
392 Proceedings of the 17th Conference of the Association for the Advancement of Animal Breeding
393 and Genetics, pp. 268-271. Armidale, Australia.
394 Bünger L, Macfarlane JM, Lambe NR, Conington J, Mclean KA, Moore K, Glasbey CA and
395 Simm G 2011. Use of X-Ray Computed Tomography (CT) in UK Sheep Production and
396 Breeding. 329-348.
397 Cavanagh CR, Jonas E, Hobbs M, Thomson PC, Tammen I and Raadsma HW 2010. Mapping
398 Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT
399 scans and aligned with a meta-assembly for sheep and cattle carcass QTL. Genetics, selection,
400 evolution : GSE 42, 36.
401 Donaldson CL, Lambe NR, Maltin CA, Knott S and Bunger L 2014. Effect of the Texel muscling
402 QTL (TM-QTL) on spine characteristics in purebred Texel lambs. Small Rumin Res 117, 34-40.
403 Fikse WF and Banos G 2001. Weighting Factors of Sire Daughter Information in International
404 Genetic Evaluations. Journal of Dairy Science 84, 1759-1767.
405 Georges M 2007. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in
406 domestic animals. Annual Review of Genomics and Human Genetics 8, 131-162.
407 Gianola D, Fariello MI, Naya H and Schon CC 2016. Genome-Wide Association Studies with a
408 Genomic Relationship Matrix: A Case Study with Wheat and Arabidopsis. G3 (Bethesda) 6,
409 3241-3256.
410 Goh L, Yap VB, Amos C, Wu X, Broderick P, Gorlov I, Gu J, Eisen T, Dong Q, Zhang Q, Gu X,
411 Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai Y, Chen W, Shete
412 S, Spitz M, Houlston R, Barrett J, Hansoul S, Nicolae D, Cho J, Duerr R, Rioux J, Brant S,
413 Silverberg M, Taylor K, Barmada M, Bitton A, Dassopoulos T, Datta L, Green T, Griffiths A,
414 Kistner E, Murtha M, Regueiro M, Rotter J, Schumm L, Steinhart A, Targan S, Xavier R, Libioulle
415 C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P,
416 Gossum AV, Zelenika D, Franchimont D, Hugot J, Vos Md, Vermeire S, Louis E, Belgian-French

417 I, Cardon L, Anderson C, Drummond H, Nimmo E, Ahmad T, Prescott N, Onnie C, Fisher S,
418 Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D,
419 Satsangi J, Mathew C, Parkes M, Georges M, Daly M, Bernardo MD, Crowther-Swanepoel D,
420 Broderick P, Webb E, Sellick G, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman A, Sunter
421 N, Hall A, Dyer M, Matutes E, Dearden C, Mainou-Fowler T, Jackson G, Summerfield G, Harris
422 R, Pettitt A, Hillmen P, Allsup D, Bailey J, Pratt G, Pepper C, Fegan C, Allan J, Catovsky D,
423 Houlston R, Frayling T, Nair R, Duffin K, Helms C, Ding J, Stuart P, Goldgar D, Gudjonsson J, Li
424 Y, Tejasvi T, Feng B, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi
425 S, Prahalad S, Guthery S, Fischer J, Liao W, Kwok P, Menter A, Lathrop G, Wise C, Begovich A,
426 Voorhees J, Elder J, Krueger G, Bowcock A, Abecasis G, Bakker Pd, Ferreira M, Jia X, Neale B,
427 Raychaudhuri S, Voight B, Feingold E, Diao G, Lin D, Labbe A, Wormald H, Peng B, Yu R,
428 Dehoff K, Amos C, Zhang F, Liu J, Chen J, Deng H, Purcell S, Neale B, Todd-Brown K, Thomas
429 L, Ferreira M, Bender D, Maller J, Sklar P, Bakker Pd, Daly M and Sham P 2009. Effects of
430 normalization on quantitative traits in association test. BMC Bioinformatics 10, 415.

431 Hayes B and Goddard ME 2010. Genome-wide association and genomic selection in animal
432 breeding. Genome 53, 876-883.

433 Hopkins A and Lobley M 2009. A Scientific Review of the Impact of UK Ruminant Livestock on
434 Greenhouse Gas Emissions. In CRPR Research Report, CRPR Research Report. University of
435 Exeter, Centre for Rural Policy Research,

436 Hu ZL, Park CA, Wu XL and Reecy JM 2013. Animal QTLdb: An improved database tool for
437 livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids
438 Research 41.

439 Jairath L, Dekkers JC, Schaeffer LR, Liu Z, Burnside EB and Kolstad B 1998. Genetic
440 Evaluation for Herd Life in Canada. Journal of Dairy Science 81, 550-562.

441 Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W,
442 Stanton JA, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SM, Adelson DL, Bian C, Cam

443 GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Fuentes-Utrilla P,
444 Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee
445 SL, Liu W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B,
446 Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB,
447 Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC,
448 Kijas JW, Wang J, Worley KC, Archibald AL, Cockett N, Xu X, Wang W and Dalrymple BP 2014.
449 The sheep genome illuminates biology of the rumen and lipid metabolism. Science 344, 1168-
450 1173.
451 Jones HE, Lewis RM, Young MJ and Wolf BT 2002. The use of X-ray computer tomography for
452 measuring the muscularity of live sheep. Animal Science 75, 387-399.
453 Lewis R 2004. Genetic Lessons from the United Kingdom. In Virginia-North Carolina Shepherds'
454 Symposium, pp. 24-34. Virginia, United States.
455 Lidauer M MK, Mantysaari E, Stranden I 2011. MiX99: solving large mixed model equations
456 manual. Jokioinen: MTT.
457 Macfarlane JM, Lewis RM, Emmans GC, Young MJ and Simm G 2006. Predicting carcass
458 composition of terminal sire sheep using X-ray computed tomography. Animal Science 82.
459 Macfarlane JM, Lewis RM, Emmans GC, Young MJ and Simm G 2009. Predicting tissue
460 distribution and partitioning in terminal sire sheep using x-ray computed tomography. J Anim Sci
461 87, 107-118.
462 Marshall K, Maddox JF, Lee SH, Zhang Y, Kahn L, Graser HU, Gondro C, Walkden-Brown SW
463 and Van Der Werf JHJ 2009. Genetic mapping of quantitative trait loci for resistance to
464 Haemonchus contortus in sheep. Animal Genetics 40, 262-272.
465 Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL and Bishop SC
466 2016. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as
467 assessed by computed tomography in Scottish Blackface lambs. Genetics, Selection Evolution
468 48, 11.

469 Pollott GE 2014. The breeding structure of the British sheep industry 2012. Defra.
470 R Core Team 2013. R: A language and environment for statistical computing. R Foundation for
471 Statistical Computing, Vienna, Austria. URL http://www.r-project.org/.
472 Raadsma HW, Thomson PC, Zenger KR, Cavanagh C, Lam MK, Jonas E, Jones M, Attard G,
473 Palmer D and Nicholas FW 2009. Mapping quantitative trait loci (QTL) in sheep. I. A new male
474 framework linkage map and QTL for growth rate and body weight. Genet Sel Evol 41, 34.
475 Royston P 1995. Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality.
476 Journal of the Royal Statistical Society. Series C (Applied Statistics) 44, 547-551.
477 Silva SR 2016. Use of ultrasonographic examination for in vivo evaluation of body composition
478 and for prediction of carcass quality of sheep. Small Ruminant Research.
479 Skinner ME, Uzilov AV, Stein LD, Mungall CJ and Holmes IH 2009. JBrowse: a next-generation
480 genome browser. Genome Res 19, 1630-1638.
481 Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F,
482 Enimil A, Chinbuah MA, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, Ngwira B,
483 Teo YY, Small K, Rockett K, Kwiatkowski D, Fine PE, Hill PC, Newport M, Lienhardt C,
484 Adegbola RA, Corrah T, Ziegler A, Morris AP, Meyer CG, Horstmann RD and Hill AVS 2010.
485 Genome-wide association analyses identifies a susceptibility locus for tuberculosis on
486 chromosome 18q11.2. Nature genetics 42, 739-741.
487 Verbeek E, Kanis E, Bett RC and Kosgey IS 2011. Optimisation of breeding schemes for litter
488 size, lambing interval, body weight and parasite resistance for sheep in Kenya. Livestock
489 Research for Rural Development. 23.
490 Walling GA, Visscher PM, Wilson AD, McTeir BL, Simm G and Bishop SC 2004. Mapping of
491 quantitative trait loci for growth and carcass traits in commercial sheep populations. Journal of
492 Animal Science 82, 2234-2245.

493 Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, Zhang S, Zhang X, Wei C, Lu G, Zheng Y and Du L
494 2013. Genome-wide association studies for growth and meat production traits in sheep. PLoS
495 One 8, e66569.

Tables

Table 1 *Descriptive statistics for the de-regressed EBVs of the analysed traits.*

Trait	**Unit**	**Acronym**	**Mean**	**SD**	**Minimum**	**Maximum**	**p value**
Growth Traits							
Birth Weight	kg	BW	0.48	0.81	-2.19	2.89	0.88
Eight Week Weight	kg	EWW	3.24	11.30	-27.01	43.26	**0.10**
Scan Weight	kg	SW	7.17	7.60	-14.69	35.22	0.17
Carcass Traits							
Fat Depth	mm	FD	-0.08	1.74	-6.1	5.78	**0.07**
Muscle Depth	mm	MD	1.73	3.42	-8.64	12.4	0.16
Fat Weight	kg	FW	0.79	1.75	-4.05	6.50	**0.10**
Lean Weight	kg	LW	2.17	2.01	-3.53	8.70	0.74
Muscularity	Ratio	MU	3.3	5.85	-12.94	18.14	0.33
Health Trait							
Faecal Egg Count	Log values	FEC	0.12	0.58	-2.72	4.77	**< 0.001**
Welfare Trait							
Lambing Ease	Score units (1-6)	LE	0.05	11.98	-70.11	24.83	**<0.001**

SD = Phenotypic standard deviation, 384 tested individuals, Significant p values, for Shapiro and Wilk's W-statistic test, ($p \leq 0.1$) in bold. Fat and Lean weights were measured by CT (as described by Bunger *et al.* (2011))

501 **Table 2** *Chromosome-wide significant SNPs associated with important economic traits*

502 *and size of estimated effects.*

SNP	Chr	Position OAR v3.1 / OAR v4.0	Allele Effect	SD	P-value	Trait	Nearest Gene (Code)	Nearest Gene (Name)
OAR17_22884911.1	17	20425356 / 20428283	-0.388	0.09	3.9E-05	EWW	PCDH18 [454.22]	Protocadherin 18
OAR16_20147789.1	16	18368560 / 18365229	-0.439	0.10	1.3E-05	FD	NDUFAF2	Ubiquinone oxidoreductase complex assembly factor 2
s26074.1	11	8271088 / 8261942	0.673	0.15	2.6E-05	LW	CUEDC1 [37.38]	CUE domain containing 1
OAR11_12972551.1	11	13110133 / 13079564	-1.115	0.25	1.7E-05	MD	ACACA	Acetyl-CoA carboxylase alpha
s30868.1	4	56089343 / 56074079	-0.336	0.07	2.0E-05	FEC	ZNF227	Zinc finger protein 227
OAR6_108683365.1	6	98702734 / 98597850	0.341	0.07	6.8E-06	LE	NKX6 [193.99]	NK6 homeobox 1

s23722.1	3	178956951 /178727572	0.519	0.11	9.3E-06	LE	MB [92.5]	Myoglobin
OAR17_11963200.1	17	10808289 / 10794783	-0.363	0.08	1.6E-05	LE	TTC29 [295.07]	Tetratricopeptide repeat domain 29

Chr (Chromosome); Allele effect (deviations from the mean); SD (standard deviation) of the allele effect; P-value for the significance of the association; Units for FEC and LE on the transformed scale; SNPs located within known ovine genes are highlighted in bold; the nearest genes were identified using the ENSEMBL Genome Browser; the number in brackets is the distance from SNP to the nearest gene.

Figure Captions

Figure 1: Manhattan plots for EWW, FD, LW, MD, FEC and LE traits, blue line refers to the genome-wise threshold and the red line to the chromosome-wise significance threshold.