Genome-wide association study of meat quality traits using whole-genome sequence data in a multi-breed sheep population

N. Duijvesteijn[1,2], S. Bolormaa[1,3], C. Gondro[2,7], S. Clark[2], M. Khansefid[1,3], N. Moghaddar[1,2], A.A. Swan[1,4], P. Stothard[5], H.D. Daetwyler[3,6],J.H.J. van der Werf[1,2] & I.M. MacLeod[1,3]

[1]Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia nduijves@une.edu.au (Corresponding Author)
[2] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia
[3]Agriculture Victoria Research, AgriBio Centre, Bundoora, VIC 3083, Australia
[4]Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia
[5]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada
[6]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia
[7]Department of Animal Science, Michigan State University, East Lansing, MI 48824, USA

Summary

This is the first study showing GWAS results from whole-genome sequence (WGS) data in sheep. The objective of this study was to fine map previously identified QTL and detect new QTL for meat quality traits in a multi-breed sheep population using WGS data. The traits measured were carcass fat depth, intramuscular fat, post-weaning eye muscle depth, post-weaning body weight and SF5. Corrected phenotypes were available on 13,363 to 26,769 Australian sheep depending on the trait.
The GWAS for all five traits for the 50K, HD, and WGS identified 18, 52 and 250 QTL regions, respectively based on a 5% FDR threshold. Together, the largest QTLs explained a substantial part of the additive genetic variance (between 10.6 - 14.3%). Previously identified QTL on chromosomes 6 and 18 were confirmed. Single breed analysis for Merino sheep only, showed a reduced significance of the large QTL, but in similar regions. We were able to identify candidate genes involved in meat quality using functional annotation. We conclude that the WGS data provided more and clearer evidence of QTL regions, potentially resulting in more predictive SNPs and more detailed functional analysis of genes underlying the QTL regions.

Keywords: GWAS, whole-genome sequence data, sheep, meat quality

Introduction

Adequate meat quality is important for the sheep meat industry to remain competitive (Russell*, et al.*, 2005). Breeding programs to improve meat quality include traits measured on the live animal and carcass. The traits are expensive to measure and often cannot be measured on the selection candidate. The Australian sheep industry is therefore focussed on overcoming these difficulties using genomic selection, while trying to develop low cost highly predictive genotyping panels. The panels can be customised to include putative causal variants or markers in high LD with variants affecting economically important traits. Previous studies have identified QTL regions for meat quality traits in sheep based on medium to high density marker panels (Walling, et al., 2004, Al-Mamun, et al., 2015, Bolormaa, et al., 2016), but few

studies have attempted to fine map QTL to the causative mutations (Knight, et al., 2012). Recently it has been possible to impute whole-genome sequence (WGS) data for more than 40,000 Australian sheep (Bolormaa, et al., 2018) and this provides the opportunity to fine map QTL to causative mutations. Under the assumption of accurate imputation and sufficient number of reference animals sequenced, all causal variants should, at least in principle, be present in the imputed sequence data.

The objective of this study was to fine map previously identified QTL and detect new QTL for meat quality traits in a multi-breed sheep population using WGS data. We used a multi-breed population because LD is conserved over much shorter distances compared to a single breed population, and this should help to more accurately fine map QTL regions.

Material and methods

Phenotypes and genotypes

Phenotypes for five meat and carcass traits measured between 2008 and 2015 on a multi-breed sheep population in Australia were used as response variables. Phenotypes originated from various research flocks (Sheep Genomics, Information Nucleus and MLA Resource Flocks) and industry and were corrected for non-genetic effects such as contemporary group. Table 1 shows the number of observations per trait and mean of the raw phenotypes. Animals with phenotypes were mostly Merino-crossbreds or purebred Merinos.

Genotypes for the animals were imputed to whole-genome sequence (WGS) using a European breed reference population of 726 sequenced sheep (Bolormaa, et al., 2018). Variants with an R^2 imputation accuracy (reported by Minimac) $\leq$ 0.4, variants with a MAF<0.01 and those on the X chromosome were excluded leaving a total of 27,896,226 variants in the GWAS for each trait. The number of SNPs overlapping the 50K SNP chip and HD SNP chip was 36,403 and 476,390 respectively.

Genome-wide association study

The GWAS was run fitting a mixed model with WOMBAT SNP Snappy (single SNP analysis) (Meyer, 2007) including the animal's breed proportion (based on pedigree, Q matrix), its additive genetic (polygenic) value and the individual SNP. Additive genetic values were distributed as $N \sim (0, G\sigma_a^2)$, where σ_a^2 is the additive genetic variance and G is a genomic relationship matrix calculated from HD SNP genotypes (Yang, et al., 2010).

Significance levels for each SNP were calculated from the resulting t-statistic assuming a two-tailed t-distribution. The significance threshold for each trait was a genome-wise 5% false discovery rate. QTL regions were defined as significant SNPs extending by a 0.5 Mb region to the left and right of the top associated variant until no other significant SNPs were detected. For each QTL region, either the gene in which the top associated SNP was in, or the closest gene to the left or right of the top associated SNP was selected for a further detailed functional annotation of the sequence variants, including GO terms, using the NGS-SNP pipeline (Grant, et al., 2011).

Test for multiple QTL or breed specific QTL

After visual inspection of LD in large QTL regions (> 2 Mb region and number of

significant SNPs within QTL >10) the GWAS was repeated with the top associated variant fitted as an additional fixed effect to check for multiple QTL.

For the top associated variant within a QTL region, the MAF within all genotyped pure breeds (Border Leicester, N=632; Poll Dorset, N=1,953; Merino, N=14,887) was calculated. For Merinos, the GWAS was repeated for specific chromosomes where large QTL were detected. The other breeds did not have sufficient phenotypes on purebreds for a single breed GWAS.

Results and Discussion

Comparing power of QTL detection with different SNP densities

The GWAS for all five traits for the 50K, HD, and WGS data identified 18, 52 and 250 QTL regions, respectively at the 5% FDR threshold, (Table 1). This demonstrates that imputed WGS provides more power to detect QTL, likely because the sequence variants are in stronger linkage disequilibrium (LD) with, or include, the causal variants. A Manhattan plot for one trait (PEMD) is shown in Figure 1 contrasting QTL discovery for the three different SNP densities. On average the LD between the most significant SNP within a QTL region from WGS data and the closest linked SNP from the 50K was 0.23. For HD the LD was higher at 0.63. There were only two cases where the most significant SNP identified from the WGS GWAS were on the HD SNP chip. The low LD between 50K SNPs and the most significant WGS SNP indicates that the 50K chip would not accurately estimate QTL effects in these regions. Therefore, we expect an increase in the accuracy of genomic prediction in sheep by including the most informative WGS SNPs in the analysis.

QTL regions from sequence data

In this study, we focused on QTL regions in the WGS analysis with > 10 significant SNPs (to guard against false positives, see Table 1). Some QTL overlapped between traits, so in total 30 independent QTL were identified. The additive genetic variance explained by each QTL varied between 0.3% and 7.3%. The SNP explaining the most variance was Chr6:36989893 and was associated with CCFAT (Table 2). The most significant SNP was Chr18:64477210 (P-value of 5.9 E-30) and was associated with SF5 (Table 2). In total all QTL explained 10.6%, 10.9%, 13.3%, 13.4% and 14.3% of the genetic variance for the traits CCFAT, IMF, PWT, SF5 and PEMD respectively. The average MAF across all breeds for the most significant variants was 0.21, but varied between breeds. Nine of the SNPs only segregated in Merinos and eight SNPs only segregated in Border Leicester or Poll Dorset (other breeds MAF <0.05). Annotation of the most significant variant in each of the 38 QTL found 17 were intergenic, 14 intronic, while 7 were within 5kb of genes (two downstream, five upstream).

Test for multiple QTL or breed specific QTL

Within the large QTL on OAR 6 and 18, the most significant SNP was included in the model while re-testing all other SNP in a GWAS, to determine whether multiple QTL could be underlying the detected QTL region. The effect on OAR 18 for multiple traits seemed to be due to a single variant, while for OAR 6, the region still showed some other SNP to be significantly associated for PEMD. In contrast, no other SNP remained significant in the CCFAT QTL region (Figure 3).

The large QTL regions on OAR 6 and 18 were also analysed within Merinos (MER). The GWAS results for OAR 6 showed expected results for traits CCFAT, PEMD and PWT where the QTL region was clear for MER (Figure 4). LD between the most significant SNP from the multi-breed and single-breed MER analyses were high for CCFAT (same SNP), PEMD (r^2 = 0.85) and PWT (r^2 = 0.81).The multi-breed GWAS, resulted in stronger associations compared to the single breed analyses. This is most likely due to the higher number of animals used in the multi-breed study, many of which were Merino crossbreds.

Functional annotation of some QTL regions

A major QTL region showing a strong association with PEMD, CCFAT, PWT and IMF was identified on chromosome 6 (Table 2). This QTL, shows very high LD across > 1 Mb in the multi-breed population for the top PEMD SNP (Figure 2B). This QTL was previously identified using 50K and HD chip in pure- and crossbred Australian sheep by Daetwyler, *et al.* (2012), Al-Mamun *et al.* (2015) and Bolormaa *et al.* (2016) for growth, bone and carcass traits. In our study the most significant SNP in the QTL region for PEMD, is within the gene *NCAPG* (non-SMC condensin I complex subunit G), known to be associated with height and weight traits across several mammalian species (Fig 2B). The most significant SNP for CCFAT (Fig 2A) is close to the gene *MED28* (Mediator Complex Subunit 28), which is involved in the pathway for regulation of lipid metabolism (Chinetti, et al., 2000). Due to the high LD in this region, it is difficult to determine if there are different causal variants/genes affecting the fat traits versus the PEMD and PWT traits.

On chromosome 18, another large QTL region was strongly associated with IMF, SF5 and PEMD (Table 2). The most significant SNPs for IMF and SF5 primarily segregated in the BL breed (SNP Chr18:64477210) and the top SNP for SF5 also segregates with low MAF in PD and MER. The effect of the alternative allele at Chr18:64477210 results in a large increase in SF5 and reduction in IMF (both undesirable). This QTL region spans two well-known major QTL: the callipyge mutation (*CLPG*) and *Carwell*. *CLPG* significantly increases lean-meat percentage and hindquarter muscling, but also increases meat thoughness (Koohmaraie, et al., 1995). The *CLPG* phenotype had an unusual inheritance pattern, known as polar-overdominance, which is only expressed in the heterozygous animals that inherit a paternal *CLPG* mutation and a normal maternal allele. To date our tests for the inheritance pattern were inconclusive, because there were insufficient alternative homozygotes with recorded phenotypes.

Unlike *CLPG*, the *Carwell* QTL inheritance mode was reported as dominance, but it appears that imprinting and polar overdominance were not explicitly tested. The *Carwell* region (~64 Mb, causal variant unknown) has been reported to have a strong effect on increasing eye muscling and has also been reported to have a small effect on meat toughness (Jopson *et al.* (2001). However, our reported QTL has a large effect on toughness and marbling, with a much smaller effect detected for eye muscling. To our best knowledge, ours is the first study to report an effect in the BL breed, similar to *Callipyge* and *Carwell* mutations. The results to date suggest that this mutation is not the same as the *CLPG* or *Carwell* mutations: similar to the example in cattle where a range of myostatin mutations cause similar double muscling phenotypes but of differing intensities. The mode of inheritance of the SNP effects is under further investigation among crossbreds in our study.

Besides known large QTL effects, other regions on the genome were significant for one or multiple traits. Genes underlying QTL for meat tenderness (measures as SF5), included calpains which are believed to play a major role in postmortem tenderization of meat, including *CAST* (calpastatin) and *CAPN1* (calpain-1 catalytic subunit). Besides calpains and cathepsins, muscle tissue also contains numerous other peptidases able to degrade a large set of muscle protein components. In our study we found *TLL1* (Tolloid like 1), which is involved in *calcium ion binding* and *metallopeptidase activity* to be associated with SF5 (OAR 17). This gene has not previously been reported to be associated with meat quality traits in sheep.

Conclusions

This is the first study showing GWAS results from whole-genome sequence (WGS) data in sheep. The use of WGS data with a multi-breed sheep population demonstrated increased power to detect QTL regions. The largest QTLs jointly explained a substantial part of the additive genetic variance (between 10.6 - 14.3%). Therefore it is likely that these sequence SNP can be added to custom SNP chips to improve the accuracy of genomic prediction. Based on this data, we aim to analyse further QTL regions to more precisely fine map and determine causative genes and mutations across a wide range of traits. Importantly, these GWAS results will provide the basis for companion studies on gene expression and functional annotation to validate candidate genes and mutations affecting economically important traits in sheep.

Acknowledgments

The authors acknowledge the contributions from breeders and many CRC participants that contributed to the Sheep CRC Information Nucleus flocks. The authors thank Sheep Genetics and MLA for providing access to phenotypic data from industry animals. Also Klint Gore is acknowledged for his help on retrieving data.

List of References

Al-Mamun, H. A., P. Kwan, S. A. Clark, M. H. Ferdosi, R. Tellam & C. Gondro, 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. Genet. Sel. Evol. 47(1): p 66.

Bolormaa, S., A. Chamberlain, J. H. J. v. d. Werf, H. D. Daetwyler & I. M. MacLeod, 2018. Evaluating the accuracy of imputed whole genome sequence in sheep. Proc. 11th World Congr. Genet. Appl. Livest. Prod.

Bolormaa, S., B. J. Hayes, J. H. van der Werf, D. Pethick, M. E. Goddard & H. D. Daetwyler, 2016. Detailed phenotyping identifies genes with pleiotropic effects on body composition. BMC Genomics 17(1): p 224.

Chinetti, G., J.-C. Fruchart & B. Staels, 2000. Peroxisome proliferator-activated receptors (PPARs): nuclear receptors at the crossroads between lipid metabolism and inflammation. Inflamm. Res. 49(10): p 497-505.

Daetwyler, H. D., A. A. Swan, J. H. van der Werf & B. J. Hayes, 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet. Sel. Evol. 44(1): p 33.

Grant, J. R., A. S. Arantes, X. Liao & P. Stothard, 2011. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. Bioinformatics 27(16): p 2300-2301.

Jopson, N., G. Nicoll, J. Stevenson-Barry, S. Duncan, G. Greer, W. Bain, E. Gerard, B. Glass, T. Broad, et al., 2001. Mode of inheritance and effects on meat quality of the rib-eye muscling (REM) QTL in sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 14:p 111-114.

Knight, M., H. Daetwyler, B. Hayes, M. Hayden, A. Ball, D. Pethick & M. McDonagh, 2012. Discovery and trait association of single nucleotide polymorphisms from gene regions of influence on meat tenderness and long-chain omega-3 fatty acid content in Australian lamb. Animal production science 52(7): p 591-600.

Koohmaraie, M., S. Shackelford, T. Wheeler, S. M. Lonergan & M. Doumit, 1995. A muscle

hypertrophy condition in lamb (callipyge): characterization of effects on muscle growth and meat quality traits. J Anim Sci 73(12): p 3596-3607.

Meyer, K., 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ Sci B 8(11): p 815-821.

Russell, B., G. McAlister, I. Ross & D. Pethick, 2005. Lamb and sheep meat eating quality—industry and scientific issues and the need for integrated research. CSIRO: p 465-467.

Walling, G., P. Visscher, A. Wilson, B. McTeir, G. Simm & S. Bishop, 2004. Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. J Anim Sci 82(8): p 2234-2245.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, et al., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42(7): p 565-569.

Table 1. The size of the population (N) and mean and standard deviation (SD) of the raw phenotypes and number of significant QTL region for the 50K, HD and WGS GWAS for five traits.

Trait[1]	N	Mean	SD	50K	HD	WGS_{all}[(2)]	$WGS_{>10}$[(3)]
CCFAT	13,644	4.04	2.27	4	9	41	4
IMF	11,772	4.36	1.14	2	8	38	8
PEMD	21,412	26.34	4.65	4	15	49	10
PWT	26,769	46.70	13.10	5	13	88	10
SF5	13,363	25.31	14.05	3	7	34	6

[1]CCFAT=carcass fat depth, IMF=intramuscular fat, PEMD=post-weaning eye muscle depth, PWT=post-weaning body weight, SF5=shear force day 5.
[2]All QTL passing a FDR of 5%.
[3] QTL regions with >10 SNPs significant.

Table 2. Information on detected QTL at chromosome 6 and 18 and the most significant variant from WGS data.

QTL description				**Most significant variant**					
Trait	Chr	Size QTL region (Mb)	N SNPs	Position (bp)	Significance	Effect size (%)[1]	Sign effect	Annotation	MAF[2] (PD, BL, MER)
PEMD	6	5.00	3841	37302018	1.34E-22	5.90	-	intron	0.36 (0, 0, 0.38)
CCFAT	6	5.33	4409	36989893	3.86E-25	7.33	-	intergenic	0.49 (0.14, 0.08, 0.35)
PWT	6	5.41	5047	37163225	5.82E-28	6.57	+	intron	0.36 (0, 0, 0.38)
IMF	6	0.28	69	36988278	1.38E-07	1.34	-	intergenic	0.49 (0.14, 0.08, 0.36)
IMF	18	6.73	1847	64477210	2.68E-29	4.50	-	upstream	0.05 (0, 0.29, 0)
SF5	18	4.36	985	64477210	5.86E-30	6.53	+	upstream	0.05 (0, 0.29, 0)
PEMD	18	1.23	84	64484335	2.35E-09	0.78	+	upstream	0.15 (0.01, 0.32, 0.13)

[1] Percentage genetic variance explained by SNP (calculated as $2pq\alpha^2$) divided by the genetic variance. *p* and *q* are the allele frequencies (across breeds) and α is the allele substitution effect.
[2] MAF is the minor allele frequency for the multi-breed population and single breed populations; PD= Poll Dorset, BL= Border Leicester and MER= Merino.

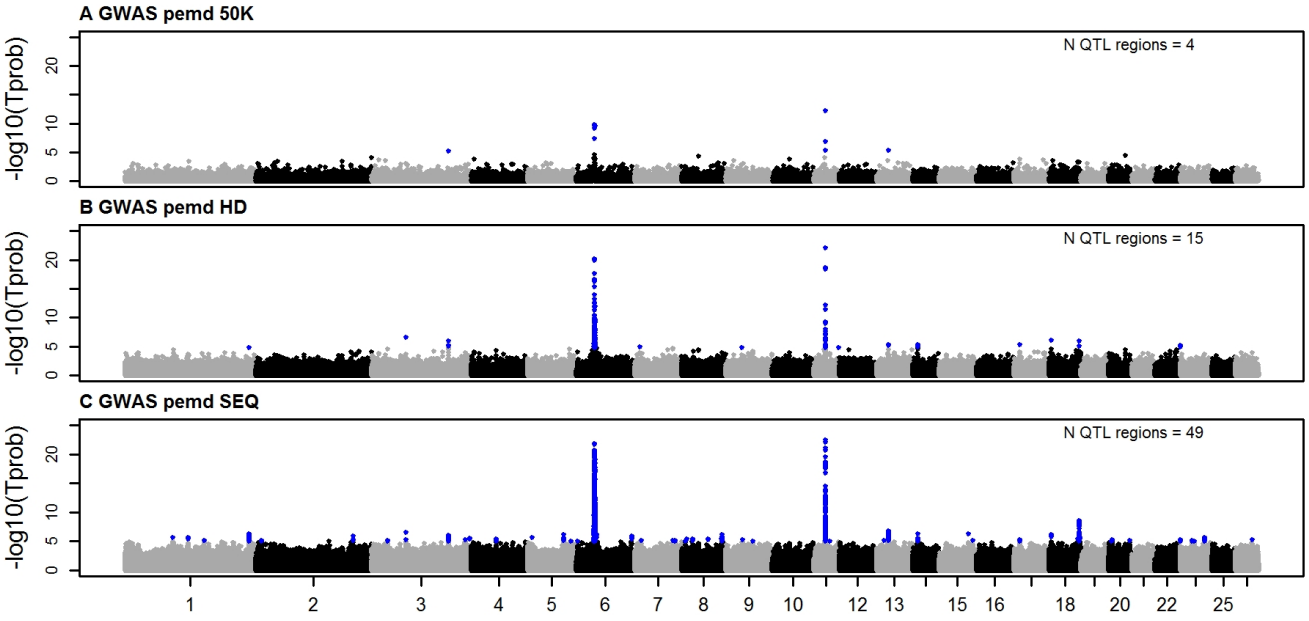

Figure 1. Manhattan plot of genome-wide –log10(Tprob) for post-weaning eye-muscle depth (PEMD) in (A) 50K SNPs, (B) HD SNPs and (C) sequence variants. The dots in blue indicate significant SNPs at a FDR of 5%.

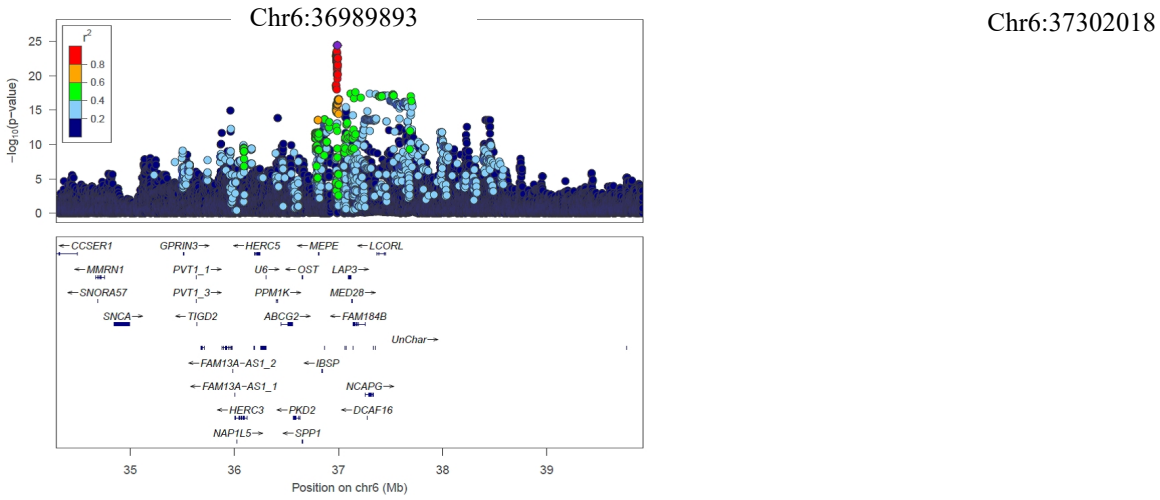

*Figure 2. Manhattan plot for **A**) CCFAT (carcass fat depth) and **B**) PEMD; where circles represent SNPs and the colours depict the squared correlation (r^2) of each SNP with the most associated SNP (i.e., CCFAT Chr6:36989893). Purple designates the SNP with the strongest association. In the bottom panel, genes are reported within the QTL region. Plots were produced using LocusZoom and show a part of OAR 6.*

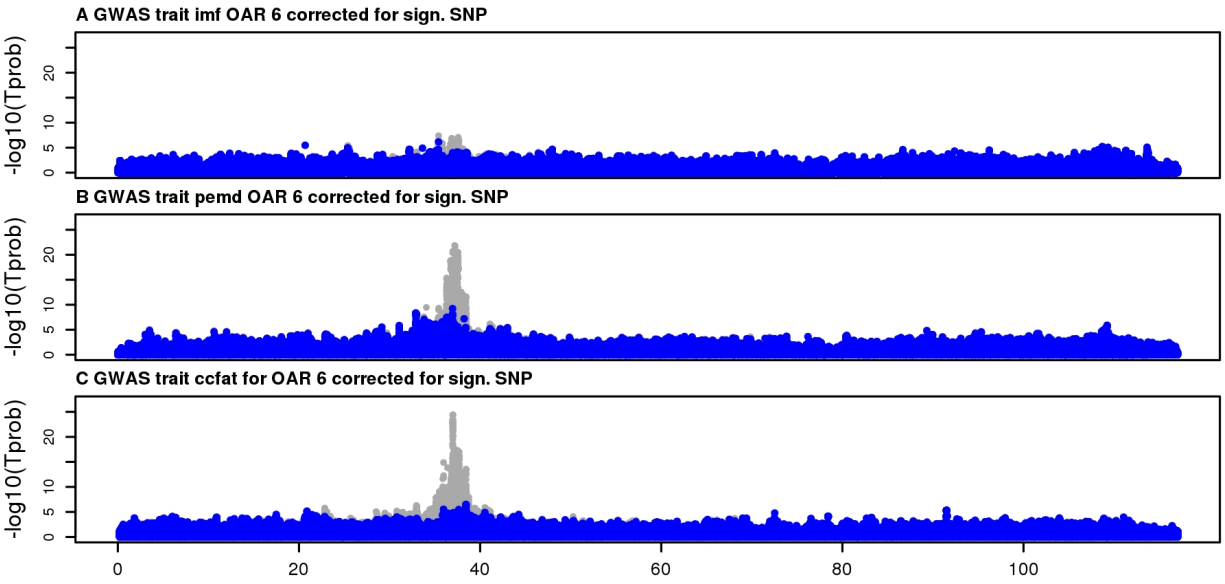

Figure 3. Comparison of the original GWAS (coloured in grey) versus a GWAS correcting for the most significant SNP from the original GWAS on OAR 6 (coloured in blue) for three traits. ***A)*** *IMF = intramuscular fat,* ***B)*** *PEMD = post-weaning eye muscle depth and* ***C)*** *CCFAT = carcass fat.*

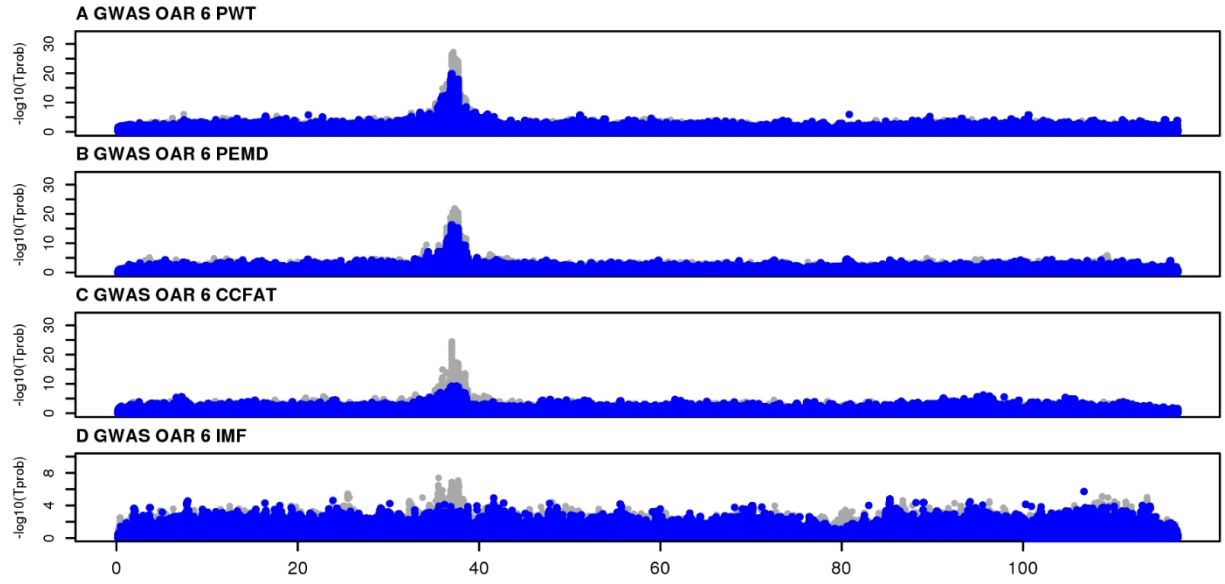

Figure 4. Comparison of the single-breed (Merino only, coloured in blue) versus multi-breed (coloured in grey) GWAS for four traits of OAR 6. ***A)*** *PWT = post-weaning weight,* ***B)*** *PEMD = post-weaning eye muscle depth,* ***C)*** *CCFAT = carcass fat and* ***D)*** *IMF = intramuscular fat. The y-axis has a different scale for IMF.*