MATN3 underlies a QTL for stature in cattle

TJ Lopdell[ab]* and MD Littlejohn[b]

[a]*School of Biological Sciences, University of Auckland, Auckland, New Zealand;* [b]*Livestock Improvement Corporation, Newstead, New Zealand*

**Corresponding author.* Email: Thomas.Lopdell@lic.co.nz

Abstract

Live weight (LWT) in dairy cattle is associated with milk production efficiency, and also has a high genetic correlation with stature. To identify genetic regions impacting stature and LWT, we conducted a genome-wide association study using 65,500 mixed-breed NZ dairy cattle and 16,243,015 imputed genotypes. We identified a quantitative trait locus (QTL) at BTA11:78-79Mbp for stature that was significant at a genome-wide level (β=3.98mm ± 0.56; P=1.6×10^{-12}) but not significant for LWT (β=1.253 ± 0.393 kg; P=1.4×10^{-3}). Ensembl's Variant Effect Predictor was used to search for coding variants to attempt to identify a candidate causative gene for this QTL, and gene expression from RNA sequence data representing 373 cows were analysed for 16 genes and 3,640 markers to highlight expression QTL (eQTL) in this region. The gene with the most significant eQTL was *MATN3* (ENSBTAG00000020893; P=2.5×10^{-38}), encoding a protein which is involved in bone development. The correlation between variant effects (-log p-values) was calculated for the stature QTL and *MATN3* eQTL (r=0.584), with the same variant (rs475277351) most significant for both QTL. This suggests a common genetic signal underlying both QTL, supporting the hypothesis that the *MATN3* gene underlies the stature QTL.

Keywords: live weight; stature; quantitative trait locus; eQTL; *MATN3*; RNAseq; cattle

Introduction

The body size traits live weight (LWT) and stature are classic quantitative genetic traits, and widely recorded due to their ease of measurement and moderately high heritabilities (approximately 0.4 for both traits in cattle; Banos & Coffey 2012). These two measures exhibit a high genetic correlation, with estimates in cattle ranging from 0.6 (Riley et al. 2007) to 0.65 (Banos & Coffey 2012), suggesting that the same genetic signals generally act on both measures. These traits are economically important to New Zealand farmers because of the association of larger animals with elevated levels of production, but also with diminished efficiency at the same level of production, due to greater feed requirements (Holmes et al. 1993).

Several quantitative trait loci (QTL) have been published previously for body size traits in cattle, including the *PLAG1* QTL (Karim et al. 2011; Fink et al. 2017) at BTA14:25.0Mbp (positions in million base-pairs on the UMD3.1 bovine reference genome) and the *HMGA2* QTL (Bolormaa et al. 2014) at BTA5:48.1Mbp. These QTL are frequently pleiotropic, affecting phenotypes in other than body size. Here, we report a QTL for stature identified on BTA11, with which LWT is not significantly associated after multiple-testing correction, and identify a candidate causative gene underlying the QTL.

Materials and methods

Genome-wide association study

The stature and LWT results reported here were part of a genome-wide association study (GWAS) performed on 65,500 mixed-breed, mixed-age New Zealand dairy cows. Cows were genotyped using the Illumina BovineSNP50, BovineHD, or Geneseek Genomic Profiler BeadChip platforms, then imputed to whole-genome sequence resolution using Beagle4 (Browning & Browning 2009) with a reference population of 565 cattle, as described by Littlejohn et al. (2016). This yielded 16,243,015 genotypic markers. Phenotypic records for LWT and stature were extracted from the national herd recording database at LIC, with data adjusted using the procedures outlined by Fink et al. (2017).

The GWAS was performed using an additive linear model implemented in the Plink software package (v1.90b3i; Chang et al. 2015). Principal components analysis (PCA; Price et al. 2006) was performed on the genomic relationship matrix (GRM) calculated from the genotypes to account for family structure and population stratification, then the first 1000 components were fitted as covariates to the phenotype using a least-squares model in JMP software (v13.2.1; SAS Institute Inc, Cary, NC, 2007). The phenotypic residual was subsequently used for association analysis.

Variant effect predictions

The GWAS analysis described above yielded a QTL peak for stature on BTA11 at approximately 78-79Mbp. All markers that were in LD (R^2>0.7) with the most significant marker (rs475277351), and within one million base pairs, were extracted and analysed using the Variant Effect Predictor (VEP) tool at Ensembl (http://ensembl.org/Bos_taurus/Tools/VEP/; Aken et al. 2016) on the *Bos taurus* UMD3.1 reference with Ensembl transcripts (build 90).

eQTL analysis

Expression QTL (eQTL) association analyses were conducted on the expression levels of the sixteen genes within 1Mbp of the top stature marker, using the genotypes of the 3,640 WGS markers in the same genomic interval, as a next step in highlighting candidate causative genes

underlying the QTL. This analysis followed the method outlined by Lopdell et al. (2017). Briefly, RNA sequencing (RNAseq) was performed on biopsies from the lactating mammary glands of 373 cows, comprising animals from the AgResearch Tokanui research farm and a smaller number of cows from the LIC Friesian-Jersey crossbred trial (Spelman et al. 2001). Biopsy samples were obtained in accordance with protocols approved by the Ruakura Animal Ethics Committee, Hamilton, New Zealand (approval AEC 12845). Samples were processed using a stranded RNAseq library preparation and sequenced on an Illumina HiSeq 2000 instrument, yielding 100bp paired-end reads. These were mapped to the reference genome using Tophat2 (version 2.0.12; Kim et al. 2013) and normalised prior to association analysis using the variance-stabilising transformation implemented in the DESeq R package (Anders & Huber 2010).

The association analysis followed the approach described by Lopdell et al. (2017). A generalised least-squares model was fitted using the numerator relationship (A) matrix to account for pedigree and population stratification. The A-matrix for the 373 animals was calculated using pedigree records extracted from the LIC herd-recording database.

Results

Stature QTL identified

A significant QTL was identified for the stature phenotype (Fig. 1) at BTA11:78-79Mbp (β=3.98 ± 0.56 mm; $P=1.6\times10^{-12}$), using the conventionally-applied genome-wide significance threshold of 5×10^{-8} (Pe'er et al. 2008). The top marker identified was rs475277351, located at BTA11: 78,916,750. This marker was an indel (genotypes T/TAAAACA) with a minor allele frequency (MAF) of 0.026. Restricting the analysis to pure-bred Holstein-Friesians and Jerseys indicated that this variant is segregating in the Holstein-Friesian breed (n=12,458; MAF=0.040; β=4.14 ± 1.05 mm; $P=8.1\times10^{-5}$), but not in Jerseys (n=7,012; MAF=0.001; β=8.82 ± 8.16 mm; P=0.28). In both breeds, the minor allele was T. Surprisingly, no co-located QTL was detected for LWT, with the association at this marker (β=1.25 ± 0.39 kg; $P=1.4\times10^{-3}$) not significant at the genome-wide level. For comparison, the *PLAG1* QTL on BTA14 (Karim et al. 2011) with all breeds yielded highly significant results for both the LWT (β=1.47 ± 0.116 kg; $P=5.6\times10^{-37}$) and stature (β=1.82 ± 0.01 mm; $P=7.17\times10^{-28}$) phenotypes.

Results from the VEP suggest that rs475277351 maps to an intergenic region, along with two of the five sites in LD (r>0.7) with it. Another variant maps downstream of the *RHOB* gene, while the remaining two variants map into introns of the *SDC1* and *LAPTM4A* genes. As no missense or nonsense mutations were identified that would predicted to impact protein function, it is likely that the QTL operates via an underlying regulatory mechanism.

Regulatory mechanisms

To investigate possible mechanisms operating via the regulation of gene expression, association analyses were undertaken for the sixteen genes positionally located within the QTL region in order to detect potential eQTL. The most highly-significant eQTL detected was for the *MATN3* gene ($P=2.5\times10^{-38}$), which maps to BTA11:78,893,508–78,904,614. Intriguingly, the same variant rs475277351 was found to be the most strongly associated with this eQTL as for the stature QTL, with other variants in LD also highly associated in both QTL (Fig. 2). This suggests that both QTL share a common underlying genetic regulation, implying that, if the two QTL truly are co-regulated, the significance levels of variants should be correlated between the two QTL. Here, the Pearson correlation between significance statistics for the variants common to the stature QTL and *MATN3* eQTL was moderately high, at 0.584 (Fig. 3a).

Another significant eQTL was observed for the *LAPTM4A* gene ($P=5.4\times10^{-12}$), which maps to BTA11:78,862,495–78,880,461. The top associated variants for this eQTL were rs110552157 and rs109993903, neither of which was associated with stature (P=0.858).

Figure 1 Manhattan plot illustrating the stature QTL detected on BTA11 at 78–79 Mbp in 65,500 mixed-age, mixed-breed NZ dairy cattle. Each point represents a whole-genome sequence variant. The dashed box indicates the most strongly associated variant, along with other variants with correlations >0.7 with that variant.

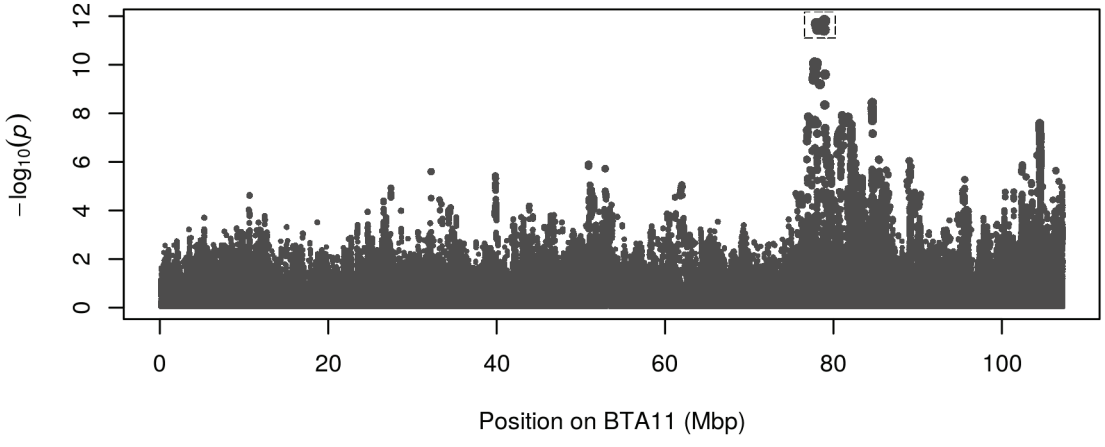


Figure 2 Zoomed Manhattan plots highlighting the region surrounding the stature QTL (a) and MATN3 eQTL (b), both on BTA11, and identified in NZ dairy cattle. The dashed box indicates the most strongly associated variant for the stature QTL, along with other variants with correlations >0.7 with that variant, showing that the same cluster of variants is strongly associated with both phenotypes.

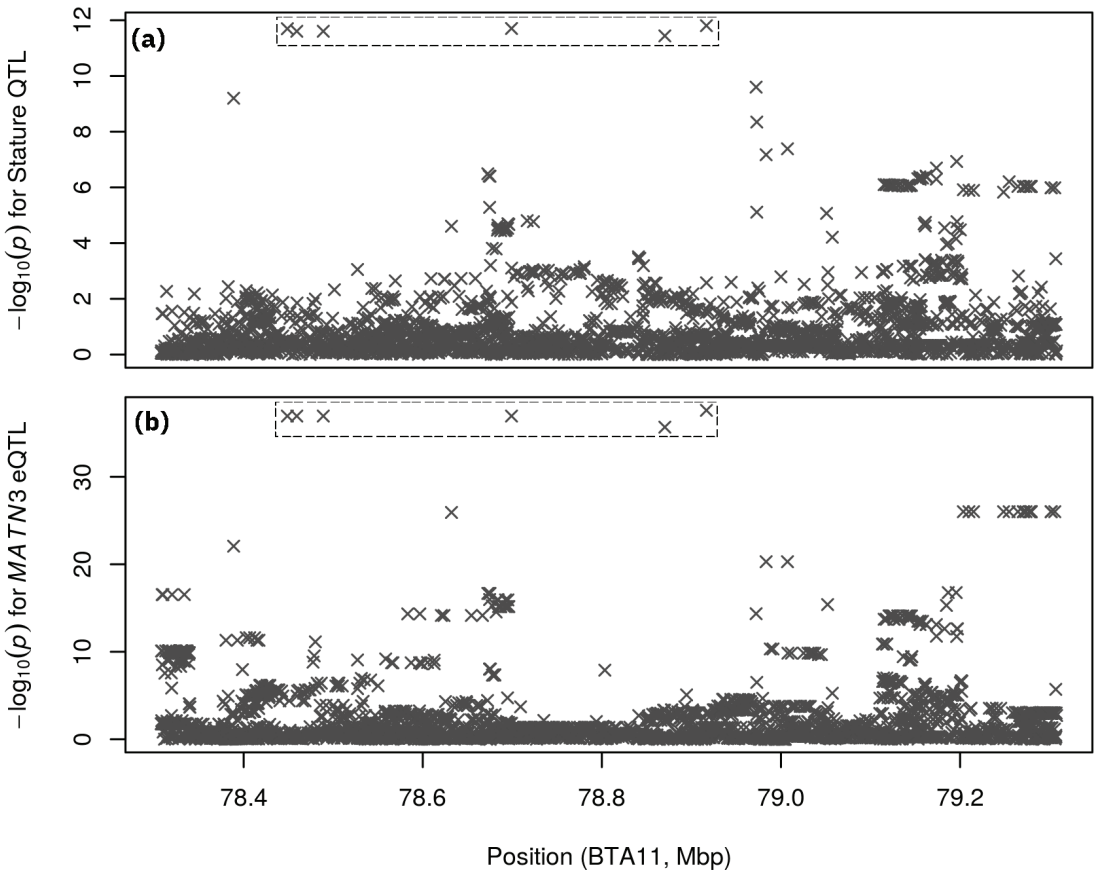
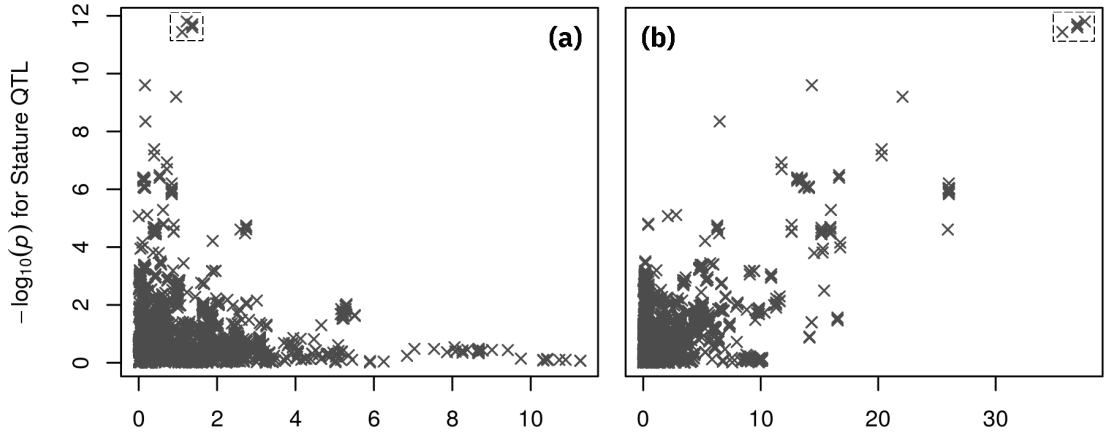


Figure 3 The log10(*p*-values) for two eQTL plotted against those for the BTA11 stature QTL. Each point represents a whole-genome sequence variant. The dashed box indicates the most strongly associated variant for the stature QTL, along with other variants with correlations >0.7 with that variant. The correlations for the (a) *LAPTM4A* eQTL, and (b) *MATN3* eQTL, respectively, with the stature QTL are -0.160 and 0.584.



The correlation between the significance statistics of the stature QTL and the *LAPTM4A* eQTL was -0.160 (Fig. 3b), as variants associated with one phenotype tended not to be associated with the other. A third eQTL was observed for the *RHOB* gene ($P=4.6\times10^{-8}$) which was marginally significant at the genome-wide threshold of 5×10^{-8}. This eQTL was also uncorrelated with the stature QTL (r=0.004).

Discussion

A QTL for stature (height at sacrum) has previously been reported near the locus reported here (Boichard et al. 2003), peaking at BTA11:82cM. However, that study included only eight markers on BTA11, and no attempt was made to assign an underlying causative gene at that time. Although the BTA11 stature QTL has been reported

subsequently (Cole et al. 2011, Bouwman et al. 2018), we are unaware of any literature assigning a causative gene to the QTL, so we consider the work presented here as the first to do so.

The largest stature QTL reported in cattle to date is at the *PLAG1* locus on BTA14, with β=2 cm (Karim et al. 2011). Here, we found that the BTA11 QTL had a larger effect (β=3.98 mm) than the *PLAG1* QTL (β=1.82 mm), although the effect size attributed to *PLAG1* in our study is much smaller than that reported by others previously, suggesting we may have been overly conservative in our application of PCA to adjust for population stratification.

As no protein-coding mutations in significant LD with the BTA11 tag variants for the stature QTL were identified by the VEP analysis, the underlying effect for this QTL is likely to be regulatory. The eQTL analysis identified three genes with significant eQTL mapping near this locus: *LAPTM4A*, *RHOB*, and *MATN3*. The *LAPTM4A* eQTL has a correlation of -0.160 with the stature QTL, suggesting that genetic regulation of this gene is independent of the stature QTL. The *RHOB* eQTL also appears to be regulated independently (r=0.004). The *MATN3* eQTL, in contrast, exhibited a correlation of 0.584 with the stature QTL, suggesting that this eQTL is co-regulated with the stature QTL. This assertion is further supported by the fact that the same variant is most strongly associated with both QTL. Therefore, we consider *MATN3* the most likely candidate causative gene.

The *MATN3* gene encodes the protein matrilin 3, a component of the extra-cellular matrix (Klatt et al. 2011) that interconnects collagen and aggrecan fibres. Matrilins have been associated with the development and maintenance of cartilage. Mutations in humans play a role in the skeletal diseases multiple epiphyseal dysplasia (Kannu et al. 2009) and osteoarthritis (Pullig et al. 2002; Klatt et al. 2009). Additionally, mutations in *MATN3* have been associated with growth plate collagen abnormalities in mice, leading to short-limbed dwarfism (Myllyharju 2014), and with shortened long bones in humans, leading to early-onset dwarfism (Borochowitz et al. 2004).

The associations of *MATN3* with bone growth and dwarfism lead us to consider *MATN3* as a candidate causal gene for this QTL. The case for this is strengthened by underlying genetic co-regulation of the stature QTL with the *MATN3* eQTL, as evidenced by the correlation between the two QTL, and by the fact that the two QTL are most highly associated with the same variant. However, as the eQTL reported here were discovered in mammary tissue, these results assume a shared regulatory architecture for this gene between mammary tissue and stature-relevant tissues such as bone or cartilage. This leaves open an alternative hypothesis that a different gene, driven by a regulatory circuit that mammary tissue does not adequately represent (though is influenced by the same underlying haplotype) is responsible for the stature QTL. Although the data used in this study cannot rule out this alternative hypothesis, we contend that the developmental function of *MATN3*, together with associated disease phenotypes in humans, makes it a strong candidate, and that the most plausible hypothesis is that *MATN3* is the causative gene at this locus.

In conclusion, therefore, we propose that genetically-regulated differences in expression of the *MATN3* gene underlie the bovine stature QTL identified by Boichard et al. (2003) on BTA11. If validated, this discovery will contribute to our understanding of the genetics of cattle growth, with potential future applications in both the dairy and beef industries.

Acknowledgements

The authors would like to acknowledge S. Morgan and staff at DairyNZ and Phil McKinnon, Ali Cullum and staff at AgResearch for facilitating mammary tissue sampling of lactating animals. We also acknowledge New Zealand Genomics Limited (NZGL) and the University of Auckland Centre for Genomics, Proteomics, and Metabolomics for RNA preparation and sequencing, and both the Australian Genome Research Facility (AGRF) and Illumina FastTrack for RNA and genomic DNA sequencing. Finally, we wish to acknowledge the computational assistance provided by Michael Keehan at LIC. This work was supported by MPI, who co-funded the work through the Primary Growth Partnership.

References

Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Gil L 2016. Ensembl 2017. Nucleic acids research 45(D1): D635–D642.

Anders S, Huber W 2010. Differential expression analysis for sequence count data. Genome biology 11(10): R106.

Banos G, Coffey MP 2012. Prediction of liveweight from linear conformation traits in dairy cattle. Journal of dairy science 95(4): 2170–2175.

Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, Rupp R, Amigues Y, Boscher MY, Levéziel H 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics Selection Evolution 35(1): 77–101.

Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ, Goddard ME 2014. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLoS genetics 10(3): e1004198.

Borochowitz ZU, Scheffer D, Adir V, Dagoneau N, Munnich A, Cormier-Daire V 2004. Spondylo-epi-metaphyseal dysplasia (SEMD) matrilin 3 type: homozygote matrilin 3 mutation in a novel form of SEMD. Journal of medical genetics, 41(5): 366–372.

Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, Govignon-

Gion A, Boitard S, Dolezal M, and others, 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics 50: 362–367.

Browning BL, Browning SR 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics 84(2): 210–223.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1): 7.

Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK, Da Y 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. BMC genomics 12(1): 408.

Fink T, Tiplady K, Lopdell T, Johnson T, Snell RG, Spelman RJ, Davis SR, Littlejohn MD 2017. Functional confirmation of *PLAG1* as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. Scientific Reports 7: 44793.

Holmes CW, Wilson GF, Kuperus W, Buvaneshwa S 1993. Liveweight, feed intake and feed conversion efficiency of lactating dairy cows. Proceedings of the New Zealand Society of Animal Production 53: 95–95).

Kannu P, Bateman JF, Belluoccio D, Fosang AJ, Savarirayan R 2009. Employing molecular genetics of chondrodysplasias to inform the study of osteoarthritis. Arthritis & Rheumatology 60(2): 325–334.

Karim L, Takeda H, Lin L, Druet T, Arias JA, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B, Harris BL 2011. Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. Nature genetics 43(5): 405–413.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology 14(4): R36.

Klatt AR, Becker AKA, Neacsu CD, Paulsson M, Wagener R 2011. The matrilins: modulators of extracellular matrix assembly. The international journal of biochemistry & cell biology 43(3): 320–330.

Klatt AR, Klinger G, Paul-Klausch B, Kühn G, Renno JH, Wagener R, Paulsson M, Schmidt J, Malchau G, Wielckens K 2009. Matrilin-3 activates the expression of osteoarthritis-associated genes in primary human chondrocytes. FEBS letters 583(22): 3611–3617.

Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, Couldrey C, Keehan M, Sherlock RG, Harland C, Scott A 2016. Sequence-based association analysis reveals an *MGST1* eQTL with pleiotropic effects on bovine milk composition. Scientific reports 6: 25376.

Lopdell TJ, Tiplady K, Struchalin M, Johnson TJJ, Keehan M, Sherlock R, Couldrey C, Davis SR, Snell RG, Spelman RJ. Littlejohn MD 2017. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. BMC genomics: 18(1): 968.

Myllyharju J 2014. Extracellular matrix and developing growth plate. Current osteoporosis reports 12(4): 439–445.

Pe'er I, Yelensky R, Altshuler D, Daly MJ 2008. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology 32(4): 381–385.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38(8): 904–909.

Pullig O, Weseloh G, Klatt AR, Wagener R, Swoboda B 2002. Matrilin-3 in human articular cartilage: increased expression in osteoarthritis. Osteoarthritis and cartilage 10(4): 253–263.

Riley DG, Coleman, SW, Chase CC, Olson TA, Hammond, AC 2007. Genetic parameters for body weight, hip height, and the ratio of weight to hip height from random regression analyses of Brahman feedlot cattle. Journal of animal science 85(1): 42–52.

Spelman RJ, Miller FM, Hooper JD, Thielen M, Garrick DJ 2001. Experimental design for QTL trial involving New Zealand Friesian and Jersey breeds. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 14: 393–396.