BMC Genomics

A new mechanism for a familiar mutation – bovine *DGAT1* K232A modulates gene expression through multi-junction exon splice enhancement

Tania Fink[1†], Thomas J. Lopdell[1,2*†] , Kathryn Tiplady[2,3], Renee Handley[1], Thomas J. J. Johnson[2], Richard J. Spelman[2], Stephen R. Davis[2], Russell G. Snell[1] and Mathew D. Littlejohn[2,3]

Abstract

Background: The *DGAT1* gene encodes an enzyme responsible for catalysing the terminal reaction in mammary triglyceride synthesis, and underpins a well-known pleiotropic quantitative trait locus (QTL) with a large influence on milk composition phenotypes. Since first described over 15 years ago, a protein-coding variant K232A has been assumed as the causative variant underlying these effects, following in-vitro studies that demonstrated differing levels of triglyceride synthesis between the two protein isoforms.

Results: We used a large RNAseq dataset to re-examine the underlying mechanisms of this large milk production QTL, and hereby report novel expression-based functions of the chr14 g.1802265AA > GC variant that encodes the *DGAT1* K232A substitution. Using expression QTL (eQTL) mapping, we demonstrate a highly-significant mammary eQTL for *DGAT1,* where the K232A mutation appears as one of the top associated variants for this effect. By conducting in vitro expression and splicing experiments in bovine mammary cell culture, we further show modulation of splicing efficiency by this mutation, likely through disruption of an exon splice enhancer as a consequence of the allele encoding the 232A variant.

Conclusions: The relative contributions of the enzymatic and transcription-based mechanisms now attributed to K232A remain unclear; however, these results suggest that transcriptional impacts contribute to the diversity of lactation effects observed at the *DGAT1* locus.

Keywords: Cattle, RNA, Transcriptomics, Milk, RNA splicing, QTL

Background

A lysine to alanine amino acid substitution (K232A) encoded by a mutation in the diacylglyercol O-acyltransferase 1 (*DGAT1*) gene has major impacts on bovine lactation traits, the most substantial being

its impact on milk fat percentage [1, 2]. This substitution, initially described by Grisart et al. (2002), results from an AA to GC dinucleotide substitution in exon eight of *DGAT1*, and likely constitutes the most widely studied and validated variant in association analyses of bovine milk characteristics. The *DGAT1* gene encodes an enzyme responsible for catalysing the terminal reaction in the mammary triglyceride synthesis pathway [3], and the *DGAT1* K allele has been shown [4] to synthesise more

* Correspondence: thomas.lopdell@lic.co.nz
†Tania Fink and Thomas J. Lopdell contributed equally to this work.
[1]School of Biological Sciences, University of Auckland, Auckland, New Zealand
[2]Livestock Improvement Corporation, Hamilton, New Zealand
Full list of author information is available at the end of the article

triglycerides in vitro when compared to the A allele. Aside from the *DGAT1* K232A mutation, an additional polymorphism 5′ of the transcription start site of the gene has also been shown to associate with milk fat percentage [5]. This variant, a variable number tandem repeat (VNTR) expansion, was speculated to play a role in bovine milk composition by increasing the number of putative transcription factor binding sites [5]. However, functional testing of the VNTR variant did not show any differences in *DGAT1* expression between QTL genotypes in cell culture [6]. This finding largely put the competing, gene expression-based hypothesis of the *DGAT1* milk fat effect to rest at the time, with enzymatic differences deriving from the K232A mutation now widely assumed as the underlying mechanism.

Since these initial analyses > 10 years ago, further functional characterisation of the K232A mutation has been largely absent. Having generated a large, mammary RNAseq dataset, however, we recently had the opportunity to re-examine this locus for potential regulatory effects impacting *DGAT1*, among other loci [7], and demonstrated a strong *DGAT1* expression-QTL (eQTL) in the mammary gland. Importantly, the expression of *DGAT1* transcripts was associated with K232A genotype (and thus milk fat percentage), and a strong correlation ($r = 0.946$) was observed between the *DGAT1* eQTL and lactose yield (a trait with high genetic correlations with other milk yield phenotypes). Based on these observations, we have investigated mechanisms by which the K232A variant might mediate this effect. For the first time, we present functional and statistical evidence suggesting splice-enhancement-based regulatory control of *DGAT1* transcripts as an explanation for the eQTL, and by inference the milk fat and lactation effects attributed to this gene and mutation.

Results

DGAT1 K232A is strongly associated with *DGAT1* transcript abundance in the lactating mammary gland

To test for regulatory effects on *DGAT1* transcript levels, we performed an association analysis using a mammary RNAseq dataset representing 375 lactating cows (Methods). Association testing was conducted using mammary *DGAT1* transcript counts transformed using the variance-stabilising transformation (VST; see Methods), and 3128 previously-imputed WGS variants [8, 9], representing a 1Mbp interval centred on (and including) the K232A mutation. This analysis revealed a highly significant eQTL for *DGAT1* (Fig. 1). Curiously, K232A was one of the top associated variants ($P = 1.59 \times 10^{-25}$; Fig. 1), explaining 29.7% of the phenotypic variance in mammary *DGAT1* expression (Supplementary Table 1). While K232A was significantly associated with *DGAT1* expression, the most highly associated markers for this signal were rs209328075 and rs209929366, located upstream of *DGAT1* at chr14: 1730455 and chr14:1747132, respectively ($P = 2.31 \times 10^{-28}$). These markers exhibited identical association statistics with mammary *DGAT1* expression, and were also highly correlated with K232A, exhibiting an R^2 value of 0.88.

Notably, the milk fat percentage-increasing K allele was the same allele associated with increased *DGAT1* expression in this analysis. Those animals homozygous for the K allele had a mean transformed read count for *DGAT1* of 9.628 (±0.024), whereas for those animals with the A allele this value was 9.244 (±0.026). Heterozygous animals had a mean transformed read count of 9.436 (±0.019), intermediate between the two opposing homozygous classes. The frequency of the *DGAT1* K allele was 0.51 in the RNAseq population.
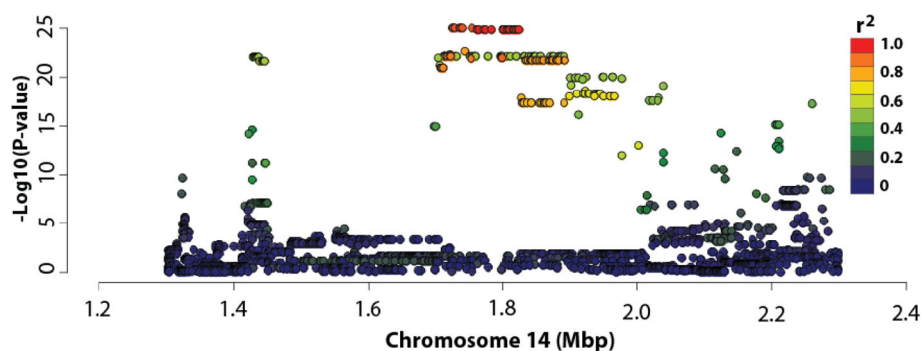
Fig. 1 Expression QTL analysis at the *DGAT1* locus in bovine lactating mammary gland. The Manhattan plot for *DGAT1* expression at the *DGAT1* locus in the RNAseq animals ($n = 375$). The X-axis shows chromosome 14 position (million base pairs); the Y-axis shows $-\log_{10}$ P-values of marker association for the 3218 WGS-derived SNPs in the 1 Mbp interval centred on *DGAT1* K232A. Markers are coloured based on their correlations (R^2) with K232A

DGAT1 K232A associates with alternative splicing of DGAT1 exon 8

The association of K232A and other local genetic variants with *DGAT1* transcripts presented several potential hypotheses to explain these effects. These included association of K232A merely based on LD (i.e. an effect driven by other *cis*-regulatory variants on the same or similar haplotype), or a direct role for K232A in regulating transcript abundance. Given a previous report highlighting the impact of *DGAT1* K232A on the generation of an alternative *DGAT1* splicing isoform [4], we aimed to investigate the latter hypothesis - under the assumption that alternative splicing might give rise to differential RNAseq read-counts and thus genetic association. In the analysis that first reported splicing effects for K232A [4], this mutation was proposed to activate an alternative intron 8 5′-splice donor (Fig. 2), so we first set out to confirm this finding. To this end, the relative usage of the reference and alternative exons was quantified by calculating the percentage of reads spliced in [10] (PSI; see Methods). Association analysis of the PSI in conjunction with K232A genotype, revealed a significant difference in the median percentage of reads splicing at the alternative splice site for exon 8 based on K232A genotype ($P = 1.658 \times 10^{-10}$; Kruskall-Wallis test).

Notably, the *DGAT1* expression-increasing K allele was the same allele associated with increased alternative splicing of *DGAT1* exon 8 (i.e., lower PSI): the median PSI was 94.1% for KK genotype animals, compared to 96.4% for AA animals. Heterozygotes were intermediate, at 94.5%. However, we reasoned that this was unlikely to explain the eQTL, as the alternative isoform differs through the 'intronification' [4] of the majority of exon 8, and is therefore expected to result in a reduction in the number of reads mapping to the exons of the gene overall.

DGAT1 K232A associates with splicing efficiency of DGAT1 introns

Having observed differential splicing at exon 8 based on calculating PSI of the alternative transcript structure, we next investigated whether this variant associated with the efficiency of splicing at the neighbouring splicing junction. To this end, a splicing efficiency phenotype was derived for intron 8, which involved quantification of the ratio of total *DGAT1* RNAseq reads that mapped to this intron junction (see Methods). Here, this ratio was taken as representative of the rate constant of splicing [11], where this in turn might be expected to influence the yield of mature mRNA available to the cell.

Association analysis between the percentage of RNAseq reads mapping to intron 8 and K232A and the 3128 WGS SNPs revealed a strong splicing efficiency effect, with K232A the 28th most significantly associated variant ($P = 7.48 \times 10^{-19}$, Fig. 3). The 27 preceding SNPs were similarly associated and statistically indistinguishable ($P = 4.37 \times 10^{-19}$) from one another. Importantly, the direction of effect for the splicing efficiency effect was consistent with a mechanism that might explain the *DGAT1* eQTL, such that the K allele was associated with an increased percentage of completely spliced transcripts at the exon 8 junction. Of the three genotype classes, the KK animals had the highest percentage of completely spliced transcripts with only 1.16% of *DGAT1* RNAseq reads mapping to intron 8, while the AA animals had the lowest percentage of completely spliced transcripts with 1.94% of reads mapping to intron 8. Heterozygous AK animals showed intermediate levels of splicing (1.53% of reads mapping to intron 8). As the K allele is also associated with increased *DGAT1* expression, this suggested splicing efficiency as a potential limiting mechanism for the production of fully spliced mRNA.

Given the observation of differential splicing efficiency of intron 8, we conducted association analyses for each of the other 13 *DGAT1* junctions, using the same 3128 markers. This revealed significant associations at eight additional *DGAT1* junctions (Table 1), with the most significant impacting the intron 2 junction ($P = 5.25 \times 10^{-46}$). For the seven significantly impacted junctions, K232A was the lead SNP for the association at two junctions (introns 2 and 7). Interestingly, intron 2 is several kb from K232A yet was the most significant splicing efficiency effect in this analysis (Table 1).

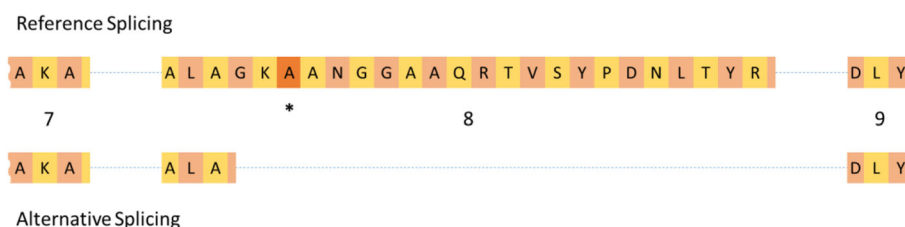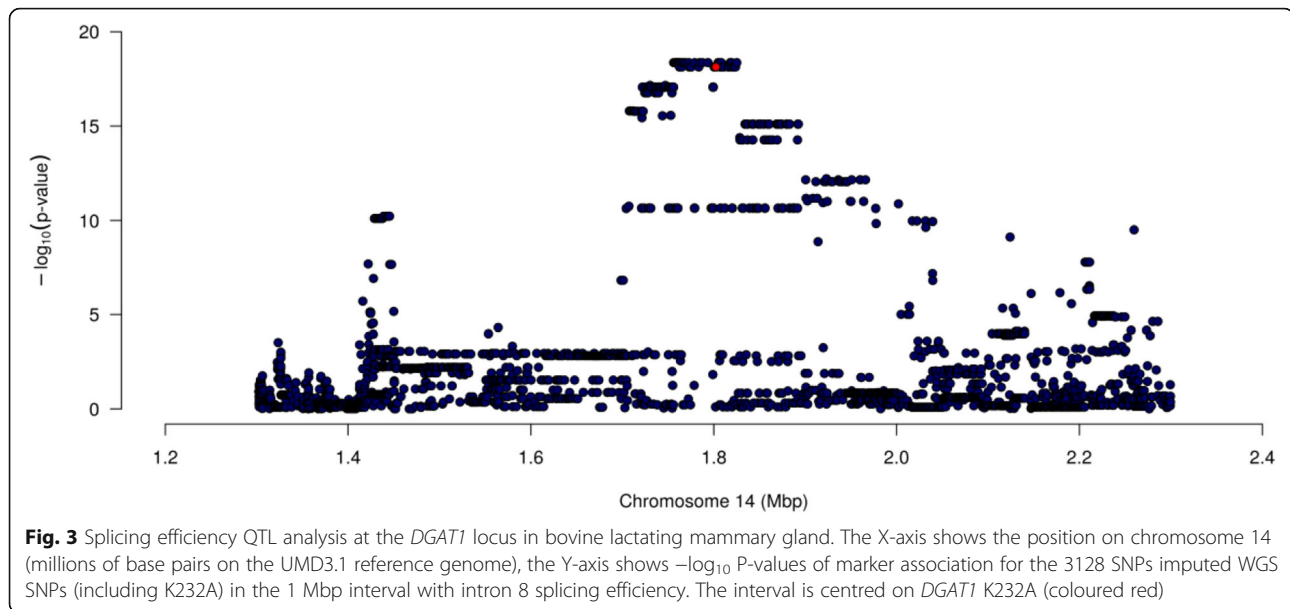**Fig. 2** Schematic structure of *DGAT1* exon 8 for the reference and alternative 5′ splice donor sites. The 3′ end of exon 7 and 5′ end of exon 9 are also shown. Blue dotted lines represent introns 7 and 8. The asterisk indicates the position of the K232A variant. Twenty-two amino acids present in the reference isoform are spliced out of the alternative isoform. Exon numbers are shown under the reference isoform

Fig. 3 Splicing efficiency QTL analysis at the *DGAT1* locus in bovine lactating mammary gland. The X-axis shows the position on chromosome 14 (millions of base pairs on the UMD3.1 reference genome), the Y-axis shows $-\log_{10}$ P-values of marker association for the 3128 SNPs imputed WGS SNPs (including K232A) in the 1 Mbp interval with intron 8 splicing efficiency. The interval is centred on *DGAT1* K232A (coloured red)

DGAT1 K232A alters *DGAT1* splicing efficiency in vitro

Despite *DGAT1* K232A being associated with the expression of *DGAT1* and splicing efficiency of multiple *DGAT1* junctions, the alternative hypothesis involving a linked *cis*-regulatory variant remained a plausible explanation for these effects. To test the possibility of K232A simply being in linkage disequilibrium (LD) with another regulatory mutation, cell-based experiments were undertaken to remove the two K232A transcript isoforms from their genomic context.

Table 1 *DGAT1* junction splicing efficiency association statistics for the top sequence variants and K232A

DGAT1 Intron	Top SNP	*P*-value	K232A Rank	K232A *P*-value
Intron 1	rs209328075	$\mathbf{3.60 \times 10^{-30}}$	69	$\mathbf{6.19 \times 10^{-29}}$
Intron 2	K232A	$\mathbf{5.25 \times 10^{-46}}$	1	$\mathbf{5.25 \times 10^{-46}}$
Intron 3	rs211223469	$\mathbf{4.63 \times 10^{-18}}$	30	$\mathbf{3.95 \times 10^{-17}}$
Intron 4	rs110825388	8.23×10^{-4}	27	0.00119
Intron 5	rs209328075	7.47×10^{-5}	61	3.17×10^{-4}
Intron 6	rs377961056	2.43×10^{-4}	1279	0.515
Intron 7	K232A	$\mathbf{9.91 \times 10^{-16}}$	1	$\mathbf{9.91 \times 10^{-16}}$
Intron 8	rs109162116	$\mathbf{4.37 \times 10^{-19}}$	28	$\mathbf{7.48 \times 10^{-19}}$
Intron 9	rs210757486	0.00742	1475	0.427
Intron 10	rs134628055	5.48×10^{-4}	2278	0.827
Intron 11	rs109050667	$\mathbf{8.23 \times 10^{-10}}$	70	$\mathbf{1.68 \times 10^{-8}}$
Intron 12	rs137017663	3.26×10^{-5}	377	0.0148
Intron 13	rs383740077	3.82×10^{-6}	459	0.0312
Intron 14	rs443459090	$\mathbf{4.25 \times 10^{-7}}$	214	1.32×10^{-4}

The fourth column displays the ranking of the association between K232A and splicing at the indicated intron, compared to all the markers. The significance of the K232A association is presented in the fifth column, compared to the *P*-value of the most-significantly associated variant in the third column (identical if the most significant variant was K232A). *P*-values smaller than the Bonferroni threshold ($P = 1.14 \times 10^{-6}$) are highlighted using bold type

For these in vitro expression experiments, plasmid-based mini-gene constructs were synthesised and cloned into a pcDNA3.1 vector backbone under the control of a cytomegalovirus (CMV) promoter (Fig. 4). These constructs were designed to represent the identical intron-exon structure of genomic *DGAT1* with the exception that introns 1 and 2 had been removed (due to plasmid size constraints, see Methods). The resultant expression vectors differed only by the AA and GC alleles that encode the K/A residues, allowing spliced and unspliced transcripts to be compared between constructs. Quantification of splicing was achieved by qPCR, targeting four individual intron/exon boundaries that either showed (introns 3 and 7), or did not show (introns 5 and 13) major differential splicing efficiency effects in vivo (Table 1; Fig. 4).

Following transfection of the *DGAT1* constructs into the bovine mammary cell line MAC-T [12], the alternate mini-gene alleles were found to recapitulate the splicing effects observed in the lactating mammary dataset. Transcripts generated from cells expressing the K allele showed higher splicing ratios at the intron 3 and intron 7 junctions compared with those cells transfected with the A allele ($P = 9.37 \times 10^{-4}$ and $P = 9.05 \times 10^{-11}$, respectively; Table 2). For the intron 5 and 13 junctions, there were no significant differences in splicing efficiency ($P = 0.256$ and $P = 0.497$, respectively; Table 2).

Having confirmed splicing efficiency differences, of further, particular interest was whether the two transcript isoforms differed in mean spliced mRNA yield, since this would support a direct role of K232A in modulating the eQTL effects seen in vivo. When analysing qPCR data in this way, the mean expression of spliced transcripts was indeed found to be significantly
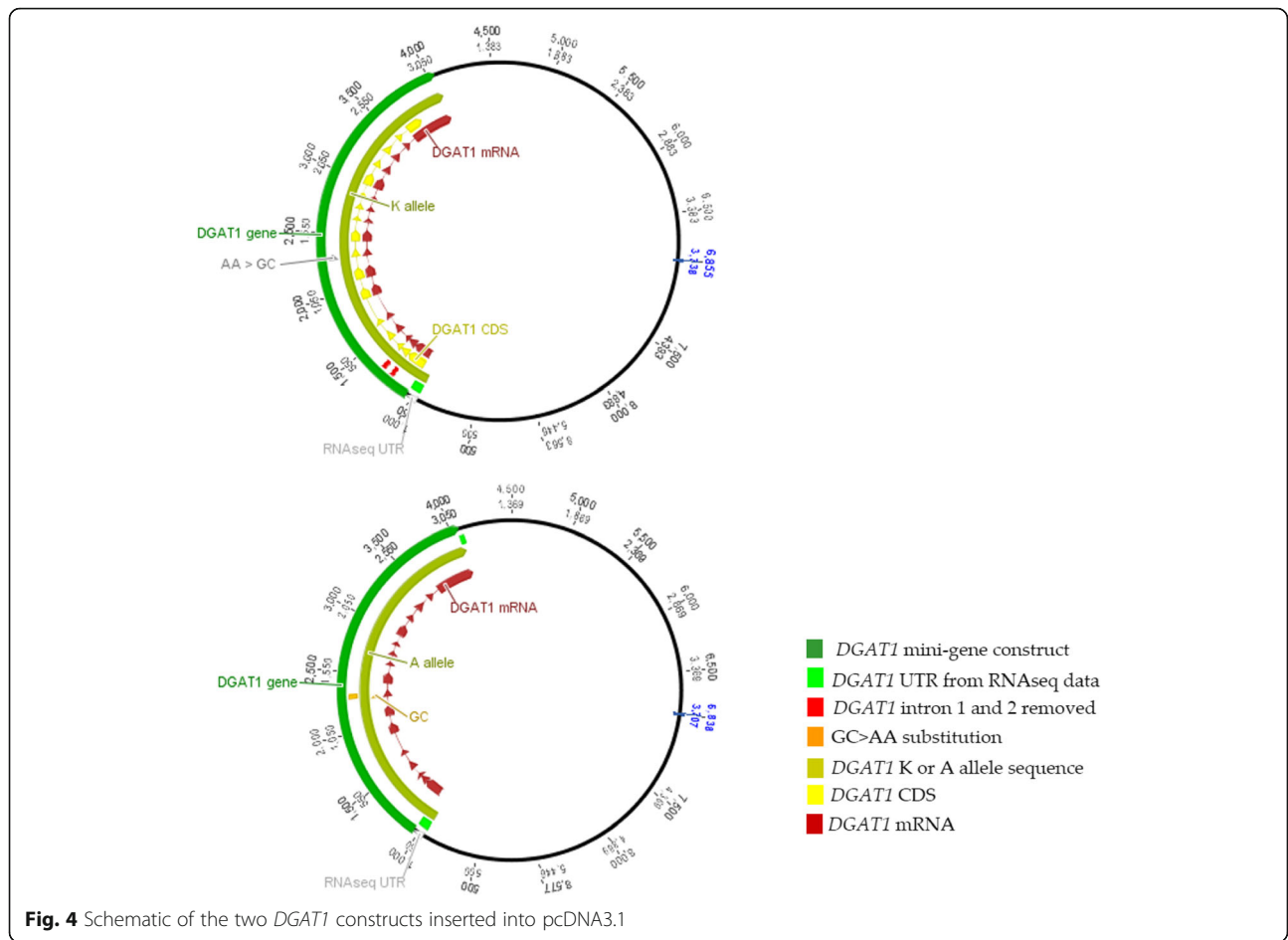
Fig. 4 Schematic of the two *DGAT1* constructs inserted into pcDNA3.1

Table 2 Measurements of splicing efficiency at *DGAT1* introns 3, 5, 7 and 13 in mammary cell culture for the 232 K and 232A *DGAT1* plasmids

Measurement	*DGAT1* Junction	K allele	A allele	*P*-value
Splicing ratio	3	0.253 (±0.009)	0.173 (±0.005)	**9.37×10^{-4}**
	5	143.18 (±10.23)	133.52 (±9.68)	0.256
	7	28.49 (±2.77)	5.55 (±0.74)	**9.05×10^{-11}**
	13	92.90 (±5.86)	90.92 (±5.83)	0.411
Spliced mRNA expression	3	0.111 (±0.002)	0.082 (±0.002)	0.015
	5	0.946 (±0.3238)	0.421 (±0.160)	0.097
	7	0.545 (±0.117)	0.255 (±0.047)	0.009
	13	0.984 (±0.219)	0.549 (±0.233)	0.022
Unspliced mRNA expression	3	0.092 (±0.012)	0.0842 (±0.227)	0.376
	5	0.286 (±0.0415)	0.180 (±0.0350)	0.069
	7	0.216 (±0.030)	0.234 (±0.059)	0.383
	13	0.821 (±0.180)	0.503 (±0.124)	0.035

P-values smaller than the Bonferroni threshold (P = 0.0042) are highlighted using bold type

higher in cells transfected with the K allele when assayed at the intron 7 junction ($P = 0.009$; Fig. 5). The other three junctions also showed numerical increases in the expression of spliced transcripts for the K construct ($P < 0.05$), though these effects were non-significant after adjustment for multiple hypothesis testing (Table 2; Bonferroni threshold $P = 0.0042$). None of the junctions showed significant differences in the mean expression of unspliced transcripts between constructs. As such, the increased yield of spliced mRNAs appeared to derive from post-transcriptional mechanisms, likely as a consequence of enhanced splicing efficiency.
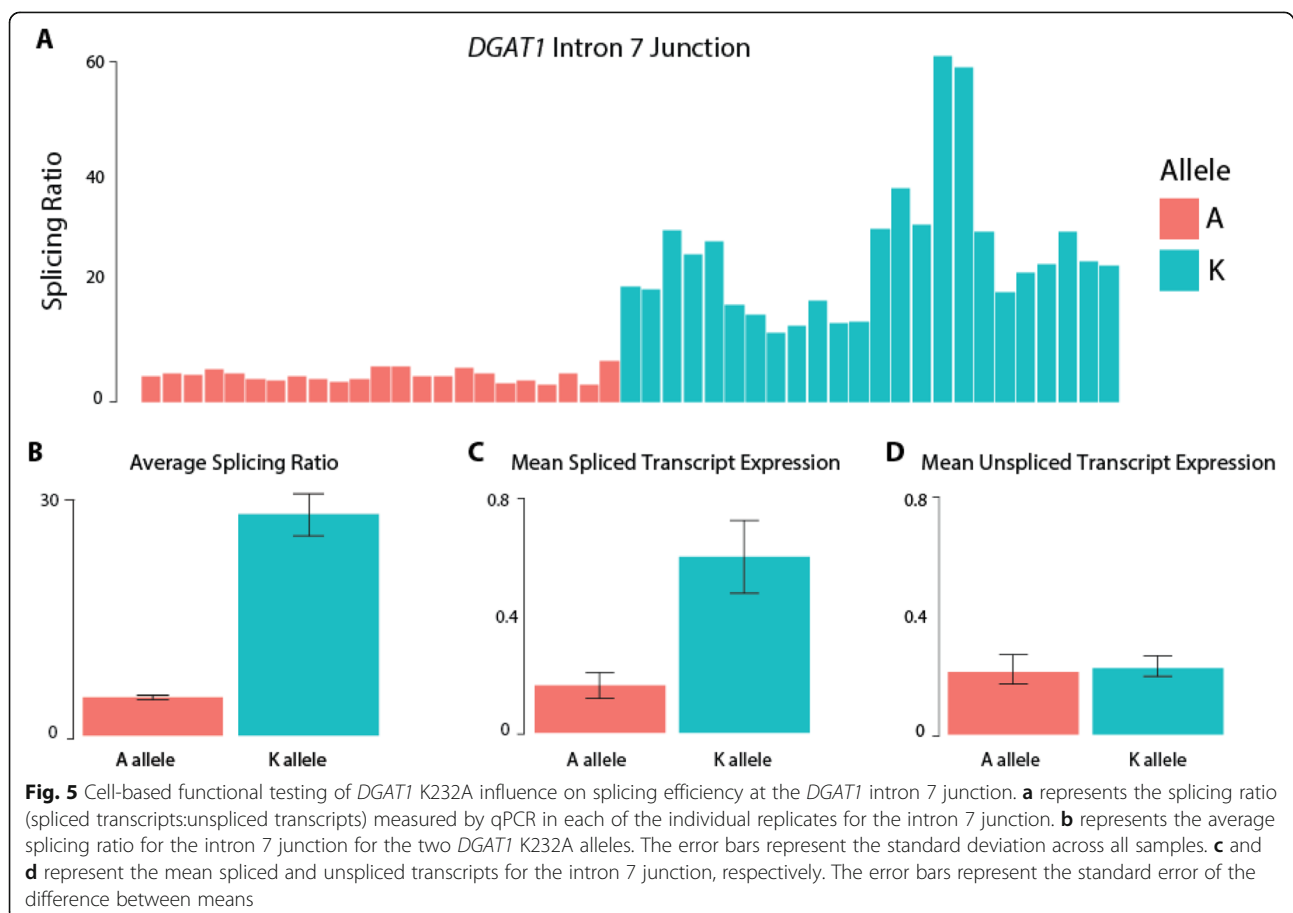
DGAT1 K232A disrupts a putative consensus exon splice enhancer

Given the observation of significant in-vitro splicing effects associated with K232A, we hypothesised that this variant may interrupt one or more exon splice enhancer (ESE) motifs. To examine this possibility, the RESCUE-ESE analysis tool [14] was used to annotate *DGAT1* exon 8 for predicted ESEs. For both *DGAT1* alleles, the first 23 nucleotides of the 5′ end *DGAT1* exon 8 were examined using the default 238 human ESE panel. This analysis revealed the presence of two predicted ESE

motifs in the 5′ end of *DGAT1* exon 8 which both overlap K232A (Supplementary Fig. 1).

Notably, these ESE motifs (AGAAGG and AAGAAG) were located at chr14: 1802263–1,802,268 and chr14: 1802262–1,802,267 (UMD3.1 genome build) respectively, and were only encoded by the K allele. These ESEs were disrupted by the AA>GC MNP (Supplementary Fig. 1). As such, this polymorphism could be expected to disrupt these two ESE motifs when the 232A allele is present (GC). This is concordant with the in vitro analysis, which showed a lower splicing efficiency for the 232A allele. This is also the same allele associated with decreased mammary *DGAT1* expression and milk fat percentage.

Having demonstrated splice enhancement and expression effects for the K232A mutation in vitro, we revisited the potential role of the variant on expression in our mammary RNA-seq dataset. To quantify the proportion of expression variance accounted for by the K232A mutation, this genotype was fitted as a covariate in the association model. The association of the two lead variants was greatly reduced (Fig. 6) in these models, and although the associations were non-significant when applying a Bonferroni correction, they remained significant
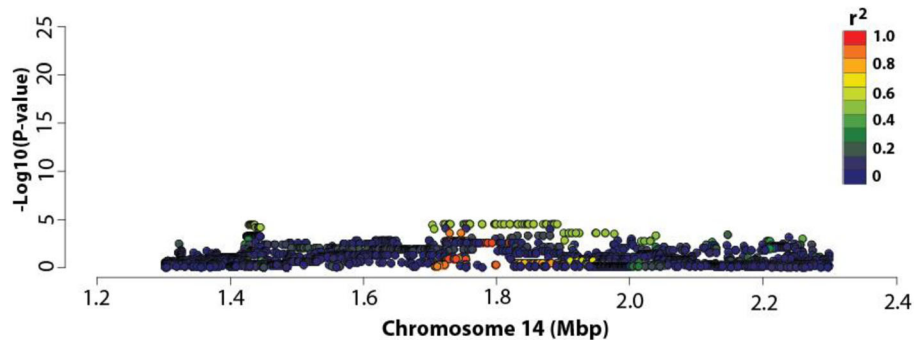

Fig. 5 Cell-based functional testing of *DGAT1* K232A influence on splicing efficiency at the *DGAT1* intron 7 junction. **a** represents the splicing ratio (spliced transcripts:unspliced transcripts) measured by qPCR in each of the individual replicates for the intron 7 junction. **b** represents the average splicing ratio for the intron 7 junction for the two *DGAT1* K232A alleles. The error bars represent the standard deviation across all samples. **c** and **d** represent the mean spliced and unspliced transcripts for the intron 7 junction, respectively. The error bars represent the standard error of the difference between means

Fink *et al. BMC Genomics*   (2020) 21:591

Page 7 of 13

Fig. 6 Expression QTL analysis at the *DGAT1* locus in bovine lactating mammary gland conditioned on K232A. The Manhattan plot for *DGAT1* expression at the *DGAT1* locus in the RNAseq animals ($n = 375$) conditioned on *DGAT1* K232A. The X-axis shows chromosome 14 position (million base pairs); the Y-axis shows $-\log_{10}$ P-values of marker association for the 3218 WGS-derived SNPs in the 1 Mbp interval centred on *DGAT1* K232A. Markers are coloured based on their correlations (R^2) with K232A

in the absence of this correction ($P = 0.000263$; Bonferroni threshold $P = 1.60 \times 10^{-5}$). Of greater note, a cluster of 39 variants in perfect linkage disequilibrium (LD) with each other, but only modestly correlated with K232A ($R^2 = 0.548$) were significant ($P = 3.58 \times 10^{-5}$) in these models. These variants explained 10.32% of the residual phenotypic variance in mammary *DGAT1* expression, suggesting the possibility of another, functionally independent regulatory effect at the locus. Interestingly, some of the markers most highly associated with the residual *DGAT1* eQTL signal reside several kb upstream of the transcription start site of the gene (chr14: 1428907−1,754,446; Supplementary Table 2), representing candidate variants that may impact an additional upstream promoter or other regulatory feature of *DGAT1*. Given these observations, and the other direct and indirect data implicating K232A presented herein, we propose a model that comprises an allelic series of multiple variants underlying the observed expression regulation of the *DGAT1* gene, with a principal role for K232A through exon splice enhancement.

Discussion

A pleiotropic QTL with a large influence on milk composition resides on the centromeric end of bovine chromosome 14, underpinned by the *DGAT1* gene. Although there was initially speculation as to the specific genetic variant and mechanism responsible for the QTL [1, 5], in vitro functional evidence showing that the two protein isoforms of *DGAT1* differed in their ability to synthesise triglycerides [4] led to the now-dominant hypothesis that the K232A amino acid substitution is the causative variant. In the current study, we sought to reassess the potential role of transcriptional regulation as a mechanism underlying the *DGAT1* effects. To this end, we herein provide genetic and functional evidence as to how the *DGAT1* K232A mutation may influence milk

composition through splice-site enhancement and consequent expression-based effects.

We had previously conducted association analysis using RNAseq-derived expression data and revealed a strong eQTL for *DGAT1* in lactating mammary tissue [7]. Importantly, the mammary *DGAT1* cis-eQTL showed a similar genetic signal underpinning the milk production QTLs reported for this locus, with K232A highly associated with the gene expression effect. We have shown in the current study that animals bearing the K allele for *DGAT1* K232A possess greater mammary *DGAT1* expression compared to those animals bearing the A allele. This is of particular note given the K allele is the same allele associated with increased milk fat percentage [1], and more *DGAT1* enzyme as a consequence of increased mRNA could be expected to increase triglyceride synthesis. It should be noted, however, that the K232A variant was not the top associated marker, and that several upstream variants were more highly associated (Supplementary Table 1), suggesting that there may also be a promoter-driven effect.

The finding that the K232A variant is one of the most highly associated genetic markers with *DGAT1* expression is surprising, since previous in vitro investigations using RT-PCR found no difference in *DGAT1* mRNA expression based on the K232A genotype, albeit with limited numbers of animals [4] ($N = 24$). Therefore, the effect of *DGAT1* K232A on milk fat production had been attributed to the enzymatic difference between the two DGAT1 isoforms, as the K allele had the greatest enzymatic activity in vitro, and associated with increased milk fat percentage [4]. Based on that previous study, it has been widely assumed that this enzymatic difference was the sole mechanism driving the effect of K232A on milk composition. In that same study, *DGAT1* K232A was shown to associate with an alternatively spliced transcript of *DGAT1* [4]. This isoform differs based on

its utilisation of an alternative splice donor site 6 bp upstream of K232A, resulting in the 'intronification' of the majority of exon 8 (Fig. 2). The protein encoded from this isoform is predicted to have an internal deletion of 22 amino acids, and is assumed to be non-functional based on its inability to synthesise triacylglycerides in vitro [4]. The proportion of this alternative isoform is approximately 10% of the total *DGAT1* transcripts, and it was shown in the same manuscript that first described this alternative transcript that the K allele results in an increase in its expression in vitro [4]. In line with this observation, the ratio of the alternative isoform to the full-length form differed by K232A genotype in the current dataset, with the animals bearing the K allele producing more of this isoform compared to those animals with the A allele.

Originally, we had hypothesised that the mammary *DGAT1* eQTL might be the result of increased alternative *DGAT1* isoform production. As the alternative isoform results in the intronification of the majority of exon 8, however, we view this as unlikely – since the K allele would present fewer sequence reads mapping to the standard exon definition, yet is the same allele associated with increased mean expression.

Regulatory control of gene expression is most commonly attributed to non-coding sequences; however, regulatory elements can also be part of the coding sequence, and coding variants, such as the dinucleotide substitution underlying *DGAT1* K232A, can influence gene expression through the modulation of auxiliary splicing elements. We hypothesised that the *DGAT1* K232A mutation may overlap one of these elements, and the data presented herein support this hypothesis. Use of the RESCUE-ESE tool to annotate the exonic sequence around K232A suggests two predicted ESE motifs (AG<u>AA</u>GG and AAGA<u>AG</u>) that overlap the K232A polymorphism, with both sequences only encoded by the K allele. Importantly, the K allele is the same allele associated with increased mammary *DGAT1* expression, which suggests a possible mechanism by which *DGAT1* K232A might exert its effect on *DGAT1* splicing, and hence mRNA expression.

The AAGAAG ESE motif has been proposed as the second-most common ESE hexamer in vertebrates [15]. Given the importance of ESEs for promoting splicing, we hypothesised that this motif could influence the splicing efficiency of *DGAT1* pre-mRNA to influence mRNA expression. It has been previously shown in humans that polymorphisms in ESEs can inhibit affinity for splicing factors and affect splicing, leading to altered mRNA and protein translation sequences that contribute to genetic disorders [16]. Additionally, the disruption of splicing has recently been reported for a novel *DGAT1* mutation in dairy cattle, whereby a non-synonymous

A > C transversion in exon 16 disrupts a putative ESE motif and causes the skipping of this exon [17]. This polymorphism results in an enzymatically inactive DGAT1, which in the homozygous state results in a severe phenotype characterised by scouring and slow growth [17].

To investigate the hypothesis that the K232A polymorphism might influence *DGAT1* pre-mRNA processing, we defined a splicing efficiency phenotype to quantitatively measure splicing of intron 8 and other junctions of the gene. Association analysis using these splicing definitions revealed strong splice enhancement for five *DGAT1* introns, providing evidence supporting the mechanism by which this variant might influence mammary *DGAT1* mRNA expression. Critically, the mammary *DGAT1* intron splicing efficiency effects appeared to bear the same genetic signature underpinning the eQTL and milk production QTLs reported for this locus, that is, the association rankings for SNPs were similar for all QTLs. The direction of effects is also consistent with this hypothesis, where animals bearing the K allele have increased milk fat percentage, *DGAT1* expression and efficiency of splicing. Conversely, the animals bearing the A allele showed decreased *DGAT1* expression and splicing efficiency, as well as decreased milk fat percentage.

Splicing efficiency is dependent on a number of factors, with the likelihood of an intron being retained in mature mRNA reflecting the strength of the splice site, intron length, GC content, splicing factor expression and changes in chromatin structure [18]. As such, polymorphisms in ESEs and other splicing elements can influence transcription levels by modifying the strength of the recruitment of the splicing machinery to the junctions in the pre-mRNA transcript [19]. The splicing efficiency effect and increased *alternative* splicing for *DGAT1* suggest that there are a number of weak splice sites in the gene, and the presence of the ESE in the K allele enhances the recruitment of the splicing machinery to increase their usage, resulting in increased splicing of these junctions.

Interestingly, the junctions that had a splicing efficiency phenotype associated with K232A were distributed throughout the gene and included introns 1–3, and 11, which are several kb from the polymorphism and ESE motif. Similar to a recent study [20], the approach taken in this study accounted for any size bias and read coverage differences across the gene and subsequently revealed no relationship between the size of the intron and the splicing efficiency at the junction. The *DGAT1* intron 1 and 2 junctions, which contain the two largest introns, both exhibit a strong splicing efficiency effect, with the intron 2 junction exhibiting the most significant effect in this analysis. An explanation for why particular

Fink *et al. BMC Genomics*   (2020) 21:591

Page 9 of 13

DGAT1 junctions appear to be influenced by K232A genotype while others remain unaffected is unknown at this stage. It is possible that during pre-mRNA processing, the *DGAT1* junctions are processed in an order such that some junctions become rate-limiting steps in the process. If such bottlenecks exist, then the presence of the ESE could influence the efficiency of the processing of the intron 8 junction and the junctions that are subsequently processed. This would result in certain junctions exhibiting a splicing efficiency difference based on the presence or absence of the ESE, while the junctions prior to the bottleneck would remain unaffected. Ultimately, further experiments are required to understand the relationship between the activation of the exon 8 ESE in *DGAT1* and its influence on the splicing efficiency at multiple junctions in the gene. Such experiments could be performed using long RNA-seq reads, generated using technologies such as pacBio, to better elucidate the interactions between splice events at multiple junctions.

Despite the strong association between *DGAT1* K232A and the expression and splicing efficiency phenotypes, there was still some possibility that one or more of these associations were due to LD effects exerted by an unknown *cis* regulatory variant. To more directly probe the function of K232A, mini-gene constructs were generated for the K and A alleles in the absence of native promoter sequence. Differing only by the dinucleotide substitution responsible for K232A, expression testing of these constructs replicated the splicing efficiency effect for a subset of the same junctions implicated in vivo, unequivocally assigning an expression-based mechanism to this variant.

Interestingly, the splicing efficiency effects appeared to result in an increase in spliced mRNA expression, rather than increased expression per se as there was no concomitant increase in unspliced transcripts at the two junctions exhibiting the splicing efficiency phenotype. The lack of increased expression of the unspliced pre-mRNA transcripts may be the result of an increased rate of pre-mRNA processing, and supports the hypothesis that splicing directly impacts mammary expression of mature *DGAT1* mRNA. A previous study [21] reported that many transcripts retain introns, and that these transcripts are retained in the nucleus without undergoing degradation via nonsense-mediated RNA decay. This study also showed that in vitro, these introns appear to be eventually spliced out at a much slower rate than other introns in the same transcript. These observations suggest that incompletely spliced *DGAT1* transcripts may also be spliced eventually, albeit at a slower rate.

While it is not the first time an expression-based effect of *DGAT1* has been proposed as the mechanism by which this gene influences milk composition [17], our study is the first to provide evidence supporting an expression-based effect associated with K232A. Association mapping at the *DGAT1* locus using imputed WGS variants showed that *DGAT1* K232A retains its status as one of the top ranking variants. However, K232A was not the marker with the smallest *P*-value, so the possibility remains that additional effects reside at the locus, or that imperfect sequence imputation or sampling error may have influenced the relative association rankings of the variants in this interval.

To attempt to address these possibilities, further association analysis was conducted to include K232A genotype as a covariate in the models. This analysis removed the majority of the association signal for *DGAT1* expression, suggesting that the *cis*-eQTL could be derived, for the most part, from *DGAT1* K232A. The clusters of highly significant markers in the previous analysis were no longer associated with *DGAT1* expression in these models, suggesting that these variants were tagging the signal from K232A. However, a seemingly distinct, marginally significant eQTL remained, signifying there may be additional effects on mammary *DGAT1* expression. A number of these highly associated markers are located upstream of the transcription start site of the gene, suggesting there may be an additional promoter driven effect on mammary *DGAT1* expression. One possibility is the previously-proposed VNTR polymorphism [5], which was hypothesised to increase the number of putative SP1 transcription factor binding sites, and stimulate an increase in *DGAT1* expression.

While we were able to convincingly demonstrate *cis*-eQTL and splicing efficiency effects at the *DGAT1* locus, an unresolved question is what proportion of the K232A impacts on milk composition are derived from differences in enzymatic activity and alternatively from the expression based effects. One possible option to delineate these two mechanisms would be to use redundant codons to create cell lines that encode identical DGAT1 proteins, yet have alternative ESE-encoding genomic sequences. Unfortunately, however, lysine and alanine amino acids have limited redundancy, precluding the design of such constructs.

Conclusions

A QTL underpinned by the *DGAT1* gene represents one of the most well-known and validated bovine milk composition and production effects, presenting profound impacts on these traits. Despite the effects being long-attributed to an enzymatic mechanism as a consequence of a K232A missense mutation, we have used a large mammary RNAseq dataset in conjunction with in vitro expression experiments to highlight an alternative functional effect of this variant. Our experiments show that *DGAT1* is differentially expressed by QTL genotypes in

Fink *et al. BMC Genomics*   (2020) 21:591

Page 10 of 13

the mammary gland, and confirm a splice enhancement role for the K232A mutation that potentially modulates these effects. Although the relative contribution of splice enhancement and differential enzymatic activities are unknown, these data suggest that the myriad lactation effects attributed to *DGAT1* K232A may, at least in part, derive from an expression-based mechanism.

Methods

A subset of the results in this work were originally published in the PhD thesis of one of the authors [22], and more detail on the methods summarised below can be found therein.

DNA extraction and high throughput genotyping

The animals used for the analysis comprised 375 mostly Holstein-Friesian NZ dairy cows, representing a subset of 406 sequenced animals described in detail previously [8, 23]. Twenty-one of these cows were F_2 animals from an earlier large Friesian-Jersey crossbreeding study [24]. The remaining animals were sampled on Tokanui Research Dairy Farm in the Waikato region of the North Island of New Zealand. Briefly, genomic DNA was extracted from ear-punch tissue or blood by GeneSeek (Lincoln, NE, USA) and processed using Qiagen BioSprints kits (Qiagen). No animals were sacrificed for this study.

The majority of the animals ($n = 354$) were genotyped with the Illumina BovineHD BeadChip by GeneSeek, with the remainder ($n = 21$) genotyped on the Illumina BovineSNP50 panel, and subsequently imputed to the BovineHD panel as described previously [8, 9]. All BovineHD genotypes were then imputed up to whole-genome sequence (WGS) for 3128 variants in the 1 Mbp interval of interest using Beagle v4 [25], with a reference population of 556 animals as described previously [8, 9]. All genotypes were recoded using PLINK [26] (version 1.90b2c) to 0, 1 or 2 to represent the number of alternative alleles for each marker (i.e. 0, 1, and 2 to represent the homozygous reference, heterozygous, and homozygous alternative genotypes, respectively) for use in the genetic association analyses described below.

RNA sequencing

RNA sequencing and informatics was conducted prior to the experiments presented herein, and has also been described elsewhere [8, 27]. The cattle in this dataset existed in three cohorts sampled at different points in time. The first cohort comprised the 21 cattle in the crossbreeding study described above, and were collected in 2004 and 2012, with sequencing conducted by NZGL (Dunedin, New Zealand) using the Illumina HiSeq 2000 instrument. For these samples, libraries were prepared using the TruSeq RNA Sample Prep Kit v2 (Illumina).

RNA sequencing of the remaining two cohorts (183 and 171 samples respectively), collected in 2013, was carried out by the Australian Genome Research Facility (AGRF; Melbourne, Australia) using the Illumina HiSeq 2000 instrument. Libraries for these two cohorts were prepared using the TruSeq Stranded Total RNA Sample Prep Kit (Illumina) with ribosomal depletion using Human/Mouse/Rat Ribo-Zero kit (Epicentre/Illumina). All samples were sequenced using a 100 base-pair (bp) paired-end protocol, with two samples multiplexed per lane.

RNA sequence data representing the 375 animals were mapped to the UMD3.1.1 bovine reference genome using Tophat2 [28] (version 2.0.12), locating an average of 88.9 million read-pairs per sample. Cufflinks software [29] (version 2.1.1) was used to quantify expressed transcripts, and yielded fragments per kilobase of exon model per million mapped (FPKM) expression values. In addition, read counts were determined using HTseq software [30] (version 0.6) and processed using the variance-stabilising transformation (VST) normalisation method in DESeq [13] (version 1.18) and adjusted for batch effects, to derive gene expression phenotypes suitable for linear model analysis, and subsequent eQTL analysis.

To investigate the influence of *DGAT1* K232A on *DGAT1* splicing efficiency, the number of reads mapping to each intron and exon of *DGAT1* was determined using HTSeq 0.6.0 [30], with the intron and exon boundaries specified by the RefSeq annotation (NM_174693.2). The splicing efficiency phenotype for *DGAT1* intron 8 was calculated as the percentage of *DGAT1* RNAseq reads mapping to the intron. The splicing efficiency phenotypes for each individual RefSeq *DGAT1* junction were calculated as the ratio of exonic reads to intronic reads corresponding to the junction (of spliced and unspliced reads, respectively), according to the phenotype defined by Pikielny and Rosbash, 1985 [11]. Reads were considered exonic if they bridged the splicing junction i.e. mapped to the 3′ end of the preceding exon and the 5′ end of the following exon. Reads were considered intronic or unspliced if they mapped to the 3′ end of the preceding exon and through the intron-exon boundary into the intron.

Genetic association analysis

Associations between 3128 SNPs in the 1 Mbp interval surrounding (and including) *DGAT1* K232A and *DGAT1* expression were quantified using pedigree-based mixed models in ASReml-R [31, 32]. Each SNP was fitted in a separate sire-maternal grandsire single trait model, with SNP treated as a quantitative variable based on the number of copies of the alternative allele, and variance components estimated in a restricted maximum-likelihood (REML) framework. Covariates for sequencing cohort,

Fink *et al. BMC Genomics*   (2020) 21:591

Page 11 of 13

the proportions of NZ Holstein-Friesian ancestry, US Holstein-Friesian ancestry, Jersey ancestry and heterosis effects were also included in the models. Association analyses for splicing efficiency were also conducted using ASReml-R with the same set of covariates. This analysis fitted the 3128 SNPs with phenotypes defined as described in the RNA sequencing methods.

The percentage spliced in (PSI) [10, 33] was used to investigate the alternative splicing of *DGAT1* exon 8, which has previously been demonstrated to be associated with the *DGAT1* K232A genotype [4]. The genetic coordinates for the alternative form of *DGAT1* exon 8 (chr14:1802251–1,802,259 alternative, 1,802,260–1,802, 325 reference) were manually added to the Ensembl gene transfer format (GTF) file that contained gene structures of all the genes in the reference genome. Because the alternative 5′ splice site results in the intronification of 22 amino acids, the PSI was calculated as (number of reference spliced reads) / (number of reference or alternative spliced reads). A Kruskal-Wallis one-way ANOVA was used to test for the influence of the K232A genotype on the alternative splicing of *DGAT1* exon 8, due to the heavily skewed distributions observed for PSI within each genotype class.

In vitro splicing efficiency assay

To test the effect of the K232A on *DGAT1* splicing efficiency in vitro, MAC-T cells [12] (ATCC CRL-10274; RRID:CVCL_U226) were transfected with *DGAT1* minigene constructs containing either the K232 or the A232 allele. The *DGAT1* alleles were based on the reference sequence (Accession number AY065621), and were identical with the exception of the AA>GC MNP that causes the K232A amino acid substitution. The *DGAT1* 5′ UTR was extended by 84 bp to represent the UTR apparent from mammary RNAseq data, and the first two

introns were removed due to constraints on total insert size. Sequences representing the two *DGAT1* isoforms were synthesised and cloned into pcDNA3.1 by GenScript (New Jersey, USA). Co-transfection of cells with pMAXGFP plasmid (Lonza) was conducted in a 1:1 ratio to provide a normalisation control for transfection efficiency.

Cells were plated in 24-well plates and grown for 24 h in proliferation media to achieve approximately 70% confluency. For cell transfection, 0.5 μL Lipofectamine® LTX (Invitrogen) was gently mixed with 25 μL Opti-MEM reduced serum media (Invitrogen). Aliquots containing 375 ng of both *DGAT1* pcDNA3.1 constructs and pMAXGFP plasmid DNA, and 0.5 μL PLUS reagent were diluted in 25 μL Opti-MEM. The diluted plasmids were combined with the Lipofectamine® LTX, gently mixed and incubated at room temperature for 5 min, after which 50 μL transfection mix was added to each well. After 24 h of incubation at 37 °C, the cells were visualised on a Nikon Ti-E inverted light microscope prior to RNA extraction. All experiments were repeated in triplicate in three separate cell preparations from MAC-T passage numbers 9, 10, 11, and 12.

RNA was extracted from each well of a 24-well plate using a TRIzol-based protocol and was subjected to two sequential DNase treatments before quantification. Following DNase treatment, Complementary DNA (cDNA) synthesis was performed using 2.5 μg of RNA as input for each 20 μL reaction. Complementary DNA was diluted 1:10 in Ultra-Pure water (Invitrogen) and used immediately for qPCR or stored at − 20 °C. Serial 5x cDNA dilutions were used to generate standard curves for each real-time PCR assay by pooling 4 μL from each experimental sample. Real-time PCR reactions were carried out in 10 μL volumes in 384-well plate format using standard qPCR cycling conditions for use with the
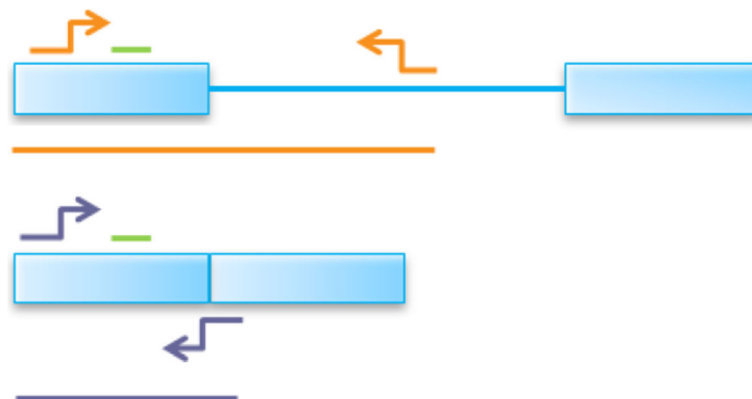

Fig. 7 Schematic of the two RT-qPCR assays for each junction in *DGAT1*. The blue boxes represent exons while the blue line represents the intron. The green line represents the probe, while the orange and purple arrows represent the primers for unspliced and spliced mRNA transcripts, respectively. The first assay quantifies the intron containing pre-mRNA transcripts (orange) while the second assay quantifies the spliced mRNA transcripts (purple). The ratio of mRNA:pre-mRNA transcripts is used to generate a splicing efficiency phenotype for each junction

Fink *et al. BMC Genomics*   (2020) 21:591

Page 12 of 13

LightCycler480® Universal Probe System. Eukaryotic translation initiation factor 3 K (*EIF3K*) was used as an endogenous control gene for normalisation of gene expression [34]. In addition, an assay was designed for the pMAXGFP plasmid as a further control to normalise for transfection efficiency.

To quantify splicing efficiency at the intron-exon junctions in *DGAT1*, PCR assays were designed using Universal Probe Library and Primer 3 software, producing two assays for each junction (Fig. 7). These assays were designed such that they had a common probe, plus one common exonic primer (either forward or reverse). The expression of spliced mRNA transcripts was measured using a second primer that bound to spliced exon/exon sequence across the junction being measured. The expression of unspliced mRNA transcripts was measured using a second primer that bound to the intron corresponding to the junction being measured. The average expression of each transcript across triplicate wells was calculated relative to the geometric mean of expression for the reference gene assays for each sample. Supplementary Table 3 lists the primers and probes used these experiments. To calculate the splicing ratio for each junction for each sample, the average expression of the spliced transcripts was divided by that of the unspliced transcripts. Student's t-test was used to determine the statistical significance of differences between the two alleles.

Supplementary information

Additional file 1 Supplementary Figure 1. Schematic of the 5′ end of *DGAT1* exon 8 with ESE motifs overlapping the K232A amino acid substitution. **Supplementary Table 1**. Mammary *DGAT1* expression association statistics for top WGS-derived variants. **Supplementary Table 2**. Mammary *DGAT1* expression association statistics for top sequence variants conditioned on *DGAT1* K232A. **Supplementary Table 3**. Primer sequences and assay design for RT-qPCR of *DGAT1* introns 3, 5, 7, and 13.

Abbreviations

DGAT1: Diacylglycerol O-acyltransferase 1; ESE: Exon splice enhancer; eQTL: Expression quantitative trait locus; NZ: New Zealand; QTL: Quantitative trait locus; VNTR: Variable number tandem repeat

Authors' contributions

Conceived and designed experiments: TF, TL, MDL, RGS, RJS, SRD. Performed the experiments: TF, MDL, KT, TL, RH. Contributed analysis tools: KT, TL, TJ, RH. Performed genotype imputation: KT, TJ. Analysed the data: TF, TL, MDL. Wrote the manuscript: TF, MDL, TL. All authors have read and approved the manuscript.

Availability of data and materials

All RNA-seq derived splicing phenotypes and associated genotypes, along with AsReml output files, are available in the Dryad data repository (title: A new mechanism for a familiar mutation – bovine DGAT1 K232A modulates gene expression through multi-junction exon splice enhancement; https://doi.org/10.5061/dryad.rn8pk0p6k). BAM files containing RNA-seq data for the *DGAT1* region of 375 animals are available in the sequence read archive (SRA) with project number SRP254673 (title: RNA sequencing of the DGAT1 gene region; PRJNA616429). Sequence data was aligned to the UMD3.1.1 reference genome (GCA_000003055.5). The sequence of the reference *DGAT1* isoform is available at accession AY065621 in the NCBI Nucleotide database.

Ethics approval and consent to participate

All animal experiments were conducted in strict accordance with the rules and guidelines outlined in the New Zealand Animal Welfare Act 1999. For the mammary tissue biopsy experiment, samples were obtained in accordance with protocols approved by the Ruakura Animal Ethics Committee, Hamilton, New Zealand (approval AEC 12845). These cows were situated on Tokanui Research Dairy Farm and written permission was sought and obtained to biopsy mammary tissue from the owner of these animals (AgResearch, NZ). All other data were generated as part of routine commercial activities and were therefore outside the scope for formal committee assessment and ethical approval (as defined by the above guidelines).

Consent for publication

Not applicable.

Competing interests

MDL, KT, TL, TJ, RJS and SRD are employees of Livestock Improvement Corporation, a commercial provider of bovine germplasm. The authors declare that no other potential competing interests exist.

Author details

[1]School of Biological Sciences, University of Auckland, Auckland, New Zealand. [2]Livestock Improvement Corporation, Hamilton, New Zealand. [3]AI Rae Centre, Massey University, Hamilton, New Zealand.

References

1. Grisart B. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on Milk yield and composition. Genome Res. 2002;12:222–31. https://doi.org/10.1101/gr.224202.
2. Schennink A, Stoop WM, Visker MHPW, Heck JML, Bovenhuis H, Van Der Poel JJ, et al. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. Anim Genet. 2007;38:467–73. https://doi.org/10.1111/j.1365-2052.2007.01635.x.
3. Mayorek N, Grinstein I, Bar-Tana J. Triacylglycerol synthesis in cultured rat hepatocytes. The rate-limiting role of diacylglycerol acyltransferase. Eur J Biochem. 1989;182:395–400. https://doi.org/10.1111/j.1432-1033.1989.tb14844.x.
4. Grisart B, Farnir F, Karim L, Cambisano N, Kim J-J, Kvasz A, et al. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc Natl Acad Sci. 2004;101:2398–403. https://doi.org/10.1073/pnas.0308518100.
5. Kühn C, Thaller G, Winter A, Bininda-Emonds ORP, Kaupe B, Erhardt G, et al. Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on Milk fat content in cattle. Genetics. 2004;167:1873–81. https://doi.org/10.1534/genetics.103.022749.

Fink *et al. BMC Genomics*   (2020) 21:591

Page 13 of 13

6. Fürbass R, Winter A, Fries R, Kühn C. Alleles of the bovine DGAT1 variable number of tandem repeat associated with a milk fat QTL at chromosome 14 can stimulate gene expression. Physiol Genomics. 2006;25:116–20. https://doi.org/10.1152/physiolgenomics.00145.2005.

7. Lopdell TJ, Tiplady K, Struchalin M, Johnson TJJ, Keehan M, Sherlock R, et al. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. BMC Genomics. 2017;18: 968. https://doi.org/10.1186/s12864-017-4320-3.

8. Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, et al. Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine Milk composition. Sci Rep. 2016;6:25376. https://doi.org/10.1038/srep25376.

9. Jivanji S, Worth G, Lopdell TJ, Yeates A, Couldrey C, Reynolds E, et al. Genome-wide association analysis reveals QTL and candidate mutations involved in white spotting in cattle. Genet Sel Evol. 2019;51:62. https://doi.org/10.1186/s12711-019-0506-2.

10. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–6. https://doi.org/10.1038/nature07509.

11. Pikielny CW, Rosbash M. mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. Cell. 1985;41:119–26. https://doi.org/10.1016/0092-8674(85)90066-2.

12. Huynh HT, Robitaille G, Turner JD. Establishment of bovine mammary epithelial cells (MAC-T): an in vitro model for bovine lactation. Exp Cell Res. 1991;197:191–9. https://doi.org/10.1016/0014-4827(91)90422-Q.

13. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.

14. Fairbrother WG. Predictive Identification of Exonic Splicing Enhancers in Human Genes. Science (80-). 2002;297: 1007–1013. doi:https://doi.org/10.1126/science.1073774.

15. Mersch B, Gepperth A, Suhai S, Hotz-Wagenblatt A. Automatic detection of exonic splicing enhancers (ESEs) using SVMs. BMC Bioinform. 2008;9:369. https://doi.org/10.1186/1471-2105-9-369.

16. Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 2015;7:45. https://doi.org/10.1186/s13073-015-0168-9.

17. Lehnert K, Ward H, Berry SD, Ankersmit-Udy A, Burrett A, Beattie EM, et al. Phenotypic population screen identifies a new mutation in bovine DGAT1 responsible for unsaturated milk fat. Sci Rep. 2015;5:8484. https://doi.org/10.1038/srep08484.

18. Wong JJ-L, Au AYM, Ritchie W, Rasko JEJ. Intron retention in mRNA: no longer nonsense. BioEssays. 2016;38:41–9. https://doi.org/10.1002/bies.201500117.

19. Ge Y, Porse BT. The functional consequences of intron retention: alternative splicing coupled to NMD as a regulator of gene expression. BioEssays. 2014; 36:236–43. https://doi.org/10.1002/bies.201300156.

20. Ni T, Yang W, Han M, Zhang Y, Shen T, Nie H, et al. Global intron retention mediated gene regulation during CD4+ T cell activation. Nucleic Acids Res. 2016;44:6817–29. https://doi.org/10.1093/nar/gkw591.

21. Boutz PL, Bhutkar A, Sharp PA. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev. 2015;29:63–80. https://doi.org/10.1101/gad.247361.114.

22. Fink TA. Functional and genetic characterisation of major effect Milk composition variants in dairy cattle [internet]. University of Auckland 2017. Available: https://researchspace.auckland.ac.nz/handle/2292/37394.

23. Littlejohn MD, Henty KM, Tiplady K, Johnson T, Harland C, Lopdell T, et al. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nat Commun. 2014;5:5861. https://doi.org/10.1038/ncomms6861.

24. Spelman RJ, Hooper JD, Stanley G, Kayis SA, Harcourt S. Friesian Jersey crossbred trial: generating phenotypes for the discovery of quantitative trait loci. Proc New Zeal Soc Anim Prod. 2004;64:92–5 Available: http://www.nzsap.org/proceedings/2004/friesian-jersey-crossbred-trial-generating-phenotypes-discovery-quantitative-trait.

25. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84:210–23. https://doi.org/10.1016/j.ajhg.2009.01.005.

26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.

27. Littlejohn MD, Tiplady K, Lopdell T, Law TA, Scott A, Harland C, et al. Expression Variants of the Lipogenic AGPAT6 Gene Affect Diverse Milk Composition Phenotypes in *Bos taurus*. Moore S, editor. PLoS One. 2014;9: e85757. doi:https://doi.org/10.1371/journal.pone.0085757.

28. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36. https://doi.org/10.1186/gb-2013-14-4-r36.

29. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc. 2012;7:562–78. https://doi.org/10.1038/nprot.2012.016.

30. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9. https://doi.org/10.1093/bioinformatics/btu638.

31. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml user guide release 3.0 [internet]. Hemel Hempstead, HP1 1ES. UK: VSN International Ltd; 2009. Available: www.vsni.co.uk.

32. Gilmour AR, Thompson R, Cullis BR. Average Information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics. 1995;51: 1440. doi:https://doi.org/10.2307/2533274.

33. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). Curr Protoc Hum Genet. 2015;87:11.16.1–11.16.14. https://doi.org/10.1002/0471142905.hg1116s87.

34. Grala TM, Phyn CVC, Kay JK, Rius AG, Littlejohn MD, Snell RG, et al. Temporary alterations to milking frequency, immediately post-calving, modified the expression of genes regulating milk synthesis and apoptosis in the bovine mammary gland. Proceedings of the New Zealand Society of Animal Production. Invercargill: NZ: New Zealand Society of Animal Production; 2011. p. 3–8. Available: http://www.nzsap.org/proceedings/2011/temporary-alterations-milking-frequency-immediately-post-calving-modified.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.