

1 **Genome variants associated with RNA splicing variation in bovine are**  
2 **extensively shared between tissues**

3 Ruidong Xiang<sup>1,2,\*</sup>, Ben J. Hayes<sup>2,3</sup>, Christy J. Vander Jagt<sup>2</sup>, Iona M. MacLeod<sup>2</sup>, Majid  
4 Khansefid<sup>2</sup>, Phil J. Bowman<sup>2,6</sup>, Zehu Yuan<sup>2</sup>, Claire P. Prowse-Wilkins<sup>2</sup>, Coralie M. Reich<sup>2</sup>,  
5 Brett A. Mason<sup>2</sup>, Josie B. Garner<sup>4</sup>, Leah C. Marett<sup>4</sup>, Yizhou Chen<sup>5</sup>, Sunduimijid Bolormaa<sup>2</sup>,  
6 Hans D. Daetwyler<sup>2,6</sup>, Amanda J. Chamberlain<sup>2</sup> and Michael E. Goddard<sup>1,2</sup>

7 <sup>1</sup>Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC 3010,  
8 Australia. <sup>2</sup>Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083,  
9 Australia. <sup>3</sup>Queensland Alliance for Agriculture and Food Innovation, Centre for Animal  
10 Science, University of Queensland, St. Lucia, QLD 4067, Australia. <sup>4</sup>Agriculture Victoria,  
11 Dairy Production Science, Ellinbank, VIC, 3821, Australia. <sup>5</sup>Elizabeth Macarthur  
12 Agricultural Institute, New South Wales Department of Primary Industries, Camden, NSW  
13 2570, Australia. <sup>6</sup>School of Applied Systems Biology, La Trobe University, Bundoora, VIC  
14 3083, Australia.

15 \*Corresponding author: Dr. Ruidong Xiang;

16 Ruidong Xiang: [ruidong.xiang@unimelb.edu.au](mailto:ruidong.xiang@unimelb.edu.au);

17 Ben J. Hayes: [Ben.Hayes@ecodev.vic.gov.au](mailto:Ben.Hayes@ecodev.vic.gov.au);

18 Christy J. Vander Jagt: [Christy.VanderJagt@ecodev.vic.gov.au](mailto:Christy.VanderJagt@ecodev.vic.gov.au);

19 Iona M. MacLeod: [Iona.MacLeod@ecodev.vic.gov.au](mailto:Iona.MacLeod@ecodev.vic.gov.au);

20 Majid Khansefid: [Majid.Khansefid@ecodev.vic.gov.au](mailto:Majid.Khansefid@ecodev.vic.gov.au);

21 Phil J. Bowman: [P.Bowman@latrobe.edu.au](mailto:P.Bowman@latrobe.edu.au);

22 Zehu Yuan [zehu.yuan@ecodev.vic.gov.au](mailto:zehu.yuan@ecodev.vic.gov.au);

23 Claire P. Prowse-Wilkins: [Claire.Prowse-Wilkins@ecodev.vic.gov.au](mailto:Claire.Prowse-Wilkins@ecodev.vic.gov.au);

24 Coralie M. Reich: [Coralie.Reich@ecodev.vic.gov.au](mailto:Coralie.Reich@ecodev.vic.gov.au);

- 25 Brett A. Mason: [Brett.Mason@ecodev.vic.gov.au](mailto:Brett.Mason@ecodev.vic.gov.au);
- 26 Josie B. Garner: [Josie.Garner@ecodev.vic.gov.au](mailto:Josie.Garner@ecodev.vic.gov.au);
- 27 Leah C. Marett: [Leah.Marett@ecodev.vic.gov.au](mailto:Leah.Marett@ecodev.vic.gov.au);
- 28 Yizhou Chen: [yizhou.chen@dpi.nsw.gov.au](mailto:yizhou.chen@dpi.nsw.gov.au);
- 29 Sunduimijid Bolormaa: Bolormaa.[Sunduimijid@ecodev.vic.gov.au](mailto:Sunduimijid@ecodev.vic.gov.au);
- 30 Hans D. Daetwyler: [Hans.Daetwyler@ecodev.vic.gov.au](mailto:Hans.Daetwyler@ecodev.vic.gov.au);
- 31 Amanda J. Chamberlain: [Amanda.Chamberlain@ecodev.vic.gov.au](mailto:Amanda.Chamberlain@ecodev.vic.gov.au);
- 32 Michael E. Goddard: [Mike.Goddard@ecodev.vic.gov.au](mailto:Mike.Goddard@ecodev.vic.gov.au)

## 33 **Abstract**

34 **Background:** Mammalian phenotypes are shaped by numerous genome variants, many of  
35 which may regulate gene transcription or RNA splicing. To identify variants with regulatory  
36 functions in cattle, an important economic and model species, we used sequence variants to  
37 map a type of expression quantitative trait loci (expression QTLs) that are associated with  
38 variations in the RNA splicing, i.e., sQTLs. To further the understanding of regulatory  
39 variants, sQTLs were compared with other two types of expression QTLs, 1) variants  
40 associated with variations in gene expression, i.e., geQTLs and 2) variants associated with  
41 variations in exon expression, i.e., eeQTLs, in different tissues.

42 **Results:** Using whole genome and RNA sequence data from four tissues of over 200 cattle,  
43 sQTLs identified using exon inclusion ratios were verified by matching their effects on  
44 adjacent intron excision ratios. sQTLs contained the highest percentage of variants that are  
45 within the intronic region of genes and contained the lowest percentage of variants that are  
46 within intergenic regions, compared to eeQTLs and geQTLs. Many geQTLs and sQTLs are  
47 also detected as eeQTLs. Many expression QTLs, including sQTLs, were significant in all  
48 four tissues and had a similar effect in each tissue. To verify such expression QTL sharing  
49 between tissues, variants surrounding ( $\pm 1$ Mb) the exon or gene were used to build local  
50 genomic relationship matrices (LGRM) and estimated genetic correlations between tissues.  
51 For many exons, the splicing and expression level was determined by the same *cis* additive  
52 genetic variance in different tissues. Thus, an effective but simple-to-implement meta-  
53 analysis combining information from three tissues is introduced to increase power to detect  
54 and validate sQTLs. sQTLs and eeQTLs together were more enriched for variants associated  
55 with cattle complex traits, compared to geQTLs. Several putative causal mutations were

56 identified, including an sQTL at Chr6:87392580 within the 5<sup>th</sup> exon of kappa casein (*CSN3*)  
57 associated with milk production traits.

58 **Conclusions:** Using novel analytical approaches, we report the first identification of  
59 numerous bovine sQTLs which are extensively shared between multiple tissue types. The  
60 significant overlaps between bovine sQTLs and complex traits QTL highlight the  
61 contribution of regulatory mutations to phenotypic variations.

62 **Keywords:** RNA splicing, gene expression, expression QTL, sQTL, genetic correlations,  
63 local genomic relationship matrices (LGRM), transcriptome meta-analysis, bovine

## 64 **Background**

65 Cattle are an important source of meat and dairy products for humans worldwide. Also, cattle  
66 can be used as clinical models to study genetic causes of human diseases [1]. To improve  
67 productivity, health performance and efficiency of cattle, traditional selective breeding has  
68 been widely used. In the last decade, genomic selection, originally developed in cattle  
69 breeding, has further increased the rate of genetic improvement of complex traits in all  
70 livestock species [2, 3]. However, genomic selection commonly uses genotyping arrays that  
71 are based on single nucleotide polymorphisms (SNPs) of which very few have known  
72 biological functions or directly impact genetic variation in production traits. Knowledge of  
73 the genes involved and polymorphic sites would increase our understanding of the biology  
74 and may further increase the rate of genetic improvement [4].

75 Many of the sequence variants that are associated with complex traits (quantitative trait loci  
76 or QTL) are not coding variants and are presumed to influence the regulation of gene  
77 expression, that is to be expression QTLs [5]. An expression QTL might be associated with  
78 the variation in overall transcript abundance from the gene, which we will refer to as a gene  
79 expression QTL or geQTL. In cattle and humans, geQTLs show significant enrichments for  
80 mutations associated with diseases and complex traits [5-7].

81 After transcription, RNA is spliced by intron removal and exon ligation to create various  
82 mature transcripts. An expression QTL associated with the changes in the expression ratio of  
83 an exon to the gene implies that it alters RNA splicing. This type of expression QTL is then  
84 defined as a splicing QTL or sQTL, which have been studied in humans by inferring the  
85 individual splicing ratio from RNA sequence data [8]. More recently, sQTLs, identified using  
86 intron information extracted from RNA sequence data, were demonstrated to have  
87 fundamental links with human diseases [9, 10]. RNA splicing also results in different

88 expression levels of exons within a gene. Thus, in theory, the type of expression QTL that  
89 change the level of expression of one or several exons, i.e., exon expression QTL or eeQTLs,  
90 may represent some sQTLs. However, the extent to which eeQTLs overlap with sQTLs  
91 and/or geQTLs remains unclear, at least in cattle.

92 Knowledge of large mammal regulatory mutations is limited mainly to humans, where there  
93 have been multiple studies reporting on expression QTLs [9, 11, 12]. In this study, we aim to  
94 identify bovine cis splicing QTLs using the abundances of genes, exons and introns estimated  
95 from RNA sequence data from hundreds of animals and multiple tissues along with imputed  
96 whole genome sequences. We examined the extent to which sQTLs can be detected in white  
97 blood cells, milk cells, liver and muscle transcriptomes and the extent to which sQTLs  
98 overlap with conventional QTL associated with complex traits overlapped. To further  
99 characterise the features of sQTLs, we used the counts of genes and exons to map another  
100 two types of cis expression QTLs: eeQTLs and geQTLs, and then analysed their relationships  
101 with sQTLs in different tissues.

102

## 103 **Results**

### 104 *Data quality*

105 In total, we analysed 378 transcriptomes of 19 tissue types from 214 cattle generated from  
106 four experiments covering major dairy and beef cattle breeds (Table 1, Figure 1a and  
107 Additional file 1: Supplementary Methods). Following recommendations from Mazzoni and  
108 Kadarmideen 2016 [13] RNA sequence quality was assessed and was detailed in Additional  
109 file 1: Supplementary Methods. Based on the results produced by Qualimap 2 [14], no  
110 significant events of RNA degradation were observed in all studied tissues (Additional file 2:

111 Supplementary Figure S1). Also, according to Qualimap 2 [14], on average, 60.3% of reads  
112 were mapped to exonic regions of the bovine reference genome (UMD3.1), 15.4% of reads  
113 were mapped to intronic regions and 24.4% were mapped to intergenic regions (Additional  
114 file 3: Supplementary Table S1). Splicing junction annotation and saturation were estimated  
115 using RSeQC [15]. As a small demonstration, no significant difference was observed in the  
116 annotated splicing junction events in the bovine reference genome (UMD3.1), using different  
117 RNA-seq alignment software including HISAT2 [16], STAR [17] and TopHat2 [18]  
118 (Additional file 4: Supplementary Table S2). Although HISAT2 and STAR outperformed  
119 TopHat2 for novel splicing junction events (do not exist in the current bovine UMD3.1  
120 genome). Also, using a splicing junction saturation analysis for all tissues we observed  
121 saturated coverage for the known splicing junctions (Additional file 2: Supplementary Figure  
122 S2), though there appeared to be more potential for splicing junction discovery for the novel  
123 category.

124 Animals with white blood cell RNA-seq data were evaluated for the consistency between  
125 imputed genotypes from the 1000 bull genomes project [19] and RNA sequence genotypes as  
126 predicted from the RNA sequence data using samtools [20]. On average, the concordance  
127 between imputed sequence genotypes and RNA sequence genotypes was 0.943 (Additional  
128 file 2: Supplementary Figure S3), which was consistent with the average imputation accuracy  
129 (0.926) of the 1000 bull genomes project [21]. The comparison of the genotypes was detailed  
130 in Additional file 1: Supplementary Methods.

131 Overall, samples from the same or similar tissues clustered together rather than clustering by  
132 experiments, based on exon expression levels (Figure 1a). This was supported by further  
133 analyses of the clusters where ellipses, drawn based on tissue types, were clearly separated  
134 (Additional file 2: Supplementary Figure S4a,c,e), whereas ellipses drawn based on different  
135 experiments overlapped (Additional file 2: Supplementary Figure S4b,d,f), at the confidence

136 interval = 0.95 [22]. Consistent with previous reports [23], milk cells and mammary gland  
 137 transcriptomes were closely related.

138

Table 1. Summary of experiments, data and analyses. Tissue splicing: variation in differential splicing associated with tissue types estimated using RNA sequence data from all experiments. Breed splicing: variation in differential splicing associated with Holstein and Jersey breeds estimated using RNA sequence data of milk cells from experiment III. sQTLs: cis splicing quantitative trait loci, sQTLs estimated using RNA sequence data and imputed whole genome and from experiment III and IV. Data from experiment III and IV were also used to estimate exon expression eeQTLs and gene expression geQTLs.

Tissue splicing	Breed splicing	sQTLs	Experiment	Tissue type	Sample No.	Breed	Individual No.
✓			I	18 various <sup>a</sup>	54	Holstein	1
✓			II	milk cells & mammary	12	Holstein	6
✓		✓	III	white blood cells	105	Holstein	105 <sup>b</sup>
✓	✓	✓	III	milk cells	131	Holstein & Jersey	105 <sup>b</sup> & 26
✓		✓	IV	liver	35	Angus	35
✓		✓	IV	muscle	41	Angus	41

<sup>a</sup>: 18 tissues from [24]. <sup>b</sup>: The same 105 Holstein cattle, each of which had both white blood and milk cell transcriptomic data.

139

#### 140 *Differential splicing between tissues and breeds*

141 We primarily defined a differentially spliced gene as a gene which contained exons whose  
 142 inclusion ratios (exon expression divided by gene expression) were significantly associated  
 143 with tissue or breed (FDR<0.1). To verify the significantly spliced exons, we imposed a  
 144 requirement that at least one adjacent intron had an excision ratio [9, 25] that was also  
 145 significantly (FDR<0.1) associated with tissue or breed (See methods). The FDR threshold of  
 146 such exon splicing events was considered as approximately 0.01 by combining the FDR  
 147 thresholds from exon and intron analyses (0.1 × 0.1). The overlaps of genes displaying  
 148 differential splicing from exon and intron analyses were shown and examined in 2 × 2 tables



149 by Chi-square tests (Additional file 5: Supplementary Table S3). Overall, the overlap of the  
150 results from exon and intron analyses was significantly more than expected by random  
151 chance.

152 Using data from all experiments (Table 1), there were 8,657 genes in which at least one exon  
153 had the variation in splicing associated with differences between tissue types. A list of these  
154 genes with the significances of differential splicing for the exons and introns was shown in  
155 Additional file 6: Supplementary Table S4. The top 10% of these significantly differentially  
156 spliced genes had a GO term enrichment (FDR<0.01) of ‘regulation of cellular process’,  
157 suggesting very general roles of these genes in cell function. There were 148 genes with  
158 significant differential splicing events in the milk cell transcriptome between breeds (Table 1,  
159 Additional file 7: Supplementary Table S5). While these genes did not show any significant  
160 GO term enrichments, they included the milk protein gene *CSN3* [26, 27], where the 5<sup>th</sup> exon  
161 (6:87,392,578-87,392,750) was more commonly included in the transcript in Holstein cattle  
162 than in Jersey cattle (Figure 1b).

### 163 *cis* splicing quantitative trait loci (sQTLs)

164 The mapping of sQTLs was based on data from 312 transcriptomes generated from  
165 experiments III and IV, including white blood cells, milk cells, liver and muscle tissues  
166 (Table 1). In total 207 individuals had imputed whole genome sequence data and in  
167 experiment III, 105 genotyped cattle had both white blood and milk cell transcriptome data  
168 (Table 1). Similar to differential splicing analyses described above, a *cis*-acting sQTL was  
169 defined as a SNP significantly (FDR<0.1) associated with the variation in the inclusion ratio  
170 of the exon (up to 1Mb away) and significantly (FDR<0.1) associated with the variation in at  
171 least one adjacent introns’ excision ratio [9, 25]. When analysed separately, the overlap  
172 between sQTLs found by exon analyses and sQTLs found by intron analyses were

173 significantly more than expected by random chance (Additional file 5: Supplementary Table  
174 S3). After requiring that the variation in the inclusion and excision ratios for adjacent exons  
175 and introns both be associated with the same SNP, 138,796 sQTLs were called in the white  
176 blood cells, 28,907 sQTLs were called in the milk cells, 11,544 sQTLs were called in the  
177 liver tissue and 5,783 sQTLs were called the muscle tissue (Figure 2, Additional file 5, 8:  
178 Supplementary Table S3 and S6).

179 The significant sQTLs in white blood and milk cells were mapped to 929 and 283 genes,  
180 respectively (Table 2). Many SNPs were significant for sQTLs due to linkage disequilibrium  
181 between SNPs close to the same gene. The results do not imply many sQTLs per gene.

182 In the milk cell transcriptome, the fifth exon of *CSN3* (6:87392578-87392750), which as  
183 described above was differentially spliced in Holstein and Jersey cattle (Figure 1b), and was  
184 strongly associated with an sQTL (Chr6:87392580,  $p = 5.0e-07$ , Additional file 8:  
185 Supplementary Table S6). This sQTL is physically located within the 5<sup>th</sup> exon of *CSN3*. Also,  
186 the B allele of this sQTL increased the expression and inclusion ratio of the 5<sup>th</sup> exon and had  
187 a higher allele frequency among Holstein cattle than Jersey cattle (0.79 vs 0.02). This  
188 predicted that the expression and inclusion ratio of the 5<sup>th</sup> exon would be significantly higher  
189 in the Holstein cattle than Jersey cattle, which was in line with the observations in Figure 1b.  
190 In addition, this sQTL was also predicted to be a splice site ('splice\_region\_variant') by  
191 Ensemble [28] and NGS-SNP software [29]. Much smaller numbers of significant sQTLs  
192 were detected in liver (11,544 SNPs) and muscle (5,783 SNPs) (Figure 2c,d). This was  
193 probably due to the smaller sample size (Table 1) and lower sequence depth of liver and  
194 muscle from experiment IV (Additional file 1: Supplementary Methods) than that of white  
195 blood and milk cells from the experiment III (Figure 2a,b, Table 2).

Table 2. Summary of expression QTLs detected. cis sQTLs: significant SNPs within  $\pm$  1Mb of the exon, associated with the variation in its inclusion ratio and also associated with the variations in the excision of an adjacent intron at the same significance level. Where the FDR threshold was approximated to 0.01 by combining FDR thresholds used in exon (FDR < 0.1) and intron (FDR < 0.1) analyses. cis eeQTLs: significant (FDR<0.01) SNPs within  $\pm$  1Mb of the exon, associated with the variation in its abundance. cis geQTLs: significant (FDR<0.01) SNPs within  $\pm$  1Mb of the gene associated with the variation in its abundance.

cis QTL type	Tissue	Expression QTL number	Gene number
sQTLs	white blood cells	138,796	929
	milk cells	28,907	283
	liver	11,544	49
	muscle	5,783	76
eeQTLs	white blood cells	802,685	6,446
	milk cells	100,844	2,102
	liver	37,322	346
	muscle	64,675	1,267
geQTLs	white blood cells	96,530	842
	milk cells	4,099	99
	liver	39,306	419
	muscle	57,054	1,150

196

197 *Comparing sQTLs with exon expression eeQTLs and gene expression geQTLs*

198 Many more significant eeQTLs than sQTLs were detected in all tissues studied (Table 2). In  
 199 white blood and milk cells, the number of geQTL was smaller than the number of significant  
 200 sQTLs in white blood and milk cells (Table 2).

201 Figure 3a showed that sQTLs were a median distance of about 200 kb from the transcription  
 202 start site (TSS) and were slightly closer to the TSS than eeQTLs and geQTLs. All 3 classes of  
 203 expression QTLs had a lower percentage of intergenic SNPs and a higher percentage of  
 204 intronic and coding SNPs, including splice sites than the same categories across all SNPs  
 205 analysed (Figure 3b). Specifically, sQTLs had the highest percentage of intronic SNPs,  
 206 compared to eeQTLs and geQTLs. However, no consistent ranking of concentrations of  
 207 ‘Splice’ SNP category for sQTLs, eeQTLs and geQTLs were observed in different tissues  
 208 (Figure 3b).

209 *Shared genetic influences between cis QTL types*

210 Within each tissue, the sharing of SNPs between all three expression QTL types was  
211 significantly more than expected by chance (Figure 4). However, in the white blood and milk  
212 cells, which had relatively large sample size ( $n \geq 105$ , Table 1), the largest absolute amount  
213 of SNP sharing appeared to be between sQTLs and eeQTLs (Figure 4). This was followed by  
214 the amount of SNP sharing between eeQTLs and geQTLs (Figure 4a,b). In liver and muscle  
215 tissue which had relatively small sample size ( $n \leq 41$ ) and low sequencing depth, the largest  
216 absolute amount of SNP sharing was between eeQTLs and geQTLs, followed by the amount  
217 of SNP sharing between sQTLs and eeQTLs (Figure 4c,d).

218

219 To further examine the relationship between sQTLs and eeQTLs, a two by two table of sQTL  
220 and eeQTL counts in white blood and milk cells, which had comparable sample sizes, was  
221 created (Additional file 9: Supplementary Table S7). This suggested that when an sQTL was  
222 found, it was highly likely to be also identified as an eeQTL. For example, of 138,796 sQTLs  
223 found in the white blood cells, 109,155 of them were also blood eeQTLs, but only 21,766 of  
224 them were identified as blood geQTLs. Again, for these 138,796 blood sQTLs, although only  
225 18,005 and 25,932 of them were milk sQTLs and eeQTLs, respectively, an even smaller  
226 number, 720, of them were identified as milk cell geQTLs.

227 *Shared genetic influences between tissues*

228 Within each type of expression QTL between different tissues, the majority of the significant  
229 expression QTL sharing was observed between white blood and milk cells and between liver  
230 and muscle (Figure 5a). This is not unexpected since most of the white blood and milk cells  
231 came from the same lactating cows and the muscle and liver from different growing Angus  
232 bulls. Nevertheless, there was significant sharing of eeQTLs between milk cells and liver and

233 between milk cells and muscle (Figure 5a). The largest amount of across-tissue expression  
234 QTL sharing was observed in eeQTLs, followed by sQTLs and geQTLs (Figure 5). Where a  
235 SNP was significantly associated with variation in expression in two tissues, the direction of  
236 effect was usually the same in both tissues (Additional file 2: Supplementary Figure S5). The  
237 correlation between effects of expression QTLs for white blood cells and milk cells  
238 (Additional file 2: Supplementary Figure S5a,c,e) was stronger than that between liver and  
239 muscle (Additional file 2: Supplementary Figure S5b,d,f). The sharing at the SNP level  
240 between white blood cells and milk cells and between liver and muscle were also evident at  
241 the exon and gene level (Figure 5b).

242 The expression QTL sharing between tissues was further examined for all types of expression  
243 QTL by using a less stringent p-value ( $p < 0.05$ ) to test their effect (Additional file 2:  
244 Supplementary Figure S6). This showed that the expression QTL sharing between tissues was  
245 stronger for sQTLs and eeQTLs (Additional file 10: Supplementary Figure S6a-d), than the  
246 sharing for geQTLs (Additional file 10: Supplementary Figure S6e-f). Again, more  
247 expression QTL sharing was found between white blood cells and milk cells than between  
248 liver and muscle. For instance, 75% of eeQTLs significant in the white blood cells at  $p < 0.002$   
249 were significant in milk cells at  $p < 0.05$  (Additional file 2: Supplementary Figure 6c).

250 The correlation between estimated SNP effects on gene splicing and expression in different  
251 tissues are lower in magnitude than the true correlation between SNP effects, because the  
252 effects are estimated with error and these errors are independent between tissues. To estimate  
253 the true correlation, we computed the genetic correlation between SNP effects in two  
254 different tissues by GREML<sup>16</sup> using a local genomic relationship matrix or LGRM built from  
255 SNPs from 1 Mb surrounding the exon or gene (Figure 5c). LGRM differs from a  
256 conventional GRM by focusing on the local SNPs (in this case within 1 Mb distance) with  
257 potential *cis* genetic associations with the variation in the splicing or expression level of the

258 exon or gene. This was in agreement with the definition of the *cis* expression QTLs which  
259 were also within 1Mb distance to the exon or gene in the current study. Out of 1,145 analysed  
260 sQTLs (inclusion ratio of the exon), eeQTLs (expression level of the exon) and geQTLs  
261 (expression level of the gene) between tissues, 598 had genetic correlations significantly  
262 ( $p < 0.05$ ) different from 0, out of which 561 had genetic correlations insignificantly ( $p \geq$   
263 0.05) different from 1. That is, in many cases, the variation in exon expression in white blood  
264 cells and milk cells was associated with the same *cis* polymorphism (s).

265 Often both splicing events and exon expression within a gene were highly correlated between  
266 white blood and milk cells, for instance *DDX19B*, *CTSD* and *EFF1A1* (Figure 5c). In liver  
267 and muscle, exons from *HLA-DQA1* encoding major histocompatibility complexes [30] also  
268 showed significant genetic correlations between tissues based on both exon expression and  
269 splicing. There were more cases of eeQTLs than sQTLs and geQTLs and so there were more  
270 estimates of genetic correlations between white blood and milk cells in Figure 5c. The  
271 genetic correlations between eeQTLs in white blood and milk cells show a range from +1 to -  
272 1 although most are close to +1. Exons with negative genetic correlations of expression  
273 between white blood cells and milk cells were mapped to *SART1* [30], a post-transcriptional  
274 regulator in epithelial tissues and *TTC4* with potential to mediate protein-protein interactions  
275 [30]. These negative genetic correlations imply that there are mutations that increase the  
276 expression of the exon in milk cells but decrease it in white blood cells. An exon within  
277 *SI00A10*, a cell cycle progress regulator, showed negative genetic correlation of expression  
278 between liver and muscle.

279 Genetic correlations between exon expression levels in two tissues can be different between  
280 exons within the same gene. For example, only the 2<sup>nd</sup> exon (5:93,942,055 -93,942,195) of  
281 *MGST1* (which is associated with the variation in dairy cattle milk fat yield [27, 31]) had a  
282 significant genetic correlation of expression between white blood cells and milk cells (Figure

283 5c). This was largely due to a few eeQTLs with relatively highly significant effects ( $p < 1e-10$ )  
284 on the expression levels of the 2<sup>nd</sup> exon in the milk cells and a similar but less significant  
285 effect ( $p < 1e-4$ ) on the 2<sup>nd</sup> exon expression in white blood cells (Additional file 2:  
286 Supplementary Figure S7). For exons 1 and 3, the significances of the eeQTLs in both milk  
287 and white blood cells were  $> 1e-3$ . For exon 4, the significance of the majority of eeQTLs in  
288 both milk and white blood cells were  $> 1e-5$  (Additional file 2: Supplementary Figure S7).

### 289 *Multi-transcriptome meta-analysis to increase power of expression QTLs detection*

290 Based on shared genetic effects of all types of expression QTLs across tissues, a multi-  
291 transcriptome meta-analysis was introduced to increase the power to detect sQTLs, eeQTLs  
292 and geQTLs (Figure 6, Table 3). For sQTLs, eeQTLs and geQTLs that had significant effects  
293 ( $p < 0.05$ ) in all of white blood cells, milk cells and muscle transcriptomes, their standardised  
294 effects (signed t values) in each transcriptome were simply combined and tested for  
295 significance against a  $\chi^2$  distribution with 1 degree of freedom. Overall, the multi-  
296 transcriptome meta-analysis based on summary statistics substantially increased the power of  
297 expression QTLs detection (Figure 6). The significance of multi-transcriptome expression  
298 QTLs was compared with their significance in the liver transcriptome (Figure 6, Table 3). For  
299 a criteria where the expression QTLs had both multi-transcriptome meta-analysis  $p < 1 \times 10^{-5}$   
300 and liver transcriptome analysis  $p < 0.05$ , all types of expression QTLs had significant  
301 overlap of the SNPs between the meta-analysis and the single transcriptome analysis in liver.  
302 In fact most of the significant sQTLs, eeQTLs and geQTLs detected by the meta-analysis  
303 were also detected by the liver analysis but at a much higher p-value (Table 3).

304

Table 3. The overlap between the multi-transcriptome meta-analysis of three tissues and the single transcriptome results of liver. sQTLs: splicing quantitative trait loci; eeQTLs: exon expression QTLs; geQTLs: gene expression QTLs. Liver: QTLs with single-transcriptome effects  $p < 0.05$  in the liver. Meta-analysis: QTLs with  $\chi^2 p < 1e-05$  for the meta-analysis of white blood cells, milk cells and muscle. '+' indicates the number of SNPs met the significance criteria while '-' indicates number of SNPs failed to meet the significance criteria.

sQTLs		Meta-analysis		Overlap p
		+	-	
Liver	+	768,650	2,842,552	$p < 2e-16$
	-	139,551	4,301,536	

eeQTLs		Meta-analysis		Overlap p
		+	-	
Liver	+	1,052,041	3,186,839	$p < 2e-16$
	-	140,033	3,575,617	

geQTLs		Meta-analysis		Overlap p
		+	-	
Liver	+	178,093	473,232	$p < 2e-16$
	-	43,578	6,663,855	

305

306 *Overlap between expression QTLs and QTL for dairy and beef traits*

307 We examined whether cis sQTLs, eeQTLs and geQTLs were significantly enriched amongst  
308 SNPs associated with economically important cattle traits. Pleiotropic SNPs significantly  
309 (FDR<0.01) associated with more than one of 24 dairy traits [32] and with more than one of  
310 16 beef traits [33] were tested for overlap with detected sQTLs, eeQTLs and geQTLs (Figure  
311 7, Additional file 10,11: Supplementary Table S8,9). Overall, sQTLs, geQTLs and eeQTLs  
312 identified in white blood and milk cells had greater overlap with SNPs associated with dairy  
313 and beef traits than sQTLs, geQTLs and eeQTLs identified in liver and muscle (Figure 7).  
314 sQTLs in white blood and milk cells were significantly enriched for dairy cattle pleiotropic  
315 SNPs, including SNPs from the *CSN3* loci on chromosome 6 (Figure 7a,b). eeQTLs in the



316 white blood cells had the largest absolute amount of SNPs overlapping with dairy cattle  
317 pleiotropic SNPs (Figure 7b) and was the only expression QTL type with significant  
318 enrichment with beef cattle pleiotropic SNPs (Figure 7b). eeQTLs in milk cells and liver also  
319 had significant enrichment for dairy cattle pleiotropic SNPs (Figure 7a). Of the geQTLs, only  
320 those from white blood cells had a significant enrichment with dairy cattle pleiotropic SNPs  
321 (Figure 7a).

322 An example of an eeQTLs that overlaps a milk production QTL is for *MGST1*, where effects  
323 of milk cell eeQTLs were highly significantly associated with their effects on milk fat yield  
324 [32] (Figure 7c). Specifically, some expression QTLs with strong associations with the  
325 variation in milk cell expression levels of exon 2 (Chr5:93,942,195-93,942,055) and exon 3  
326 (Chr5: 93,939,244-93,939,150) of *MGST1* (Additional file 2: Supplementary Figure S7) also  
327 had strong associations with the variation in milk fat yield (Figure 7c). Littlejohn et al.[31]  
328 identified SNPs associated with milk yield percentage and *MGST1* expression in the  
329 mammary gland, including 17 putative causal variants. Most SNPs identified by Littlejohn et  
330 al. originated from whole genome sequence and so were not present on the high density SNP  
331 chip data we analysed for dairy cattle pleiotropy [32] (Additional file 12: Supplementary  
332 Table S10). However, 53 significant milk cell eeQTLs identified by the current study  
333 overlapped with the top 200 SNPs from Littlejohn et al [31] (Additional file 12:  
334 Supplementary Table S10), which was significantly more than expected by random chance.  
335 The 53 eeQTLs included the SNP suggested as a putative causal candidate (Chr5:93945738)  
336 [31], which was significantly associated with the variation in expression level of the third  
337 exon (5:93939150-93939244) of *MGST1* in milk cells (Additional file 12: Supplementary  
338 Table S10). No milk cell geQTLs was called for *MGST1*, as all of them had weak effects on  
339 the whole *MGST1* gene expression in milk cells, resulting in a large FDR (Additional file 12:  
340 Supplementary Table S10).

341

## 342 **Discussion**

343 We performed a systematic analysis of cis expression QTLs ( $\leq 1\text{Mb}$ ) in multiple tissues  
344 centred around RNA splicing events, using a large number of RNA and whole genome  
345 sequence data from an important domestic animal species. Overall, differential splicing  
346 between tissues is ubiquitous and between breeds is common. Differential splicing between  
347 individuals due to SNPs (sQTLs) occurs for many genes and is enriched with cattle complex  
348 trait QTL. Within each tissue, all cis expression QTLs types showed significant overlap. Most  
349 geQTLs and sQTLs were detected as eeQTLs indicating that the exon expression can be  
350 altered by changing the expression of the whole gene or by changing the splicing. However,  
351 an sQTL was likely to be an eeQTLs and, to a lesser extent, geQTLs. Between tissues, while  
352 all QTLs types showed significant overlap between white blood cells and milk cells and  
353 between liver and muscle, the strongest cross-tissue sharing appeared to be at the exon level  
354 (sQTLs and eeQTLs). This is supported by many significant tissue pair genetic correlations.  
355 Such cross-tissue expression QTL sharing allowed the multi-transcriptome meta-analysis of  
356 expression QTL effects which substantially increases power to detect significant expression  
357 QTLs.

358 The majority of significant sQTLs were detected from white blood and milk cells (Figure  
359 2a,b) which also overlapped with SNP chip based complex trait QTL (Figure 7), compared to  
360 sQTLs detected from liver and muscle. This is probably due to the larger sample size for  
361 white blood and milk cells than for liver and muscle (Table 1) and the higher sequencing  
362 depth (Additional file 1: Supplementary Methods). One of the significant white blood cell  
363 sQTLs (29:44585782) for *CAPNI* is also a SNP chip based significant pleiotropic SNP for 16

364 beef cattle traits (Additional file 11: Supplementary Table S9). This SNP is associated with  
365 shear force in multiple taurine breeds [34].

366 In the milk cell transcriptome, a significant sQTL (Chr6:87392580, Figure 2a) with predicted  
367 splicing function [28] within the fifth exon (6:87,392,578-87,392,750) of *CSN3* is strongly  
368 associated with differential splicing between Holstein and Jerseys (Figure 1b). Variants  
369 within *CSN3* have long been found to be associated with milk traits [35, 36] but only recently  
370 have putative causal variants been prioritised [26]. The milk cell sQTL 6:87392580 had  
371 perfect linkage disequilibrium ( $r=1$ ) with the variant 6:87390576 which has been suggested  
372 as a putative causal variant for effects on milk protein yield and percentage [26, 27]. Given it  
373 is at a splicing site, 6:87392580 could be a putative causal variant contributing to milk  
374 production in dairy cattle by altering exon splicing.

375 Compared to identified bovine cis geQTLs, cis sQTLs tended to be closer to the transcription  
376 starting site (TSS) and had highest concentrations of intronic SNPs (Figure 3). In humans, cis  
377 sQTLs [9, 37] were more enriched for intron SNPs than other types of QTLs. However,  
378 reports of the distance between human QTLs and TSS appear to be inconsistent. While no  
379 difference in enrichment of SNPs near TSS between sQTLs and geQTLs were found by the  
380 human GTEx project [8], a more recent study [9] found that human geQTLs were more  
381 enriched near TSS than sQTLs. Our results appear to stand in between the results of GTEx  
382 project and the later findings from Li et. al. [9], where cattle sQTLs were slightly closer to  
383 TSS than geQTLs. However, this difference is not significant in all tissues (Figure 3a). On  
384 the other hand, significant overlap between sQTLs and geQTLs was found in this study  
385 (Figure 4) and by the human GTEx project [8]. However, Li et. al. [9] found that human cis  
386 sQTLs were independent of geQTLs. These inconsistent observations are likely to be due to a  
387 number of differences between these studies, including definition of sQTLs, choice of tissues

388 and populations and computational procedures. Also, these inconsistent observations also  
389 suggest that we are still at the very early stage of understanding of sQTLs features.

390 Within each studied bovine tissue, the largest amount of overlap between expression QTL  
391 types was found either between exon expression eeQTLs and sQTLs or between eeQTLs and  
392 geQTLs (Figure 4). Further, the largest amount of enrichments of cattle pleiotropic SNPs was  
393 found for eeQTLs, followed by sQTLs and geQTLs. The white blood cell eeQTLs showed  
394 particularly strong enrichments of pleiotropic SNPs for dairy and beef cattle. In a large scale  
395 human blood cell expression QTLs study [12], eeQTLs also showed the strongest  
396 enrichments of GWAS variants, followed by sQTLs and geQTLs. Thus, focusing on exon-  
397 level QTLs, including eeQTLs and sQTLs, could increase the chance of finding regulatory  
398 variants for complex traits, as proposed by Guan et. al.[38].

399 A hypothesis to explain these results is that mutations in regulatory DNA may increase the  
400 expression of one or more transcripts from a gene. If they increase expression of one  
401 transcript then they may be detected as an eeQTL for the exons in that transcript, as a sQTL  
402 for exons spliced out of that transcript or as a geQTL if this transcript forms a large part of  
403 the total transcription from the gene. Thus, there is expected to be overlap between eeQTLs,  
404 geQTLs and sQTLs, but at least sQTLs and eeQTLs should overlap and this is what we found  
405 (Figure 4, Additional file 9: Supplementary Table S7). It appears that eeQTLs detect the  
406 largest proportion of these regulatory polymorphisms provided sequencing depth is high.

407 In humans, significant cross-tissue sharing of sQTLs and geQTLs was reported [8, 39]. In our  
408 study of cattle, the strongest evidence of expression QTL sharing appeared to be at the exon  
409 level. This included sQTLs and eeQTLs sharing between white blood and milk cells and  
410 between liver and muscle (Figure 5). When extending the examination of expression QTLs to

411 include those with  $p < 0.05$  (Additional file 2: Supplementary Figure S6), the exon-level  
412 expression QTLs cross-tissue sharing is also the greatest.

413 We highlighted a few examples of cross-tissue shared eeQTLs along with the related exons,  
414 of which the genetic correlations of the expression and splicing in different tissues were  
415 significant (Figure 5c). One of these eeQTLs is located within the milk fat yield [27, 31] QTL  
416 *MGST1* (Figure 5c, Additional file 2: Supplementary Figure S7). For eeQTLs associated with  
417 *MGST1*, a strong positive relationship of SNP effects was observed between milk cell  
418 eeQTLs and dairy milk fat yield SNPs (Figure 7c). Furthermore, the identified milk cell  
419 eeQTL overlaps with previously identified putative causal variants [31] within *MGST1* for  
420 milk fat percentage, thus supporting their candidacy. This overlap further supports the top  
421 candidate SNP 5:93945738 with significant effects on the abundance of the third exon of  
422 *MGST1* (Additional file 12: Supplementary Table S10) for milk fat traits. Overall, our  
423 analysis demonstrates the significant potential of using detailed exon analysis to aid in  
424 identification of putative causative mutations.

425 Based on the sharing of expression QTLs between tissues, a multi-transcriptome meta-  
426 analysis which simply combined expression QTL effects to substantially increase the power  
427 (Figure 6) was introduced. Using this approach, combined expression QTL effects of white  
428 blood cells, milk cells and muscle were validated in the liver (Table 3). This also  
429 demonstrated the significant extent of QTL sharing across tissues. Previously, Flutre et. al.  
430 [39] combined data from human fibroblasts, lymphoblastoid cell lines and T-cells and found  
431 that up to 88% of geQTLs were shared across tissues at  $FDR < 0.05$  level. We checked the  
432 existing results of the meta-analysis combining SNP effects from tissues of white blood cells,  
433 milk cells and muscle at the FDR threshold  $< 0.05$ . We found that the meta-analysis  
434 identified 585,406 geQTLs with  $FDR < 0.05$  in more than one tissue. This accounted for 69.2%  
435 of total geQTLs (845,431) that were called and common in the individual geQTL analysis of

436 white blood cells, milk cells and muscle. While there were differences in the selection of  
437 tissue/cell type between our experiment and Flutre et. al, it is possible that the analysis  
438 proposed by Flutre et. al with more complex procedures would be more powerful than the  
439 meta-analysis introduced by us. Flutre et. al applied principal components analysis to  
440 normalise their gene expression data while we used quantile normalisation which appeared to  
441 show good performances in combining different transcriptome datasets [40]. However, our  
442 meta-analysis is powerful for detecting and validating many expression QTLs that have an  
443 effect in the same direction in multiple tissues, and is simpler to implement than that of Flutre  
444 et al. A future systematic comparison of different approaches of analysing expression QTL in  
445 multiple tissues would be very useful.

446 As one of earliest investigations of large animal expression QTLs, our study has its potential  
447 limitations. While the overlaps between sQTLs detected with exon and intron analyses were  
448 significantly more than expected by random chance, the absolute amount of overlap was still  
449 small. Through all analyses, there were always many more splicing events detected by intron  
450 analyses implemented by leafcutter [10] than the exon analysis (Additional file 5:  
451 Supplementary Table S3). This appears to be consistent with Li et al [10], the authors of  
452 leafcutter. They suggested that intron-centred analyses can be both more sensitive (lower  
453 proportion of false negatives) and more accurate (lower proportion of false positives) than the  
454 exon based splicing mapping methods, such as Altrans [41].

455 We found that the strongest sharing of expression QTLs was either between white blood and  
456 milk cells or between liver and muscle tissues, at the threshold of  $FDR < 0.01$  (Figure 5). The  
457 white blood and milk cells sampled from the same Holstein and/or Jersey cattle of experiment  
458 III had a larger sample size and higher read coverage, compared to the liver and muscle  
459 tissues sampled from different Angus bulls of experiment IV. The reduced expression QTL  
460 sharing detected between, e.g., muscle and milk cells, could be due to differences in the tissue,

461 the physiological state of the cattle or the breed. However, it can be also due to different  
462 power in the milk cells, liver and muscle datasets compared to the white blood cell data.  
463 Nevertheless, in the multi-transcriptome meta-analysis where expression QTLs with low  
464 threshold were examined ( $p < 0.05$ ), the combined effects of all types of expression QTLs of  
465 the three tissues from different experiments were highly significant (Figure 6). Many of these  
466 expression QTLs were also found in liver with  $p < 0.05$  (Table 3). This evidence supports the  
467 proposal that the sharing of cis expression QTL is extensive across tissues, but these shared  
468 expression QTLs may not necessarily have strong effects in each studied tissue. In the latest  
469 human expression QTLs mapping study (GTEx consortium) where RNA seq data of 44  
470 tissues from up to 450 individuals were analysed, cis expression QTL tended to be either  
471 shared across most tissues or specific to a small subset of tissues [11]. As sample numbers for  
472 each tissue increased, GTEx consortium identified more tissue specific expression QTLs [11].  
473 Future studies with significantly increased power and selection of cattle tissues and breeds  
474 may update our current results.

475 Another potential limitation of our study is the use of imputed sequence data, which may  
476 introduce imputation errors that lead to inaccurate identification or exclusion of expression  
477 QTLs. However, the average imputation accuracy of the 1000 bull genome project data used  
478 in this study was high (0.926) [21] and there was a good consistency between the imputed  
479 sequence genotypes and RNA sequence genotypes (average concordance = 0.943, Additional  
480 File 2: Supplementary Figure S3). Stringent thresholds were also imposed to control the false  
481 discovery rate of expression QTLs mapping (either  $FDR < 0.01$  or  $FDR$  approximately  $< 0.01$   
482 for sQTLs). In the current study, we did not consider the case where a haplotype can be  
483 potentially associated with expression phenotypes. While a haplotype analysis can be  
484 informative, it would require a very large sample size to achieve reliable results due to testing  
485 a large number of combinations of haplotype blocks. In a human study where over 2,000

486 individuals were analysed, expression QTLs conditioning on expression levels of  
487 transcription factor genes were reported [12]. Finally, our results obtained from genome-wide  
488 associations do not necessarily contain causal relationships. However, our findings are  
489 important for prioritising informative SNP candidates for future validation of causal  
490 relationships.

491

## 492 **Conclusions**

493 We found that eeQTLs overlapped with both geQTLs, due to polymorphisms affecting the  
494 level of expression of the whole gene, and with sQTLs, due to polymorphisms affecting the  
495 exon usage within the gene. sQTLs tended to be closer to the transcription start sites more  
496 often located in introns than geQTLs. We found the largest number of sQTLs in white blood  
497 cells probably because the power to find them was greatest in this dataset. However, many of  
498 the sQTLs found in other tissues were also detected in blood cells and many sQTLs found in  
499 blood could be detected in other tissues at higher p-values. The genetic correlation between  
500 expression QTLs in different tissues was often indistinguishable from 1.0 indicating that  
501 many expression QTLs operate in a similar way across tissues. Consequently, combining  
502 results from several tissues using the multi-transcriptome meta-analysis increased power to  
503 detect all 3 types of expression QTLs. The potential of exon-level QTLs information was  
504 demonstrated by the identification of several strong candidates of putative causal mutations  
505 for complex traits: sQTL 6:87392580 within *CSN3* for milk production and eeQTL  
506 5:93945738 within *MGST1* for milk fat yield.

507

## 508 **Methods**



509 **Sample collection.** For Experiment I, the sampling of 18 tissues from one lactating Holstein  
510 cow followed procedures described by Chamberlain et al [24]. For Experiment II and III, the  
511 sampling and processing of all tissues including white blood and milk cells is detailed in  
512 Additional file 1: Supplementary Methods. Briefly, animals of Experiment II and III were  
513 selected from Agriculture Victoria Research dairy herd at Ellinbank, Victoria, Australia. In  
514 Experiment II, milk and mammary tissue samples were taken from six Holstein cows. In the  
515 Experiment III, milk and blood samples were originally taken from 112 Holstein and 29  
516 Jersey cows, but only RNA sequence data of 105 Holstein and 26 Jersey with > 50 million  
517 reads for milk cells or >25 million reads for white blood cells and had a concordant alignment  
518 rate [18] >80% were used. For Experiment IV, the sampling of 41 *semitendinosus* muscle and  
519 35 liver from Angus bulls was previously described by [42, 43]. As recommended by  
520 ENCODE guidelines (<https://www.encodeproject.org/about/experiment-guidelines/>)  
521 biological replicates were favoured over technical replicates for experiments II-IV. However  
522 Chamberlain et al [24] assessed technical replicates for experiment I.

523 **RNA seq data.** For Experiment I, RNA extraction and sequencing followed the procedures  
524 described by Chamberlain et al [24]. For Experiment II and III, the RNA extraction and  
525 sequencing procedure is detailed in Additional file 1: Supplementary Methods. For  
526 Experiment IV, RNA extraction and sequencing is previously described by Khansefid et al.  
527 [43]. For all experiments sequence quality was checked and were aligned to the Ensembl  
528 UMD3.1 bovine genome assembly using TopHat2 [18]. The RNA sequence data processing  
529 and quality checking are detailed in Additional file 1: Supplementary Methods.

530 **Whole genome seq data.** Experiments III and IV had whole genome sequence genotypes  
531 imputed from the SNP chip genotypes using FImpute [44] based on the 1000 bull genomes  
532 project [19]. The overall imputation accuracy of the most recent genome sequence data  
533 ranged from 0.898 to 0.952 depending on chromosomes [21]. 50K Illumina genotypes were

534 used for imputation for experiment III with previous protocols [27, 45]. For experiment IV,  
535 800K and 50K Illumina genotypes were used with procedures following [33]. SNPs were  
536 filtered for minor allele frequency  $> 0.01$  and resulted in 14,302,604 and 13,632,145 SNPs  
537 used in the analysis in experiment III and IV, respectively. There were 10,242,837 SNPs  
538 shared between experiment III and IV.

539 **Gene/exon analysis.** Gene count data were generated by Python package HTSeq [46] using  
540 default settings. The exon count data were generated by Bioconductor package featureCounts  
541 [47] in R v3.3.2 [48]. The Ensembl based bovine genome reference (UMD3.1) was used to  
542 define genes and exons. Genes and exons with count per million  $> 0$  in more than 40% of  
543 RNA samples were used for all the following analyses. This filtering allowed the analysis to  
544 focus on exons or genes with relatively robust expression in many RNA sequencing samples.  
545 The exon-based tissue principal components analysis used DEseq2 based on the 250 exons,  
546 the expression of which were most variable across studied tissues [49]. The usage of 500 and  
547 1,000 exons with the most variable expression across tissue samples were also tested.  
548 Consistent with [49], the selection of different numbers of exons had little impact on the  
549 clustering patterns (Supplementary Figure S4). The significance of the clustering was  
550 determined using ellipse method proposed by [22] and implemented in ggplot2 [50]. The  
551 confidence interval was set to 0.95 to which ellipses were drawn based on data categorised by  
552 tissue types or by experiments. The separation of ellipses indicated independence of  
553 categories of data. The phenotype of exon inclusion was calculated as the exon to gene  
554 expression ratio. The phenotype of intron excision was estimated using the publically  
555 available software leafcutter [9, 10]. Briefly, leafcutter used RNA seq BAM files as input and  
556 generated ratios of reads supporting each alternatively excised intron as the intron excision  
557 phenotype [10](<http://davidaknowles.github.io/leafcutter/>). Those exons and introns with  
558 ratio values  $< 0.001$  were removed and the remaining ratio values were transformed to  $\log_2$

559 scale, then underwent exon/intron –wise quantile normalisation and individual –wise zscore  
560 standardisation [51].

561 **Gene differential splicing.** Both exon inclusions and intron excisions were analysed and  
562 used in combination for gene differential splicing for (1) the overall tissue effects and (2) the  
563 breed effects. Primarily, differential splicing was defined for the gene containing exons  
564 whose variation in inclusion ratios were significantly (FDR<0.1) associated with the tissue or  
565 breed variable. To be called as significantly spliced exons, they were required to have at least  
566 one adjacent intron whose variation in excision ratios were also significantly (FDR<0.1)  
567 associated with the tissue or breed variable. The tissue effects were analysed in a linear mixed  
568 model in lme4 [52] in R as:

$$569 \quad y_{ijk} = b_i + x_j + t_k + e \quad (1)$$

570 Where  $y$  = exon inclusion or intron excision ratios,  $b_i$  = the animal random effects ( $i=214$ ),  
571  $x_j$  = the experiments ( $j=4$ ),  $t_k$  = tissue type ( $k=19$ ),  $e$  = random residual term. The fitting of the  
572 animal random effects accounted for the fact that only 1 animal was used in experiment I.  
573 The P values of F tests were calculated using Satterthwaite approximation implemented in  
574 lmerTest [53]. The breed effects for the milk transcriptome data were analysed in a linear  
575 model in R as:

$$576 \quad y_l = breed_l + e \quad (2)$$

577 Where  $y$  = exon inclusion or intron excision ratios in the milk cell transcriptome,  $breed_l$  =  
578 breeds ( $l=2$ , Holstein and Jersey). p values of exons/introns for the tissue effects in equation  
579 (1) or the breed effects in equation (2) were used to calculate the false discovery rates (FDR)  
580 using qvalue [54] in R. The FDR threshold of such detected exon/intron group was  
581 considered as approximately  $0.1 \times 0.1 = 0.01$ , as a combination of FDR thresholds of exon

582 and intron analyses to reflect our selection criteria for significant splicing events. For genes  
583 showing significant differential splicing for (1) the overall tissue effects and (2) the breed  
584 effects as described above, enrichments of biological pathway were tested using GOrilla [55].  
585 As many genes had differential splicing events associated with tissue differences, top 10% of  
586 the genes with significant differential splicing were selected based on the approximate FDR  
587 with combined FDR values of both exon and intron analyses.

588 **cis expression splicing QTLs.** Only transcriptomic data of experiment III and IV were used  
589 in sQTLs mapping. Similar to differential splicing analysis described above, a significant  
590 (FDR<0.01) cis splicing QTLs was expected to satisfy two conditions simultaneously: (1) a  
591 SNP, within or up to  $\pm 1$ Mb away from the exon, was significantly (FDR<0.1) associated  
592 with the variation in the exon inclusion ratio and (2) the same SNP was significantly  
593 (FDR<0.1) associated with at least one event of the excision of the intron next to the same  
594 exon at the same significance level. Both individual exon inclusion and intron excision values  
595 were used as phenotype to map associated QTLs with widely used [8] Matrix eQTL[56]  
596 package in R. For each cell type of the experiment III (white blood and milk cells) and  
597 experiment IV (liver and muscle), SNPs  $\pm 1$ Mb from the exon or intron were tested for  
598 regressions with the exon inclusion or intron excision phenotype. For milk cell transcriptome,  
599 breed was fitted as a covariate.

600 To compare cis sQTLs with exon expression cis eeQTLs and cis gene expression geQTLs,  
601 the expression count data were normalised by voom [57] estimating mean-variance  
602 relationship to calculate observation-level weighted expression values. Normalised  
603 expression values of exons and genes were used as phenotype to map cis expression QTLs  
604 (within  $\pm 1$ Mb) at FDR <0.01 level as described above.

605 **SNP annotation.** The gene transcription start site coordinates were downloaded from  
606 Ensembl (<http://www.ensembl.org>) and the absolute difference between the position of a SNP  
607 and the transcription start site of the gene were calculated for the SNP with significant cis  
608 effects. The SNP functional categories were generated using predictions from Ensembl  
609 Variant Effect Predictor [28] in conjunction with NGS-SNP [29]. All analysed SNPs were  
610 assigned a functional category.

611 **Dairy and beef cattle pleiotropic QTL.** To test the significance of overlap between cis  
612 expression QTLs and SNPs associated with cattle phenotype, meta-analyses of dairy and beef  
613 cattle pleiotropy were performed using single-trait GWAS results from Xiang et al [32] and  
614 Bolormaa et al [33]. HD 800K SNP chip genotypes were used for trait GWAS. 24 dairy cattle  
615 traits with matching phenotype in 9,662 bulls and cows and 16 beef cattle traits with animal  
616 numbers >2,000 were selected. Briefly, the multi-trait  $\chi^2$  statistic for the  $i$ th SNP was  
617 calculated based on its signed t values generated from each single trait GWAS [33]:

$$618 \chi^2 = t'_i V^{-1} t_i \quad (3)$$

619 For dairy cattle, the meta-analysis was based on the weighted SNP effects  $t_w$  combining SNP  
620 effects calculated separately in bulls and cows. The  $t_w$  accounting for the phenotypic error  
621 differences between bulls and cows [27] was calculated as:

$$622 t_w = \frac{B_w}{se_w} = \frac{\frac{\frac{B_{bull}}{se_{bull}^2} + \frac{B_{cow}}{se_{cow}^2}}{\frac{1}{se_{bull}^2} + \frac{1}{se_{cow}^2}}}{\sqrt{\frac{1}{\frac{1}{se_{bull}^2} + \frac{1}{se_{cow}^2}}}} \quad (4)$$

623 Where the weighted SNP t value  $t_w$  was the quotient of the weighted SNP effects  $B_w$  and the  
624 weighted effect error  $se_w$ .  $B_{bull}$  and  $se_{bull}$  were the SNP effects and error obtained from single-

625 trait GWAS in bulls and  $B_{cow}$  and  $se_{cow}$  were the SNP effects and error of cows. Those SNPs  
626 which had meta-analysis  $FDR < 0.01$  were chosen to be compared with cis expression QTLs.  
627 The lead SNP loci were defined as  $\pm 1$  Mb from the lead SNPs identified in the previous  
628 analysis [32] and [33].

629 **The significance of overlaps.** The significance of overlaps were compared with the expected  
630 number using the Fisher's exact test ( $p$ ) implemented in GeneOverlap [58] in R. This analysis  
631 required four types of counts: the size of overlap between set A (e.g., SNPs that were blood  
632 sQTLs) and set B (e.g., SNPs that were milk sQTLs), the size of set A, the size of set B and  
633 the size of background. The union number of whole genome sequence SNPs with  $MAF > 0.01$   
634 in each breed and the bovine high density chip SNPs were used as the background. Where  
635 expression QTL categories from different breeds of dairy and beef cattle were tested for  
636 overlap, the number of common SNPs between breeds was used.

637 **Genetic correlations using local genomic relationship matrices.** The cross-tissue sharing  
638 of SNPs were confirmed by bivariate GREML analysis using GCTA [59]. For an exon or a  
639 gene of interest, its inclusion ratios or expression levels in two different tissues were treated  
640 as two different phenotype,  $tr_1$  and  $tr_2$ . The SNPs within 1Mb of this exon or gene were used  
641 to make a local genomic relationship matrix, i.e., LGRM, representing the local polygenic  
642 component  $a$  with potential associations with the variation in the splicing or expression level  
643 of the exon or gene. This allowed linear mixed modelling of the local additive genetic  
644 variances of  $tr_1$ ,  $var_{lg}(tr_1)$  and of  $tr_2$ ,  $var_{lg}(tr_2)$  and the local additive genetic covariance  
645 between  $t_1$  and  $t_2$ ,  $cov_{lg}(tr_1, tr_2)$  using GREML [59]. This approach agreed with the definition  
646 of cis expression QTLs defined in this study (also within 1Mb distance to the exon or gene)  
647 and allowed the estimation of genetic correlation  $r_{lg} = \frac{cov_{lg}(tr_1, tr_2)}{\sqrt{var_{lg}(tr_1)var_{lg}(tr_2)}}$  (5). Genetic

648 correlations were also tested for their significance of being different from 0 or 1, by fixing the  
649 correlation value to 0 and 1 using GCTA [59].

650 **Validation by multi-transcriptome meta-analysis.** The validation based on expression QTL  
651 effect commonality across tissues was conducted by comparing the combined expression  
652 QTL effects from white blood cells (experiment III), milk cells (experiment III) and muscle  
653 (experiment IV) transcriptomes with their effects in the liver transcriptome (experiment IV).  
654 The standardised expression QTL effects, b/se, signed t values were calculated from single-  
655 transcriptome results of white blood cells ( $t_1$ ), milk cells ( $t_2$ ) and muscle ( $t_3$ ). The significance  
656 of multi-transcriptome effects of an expression QTL was tested by  $\chi^2$  distribution with 1  
657 degree of freedom:

$$658 \quad \chi_{(1)}^2 = \left[ \sum_{n=1}^N \frac{t_n}{\sqrt{N}} \right]^2 \quad (6)$$

659  $N$  = the number of studied tissues ( $N=3$  in this case) where the original SNP t values were  
660 estimated. Provided the individual t-values followed a t-distribution under the null hypothesis,  
661 the properties of the average t value in the current study was a simple mathematical result  
662 which approximated the chi square distribution with 1 degree of freedom, the null hypothesis  
663 of which was that the SNP does not have any significant associations in any of the 3 tissue  
664 types. Previously, the concept of meta-analysis combining SNP t values estimated from  
665 different datasets has been also applied to analyse multiple quantitative phenotypic traits in  
666 large animals to increase power (see [33, 60] and equation (3)). The expression QTLs that  
667 participated in the validation analysis had single-transcriptome effect  $p < 0.05$  in each tissue  
668 and the significance of the multi-transcriptome effects was defined as  $p < 1e-05$ . Significant  
669 multi-transcriptome expression QTLs were compared with the liver single-transcriptome  
670 effects at  $p < 0.05$  level. We chose to combine two tissues which appeared to display strong

671 power (white blood and milk cells, experiment III) with the third tissue from a different  
672 experiment with relatively weak power (muscle, experiment IV). The single tissue left to be  
673 compared with was liver, a tissue which also appeared to show weak power and was from  
674 experiment IV. These choices intended to create enough differences for the meta-analysis to  
675 combine the SNP effects and for the combined SNP effects to be compared with the SNP  
676 effects in the single tissue.

677

### 678 **List of abbreviations**

679 sQTLs: splicing quantitative trait loci;  
680 geQTLs: gene expression quantitative trait loci;  
681 eeQTLs: exon expression quantitative trait loci.

682

### 683 **Declarations**

#### 684 **Ethics approval and consent to participate**

685 For experiment I-III, the animal ethics was approved by the Victoria Animal Ethics  
686 Committee (application number 2013-14), Australia. For experiments IV, the animal ethics  
687 was approved by the University of New England Animal Ethics Committee (AEC 06/123,  
688 NSW, Australia) and Orange Agriculture Institute Ethics Committee (ORA09/015, NSW,  
689 Australia).

690

#### 691 **Consent for publication**

692 Not applicable.



693

694 **Availability of data and materials**

695 The RNA sequence data for experiment I was published [24] (NCBI Sequence Read Archive,  
696 SRA, accession SRP042639); For experiment II: SRA accession SRP111067; For experiment  
697 III: SRA accession PRJNA305942; For experiment IV: SRA accession PRJNA392196; The  
698 imputed whole genome sequence data was part of the 1000 bull genome project [19].

699

700 **Competing interests**

701 The authors declare that they have no competing interests.

702

703 **Funding**

704 Australian Research Council's Discovery Projects (DP160101056) supported R.X. and  
705 M.E.G. Dairy Futures CRC and DairyBio (a joint venture project between Agriculture  
706 Victoria and Dairy Australia) supported the generation of the transcriptome data of  
707 experiment I-III. No funding bodies participated in the design of the study and collection,  
708 analysis, and interpretation of data and in writing the manuscript.

709

710 **Author contributions**

711 A.J.C., R.X., M.E.G., B.J.H and H.D.D. conceived the experiments. C.P.P., C.M.R., B.A.M.,  
712 J.B.G., L.C.M., Y.C. and A.J.C. performed sample collections and RNA sequencing  
713 experiments. S.B., I.M.M., M.K. and H.D.D. provided data and assisted with study design.

714 R.X., B.J.H., C.J.V., A.J.C., M.E.G, P.J.B and Z.Y. analysed data. R.X. and M.E.G. wrote the  
715 paper. R.X., M.E.G., B.J.H, A.J.C., I.M.M and H.D.D.. revised the paper. All authors read  
716 and approved the final manuscript.

717

## 718 **Acknowledgements**

719 We thank DataGene for access to data used in this study and Gert Nieuwhof, Kon  
720 Konstantinov and Timothy P. Hancock for preparation and provision of data.

721

## 722 **Reference**

- 723 1. Bourneuf E, Otz P, Pausch H, Jagannathan V, Michot P, Grohs C, Piton G, Ammermüller S,  
724 Deloche M-C, Fritz S: **Rapid Discovery of De Novo Deleterious Mutations in Cattle**  
725 **Enhances the Value of Livestock as Model Species.** *Sci Rep* 2017, **7**.  
726 2. Hayes BJ, Lewin HA, Goddard ME: **The future of livestock breeding: genomic selection for**  
727 **efficiency, reduced emissions intensity, and adaptation.** *Trends Genet* 2013, **29**:206-214.  
728 3. Meuwissen T, Hayes B, Goddard M: **Genomic selection: A paradigm shift in animal**  
729 **breeding.** *Animal frontiers* 2016, **6**:6-14.  
730 4. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW: **Coordinated**  
731 **international action to accelerate genome-to-phenome with FAANG, the Functional**  
732 **Annotation of Animal Genomes project.** *Genome Biol* 2015, **16**.  
733 5. Albert FW, Kruglyak L: **The role of regulatory variation in complex traits and disease.**  
734 *Nature reviews Genetics* 2015, **16**:197.  
735 6. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M: **Linking disease associations**  
736 **with regulatory information in the human genome.** *Genome Res* 2012, **22**:1748-1759.  
737 7. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS,  
738 Sahana G, Govignon-Gion A, Boitard S, Dolezal M, et al: **Meta-analysis of genome-wide**  
739 **association studies for cattle stature identifies common genes that regulate body size in**  
740 **mammals.** *Nat Genet* 2018, **50**:362-367.  
741 8. Consortium G: **The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene**  
742 **regulation in humans.** *Science* 2015, **348**:648-660.  
743 9. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK: **RNA**  
744 **splicing is a primary link between genetic variation and disease.** *Science* 2016, **352**:600-604.  
745 10. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK:  
746 **Annotation-free quantification of RNA splicing using LeafCutter.** *Nat Genet* 2018, **50**:151.  
747 11. Consortium G: **Genetic effects on gene expression across human tissues.** *Nature* 2017,  
748 **550**:204.  
749 12. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, van't Hof P,  
750 Mei H, van Dijk F, Westra H-J: **Identification of context-dependent expression quantitative**  
751 **trait loci in whole blood.** *Nat Genet* 2017, **49**:139-145.

- 752 13. Mazzoni G, Kadarmideen HN: **Computational Methods for Quality Check, Preprocessing**  
753 **and Normalization of RNA-Seq Data for Systems Biology and Analysis**. In *Systems Biology*  
754 *in Animal Production and Health, Vol 2*. Springer; 2016: 61-77
- 755 14. Okonechnikov K, Conesa A, García-Alcalde F: **Qualimap 2: advanced multi-sample quality**  
756 **control for high-throughput sequencing data**. *Bioinformatics* 2015, **32**:292-294.
- 757 15. Wang L, Wang S, Li W: **RSeQC: quality control of RNA-seq experiments**. *Bioinformatics*  
758 2012, **28**:2184-2185.
- 759 16. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory**  
760 **requirements**. *Nat Methods* 2015, **12**:357.
- 761 17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,  
762 Gingeras TR: **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics* 2013, **29**:15-21.
- 763 18. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate**  
764 **alignment of transcriptomes in the presence of insertions, deletions and gene fusions**.  
765 *Genome Biol* 2013, **14**:R36.
- 766 19. Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, Liao X,  
767 Djari A, Rodriguez SC, Grohs C: **Whole-genome sequencing of 234 bulls facilitates mapping**  
768 **of monogenic and complex traits in cattle**. *Nat Genet* 2014, **46**:858-865.
- 769 20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R:  
770 **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**:2078-2079.
- 771 21. Pausch H, MacLeod IM, Fries R, Emmerling R, Bowman PJ, Daetwyler HD, Goddard ME:  
772 **Evaluation of the accuracy of imputed sequence variant genotypes and their utility for**  
773 **causal variant detection in cattle**. *Genet Sel Evol* 2017, **49**:24.
- 774 22. Fox J, Weisberg S: *An R companion to applied regression*. Sage Publications; 2011.
- 775 23. Cánovas A, Rincón G, Bevilacqua C, Islas-Trejo A, Brenaut P, Hovey RC, Boutinaud M,  
776 Morgenthaler C, VanKlumpenberg MK, Martin P: **Comparison of five different RNA sources**  
777 **to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing**.  
778 *Sci Rep* 2014, **4**.
- 779 24. Chamberlain AJ, Vander Jagt CJ, Hayes BJ, Khansefid M, Marett LC, Millen CA, Nguyen TTT,  
780 Goddard ME: **Extensive variation between tissues in allele specific expression in an**  
781 **outbred mammal**. *BMC Genomics* 2015, **16**:993.
- 782 25. Li YI, Knowles DA, Pritchard JK: **LeafCutter: Annotation-free quantification of RNA**  
783 **splicing**. *bioRxiv* 2016:044107.
- 784 26. MacLeod I, Bowman P, Vander Jagt C, Haile-Mariam M, Kemper K, Chamberlain A,  
785 Schrooten C, Hayes B, Goddard M: **Exploiting biological priors and sequence variants**  
786 **enhances QTL discovery and genomic prediction of complex traits**. *BMC Genomics* 2016,  
787 **17**:1.
- 788 27. Kemper KE, Reich CM, Bowman PJ, vander Jagt CJ, Chamberlain AJ, Mason BA, Hayes BJ,  
789 Goddard ME: **Improved precision of QTL mapping using a nonlinear Bayesian method in**  
790 **a multi-breed population leads to greater accuracy of across-breed genomic predictions**.  
791 *Genet Sel Evol* 2015, **47**:1.
- 792 28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F: **The**  
793 **Ensembl Variant Effect Predictor**. *Genome Biol* 2016, **17**:1.
- 794 29. Grant JR, Arantes AS, Liao X, Stothard P: **In-depth annotation of SNPs arising from**  
795 **resequencing projects using NGS-SNP**. *Bioinformatics* 2011, **27**:2300-2301.
- 796 30. **National Center for Biotechnology Information (NCBI)** [<https://www.ncbi.nlm.nih.gov/>]
- 797 31. Littlejohn MD, Tiplady K, Fink TA, Lehnert K, Lopdell T, Johnson T, Couldrey C, Keehan M,  
798 Sherlock RG, Harland C: **Sequence-based Association Analysis Reveals an MGST1 eQTL**  
799 **with Pleiotropic Effects on Bovine Milk Composition**. *Sci Rep* 2016, **6**.
- 800 32. Xiang R, MacLeod I, Bolormaa S, Goddard M: **Genome-wide comparative analyses of**  
801 **correlated and uncorrelated phenotypes identify major pleiotropic variants in dairy cattle**.  
802 *Sci Rep* 2017, **7**:9248.
- 803 33. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes  
804 BJ, Goddard ME: **A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for**  
805 **stature, fatness and reproduction in beef cattle**. *PLoS Genet* 2014, **10**:e1004198.

- 806 34. McClure M, Ramey H, Rolf M, McKay S, Decker J, Chapple R, Kim J, Taxis T, Weaber R,  
807 Schnabel R: **Genome-wide association analysis for quantitative trait loci influencing**  
808 **Warner–Bratzler shear force in five taurine cattle breeds.** *Anim Genet* 2012, **43**:662-673.
- 809 35. Kühn C, Freyer G, Weikard R, Goldammer T, Schwerin M: **Detection of QTL for milk**  
810 **production traits in cattle by application of a specifically developed marker map of BTA6.**  
811 *Anim Genet* 1999, **30**:333-339.
- 812 36. Glantz M, Gustavsson F, Bertelsen HP, Stålhammar H, Lindmark-Månsson H, Paulsson M,  
813 Bendixen C, Gregersen VR: **Bovine chromosomal regions affecting rheological traits in**  
814 **acid-induced skim milk gels.** *J Dairy Sci* 2015, **98**:1273-1285.
- 815 37. Takata A, Matsumoto N, Kato T: **Genome-wide identification of splicing QTLs in the**  
816 **human brain and their enrichment among schizophrenia-associated loci.** *Nature*  
817 *Communications* 2017, **8**:14519.
- 818 38. Guan L, Yang Q, Gu M, Chen L, Zhang X: **Exon expression QTL (eeQTL) analysis**  
819 **highlights distant genomic variations associated with splicing regulation.** *Quantitative*  
820 *Biology* 2014, **2**:71-79.
- 821 39. Flutre T, Wen X, Pritchard J, Stephens M: **A statistical framework for joint eQTL analysis**  
822 **in multiple tissues.** *PLoS Genet* 2013, **9**:e1003486.
- 823 40. Thompson JA, Tan J, Greene CS: **Cross-platform normalization of microarray and RNA-**  
824 **seq data for machine learning applications.** *PeerJ* 2016, **4**:e1621.
- 825 41. Ongen H, Dermitzakis ET: **Alternative Splicing QTLs in European and African**  
826 **Populations.** *Am J Hum Genet* 2015, **97**:567-575.
- 827 42. Chen Y, Gondro C, Quinn K, Herd R, Parnell P, Vanselow B: **Global gene expression**  
828 **profiling reveals genes expressed differentially in cattle with high and low residual feed**  
829 **intake.** *Anim Genet* 2011, **42**:475-490.
- 830 43. Khansefid M, Millen CA, Chen Y, Pryce JE, Chamberlain AJ, Vander Jagt CJ, Gondro C,  
831 Goddard ME: **Gene expression analysis of blood, liver, and muscle in cattle divergently**  
832 **selected for high and low residual feed intake.** *J Anim Sci* 2017, **95**:4764-4775.
- 833 44. Sargolzaei M, Chesnais JP, Schenkel FS: **A new approach for efficient genotype imputation**  
834 **using information from relatives.** *BMC Genomics* 2014, **15**.
- 835 45. Erbe M, Hayes B, Matukumalli L, Goswami S, Bowman P, Reich C, Mason B, Goddard M:  
836 **Improving accuracy of genomic predictions within and between dairy cattle breeds with**  
837 **imputed high-density single nucleotide polymorphism panels.** *J Dairy Sci* 2012, **95**:4114-  
838 4129.
- 839 46. Anders S, Pyl PT, Huber W: **HTSeq—a Python framework to work with high-throughput**  
840 **sequencing data.** *Bioinformatics* 2015, **31**:166-169.
- 841 47. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for**  
842 **assigning sequence reads to genomic features.** *Bioinformatics* 2013, **30**:923-930.
- 843 48. Team RC: **R: A language and environment for statistical computing.** 2013.
- 844 49. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for**  
845 **RNA-seq data with DESeq2.** *Genome Biol* 2014, **15**:550.
- 846 50. Wickham H: *ggplot2: elegant graphics for data analysis.* Springer; 2016.
- 847 51. Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, de Leon S,  
848 Michelini K, Lewellen N, Crawford GE: **DNase I sensitivity QTLs are a major determinant**  
849 **of human expression variation.** *Nature* 2012, **482**:390-394.
- 850 52. Bates D, Mächler M, Bolker B, Walker S: **Fitting Linear Mixed-Effects Models Using lme4.**  
851 *Journal of Statistical Software* 2015, **67**:48.
- 852 53. Kuznetsova A, Brockhoff PB, Christensen RHB: **lmerTest Package: Tests in Linear Mixed**  
853 **Effects Models.** *Journal of Statistical Software* 2017, **82**:26.
- 854 54. Bass AJ, Dabney A, Robinson D: **qvalue: Q-value estimation for false discovery rate control.**  
855 **R package version 2.9.0, <http://github.com/jdstorey/qvalue>.**, vol. 2017; 2015.
- 856 55. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GOrilla: a tool for discovery and**  
857 **visualization of enriched GO terms in ranked gene lists.** *BMC Bioinformatics* 2009, **10**:48.
- 858 56. Shabalin AA: **Matrix eQTL: ultra fast eQTL analysis via large matrix operations.**  
859 *Bioinformatics* 2012, **28**:1353-1358.

- 860 57. Law CW, Chen Y, Shi W, Smyth GK: **Voom: precision weights unlock linear model analysis**  
861 **tools for RNA-seq read counts.** *Genome Biol* 2014, **15**:R29.  
862 58. Shen L: **GeneOverlap: An R package to test and visualize gene overlaps.** 2014.  
863 59. Yang J, Lee SH, Goddard ME, Visscher PM: **GCTA: a tool for genome-wide complex trait**  
864 **analysis.** *Am J Hum Genet* 2011, **88**:76-82.  
865 60. Bolormaa S, Hayes BJ, van der Werf JH, Pethick D, Goddard ME, Daetwyler HD: **Detailed**  
866 **phenotyping identifies genes with pleiotropic effects on body composition.** *BMC Genomics*  
867 2016, **17**:1.
- 868

## **Additional files:**

Additional file 1 (DOCX): Supplementary Methods.

Additional file 2 (PDF): Supplementary Figure S1-7.

Additional file 3 (XLSX): Supplementary Table S1. RNA-seq reads mapped to different bovine genome origins.

Additional file 4 (XLSX): Supplementary Table S2. Splicing junction annotation analysis using BAM files generated by different alignment software.

Additional file 5 (XLSX): Supplementary Table S3. 2x2 tables for the overlap between exon and intron analyses.

Additional file 6 (XLSX): Supplementary Table S4. Genes that display significant differential splicing cross tissues.

Additional file 7 (XLSX): Supplementary Table S5. Genes that display significant differential splicing between breeds.

Additional file 8 (XLSX): Supplementary Table S6. Summary of significant cis splicing sQTLs (within 1Mb distance to the exon).

Additional file 9 (XLSX): Supplementary Table S7. Summary table for count values of sQTLs and eeQTLs in white blood and milk cells.

Additional file 10 (XLSX): Supplementary Table S8. SNP overlap between expression QTLs and SNPs with pleiotropic effects on conventional traits of dairy cattle.

Additional file 11 (XLSX): Supplementary Table S9. SNP overlap between expression QTLs and SNPs with pleiotropic effects on conventional traits of beef cattle.

Additional file 12 (XLSX): Supplementary Table S10. SNP overlap between blood eeQTLs and putative causal variants identified by Littlejohn et al (2016).

## Figure Legends:

**Figure 1. a:** Sample principal components clustering based on exon expression. Circles on the plot were ellipses drawn based on tissue types at the confidence interval = 0.95. Tissue types with which the non-overlapping ellipses were drawn were emphasised with underscored text labelling. Ellipses that were drawn based on experiments can be found in Supplementary Figure S4. **b:** The significant splicing events between breeds and between genotypes (cis splicing quantitative trait loci, sQTLs) for *CSN3* in the milk cell transcriptome. In the upper panel, from left to right: the 1<sup>st</sup> pair of bars are the least square means of normalised expression level of the gene (ENSBTAG00000039787) in Holstein and Jersey breeds; the 2<sup>nd</sup> pair of bars are the normalised expression level of the 5<sup>th</sup> exon (6:87392578-87392750) in Holstein and Jersey breeds; the 3<sup>rd</sup> pair of bars are the normalised inclusion ratio of the 5<sup>th</sup> exon in Holstein and Jersey breeds; and 4<sup>th</sup> pair of bars is the frequency of the B allele of the sQTL (Chr6:87392580) for *CSN3* in Holstein and Jersey breeds. The standard errors bars are indicated. In the lower panel, from left to right: the 1<sup>st</sup> bar is the effects (signed t values, b/se) of the sQTL (Chr6:87392580) B allele on the normalised expression of the gene; the 2<sup>nd</sup> bar is the sQTL B allele effect on the normalised expression of the 5<sup>th</sup> exon; and the 3<sup>rd</sup> bar was the sQTL B allele effects on the normalised inclusion ratio of the 5<sup>th</sup> exon.

**Figure 2.** Manhattan plots of significant cis splicing quantitative trait loci (sQTLs, approximate FDR<0.01 and within 1Mb of the exon) in white blood cells (**a**), milk cells (**b**), liver tissue (**c**) and muscle tissue (**d**). A significant sQTLs was defined as a SNP associated with the variation in the exon inclusion ratio and also variation in at least one excision of an adjacent intron at the same significance level. The input SNPs had significance  $p < 0.0001$ .



sQTLs in all tissues with their associated genes and significance are given in Supplementary Table S6.

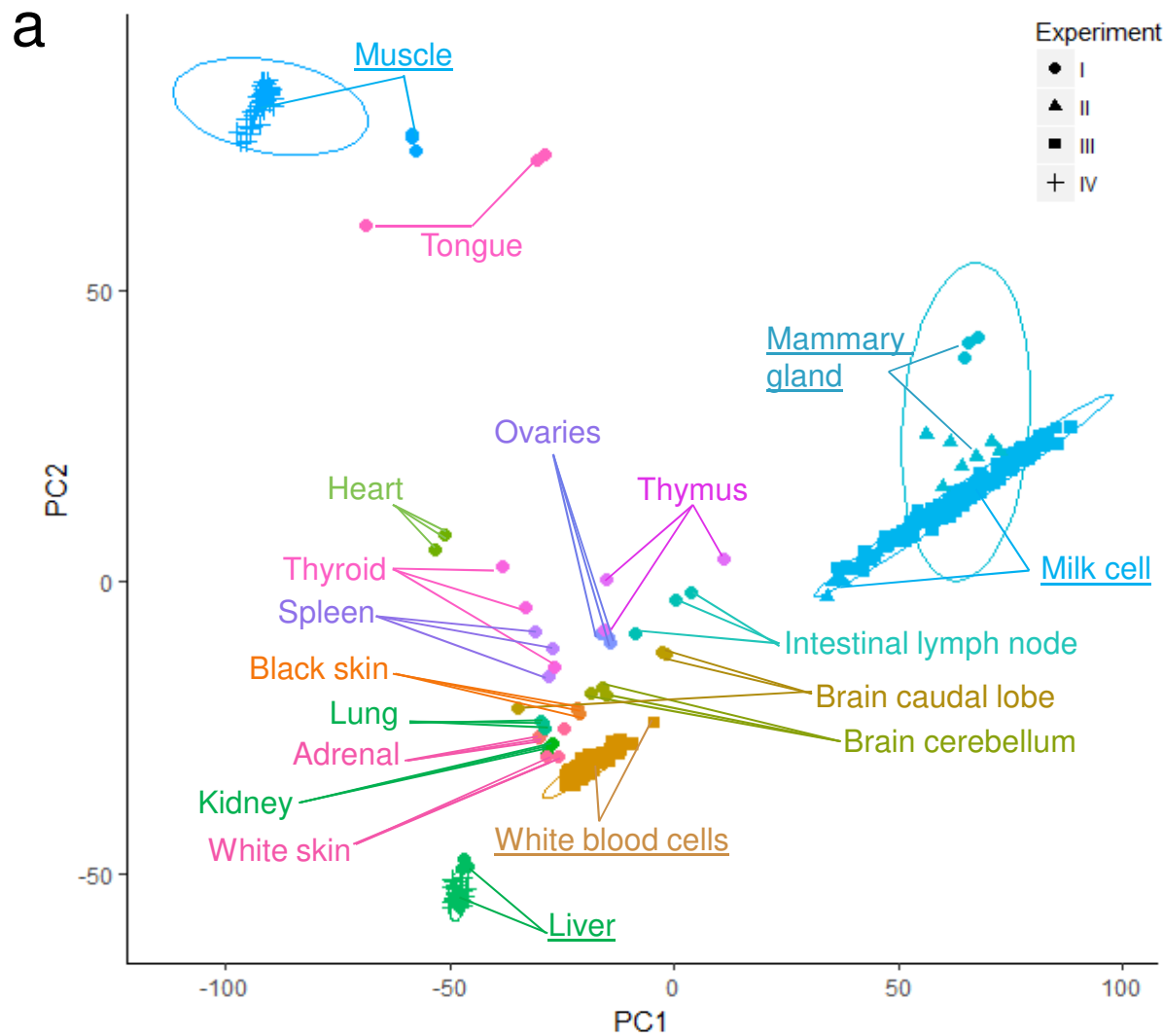
**Figure 3.** Features of cis splicing quantitative trait loci (sQTLs) compared to exon expression QTL (eeQTLs) and gene expression QTL (geQTLs). **a:** The distance between the transcription start site (TSS) and the expression QTLs. TSS information was downloaded from Ensembl (bovine reference UMD3.1). **b:** The proportion of expression QTLs annotated as splice, UTR, gene\_end, synonymous, missense, intron, intergenic or other. SNP annotations were based on Variant Effect Predictor. ‘Splice’ included all SNP annotations containing the word ‘splice’. ‘UTR’ included 3’ and 5’ untranslated region. ‘Gene\_end’ included upstream and downstream.

**Figure 4.** Overlaps of different expression QTL types for white blood cells (**a**), milk cells (**b**), liver(**c**) and muscle (**d**). Within each panel, y-axis was the number of significant expression QTLs; from left to right as guided by the green dots, the 1<sup>st</sup> bar indicated the number of significant cis splicing QTL (sQTLs); the 2<sup>nd</sup> bar indicated the number of significant exon expression QTL (eeQTLs); the 3<sup>rd</sup> bar indicated the number of significant gene expression QTL (geQTLs); the 4<sup>th</sup> bar indicated the number of SNPs identified as both geQTL and eeQTL; the 5<sup>th</sup> bar indicated the number of SNPs identified as both geQTL and sQTL; the 6<sup>th</sup> bar indicated the number of SNPs identified as both eeQTL and sQTL; and the 7<sup>th</sup> bar indicated the number of SNPs identified as geQTL and eeQTL and sQTL. The red colour indicates that the overlap between categories of expression QTLs was significantly more than expected by random chance based on Fisher’s exact test.

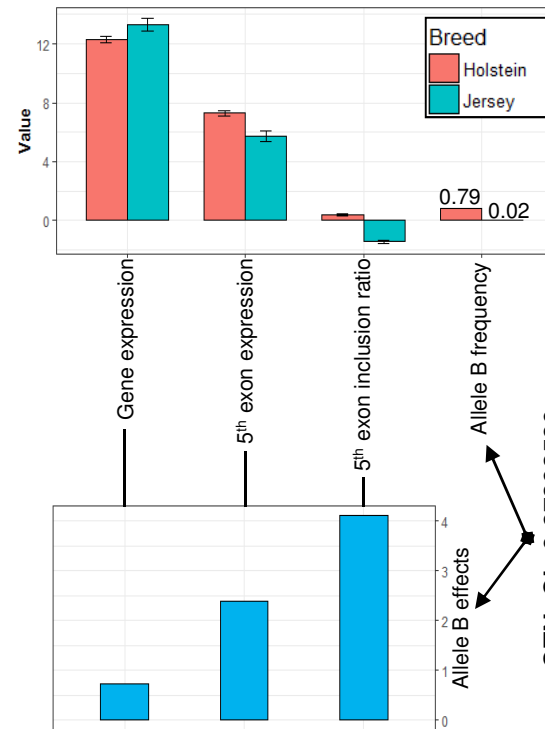
**Figure 5.** Shared genetic influence on the splicing, exon and gene expression between tissues. Blood refers to white blood cells and milk refers to milk cells. **a:** Each matrix shows the pair wise comparison of the numbers of significant SNP and the total number of significant SNPs detected for each analysis shown in parentheses. The significance of each overlap was tested by Fisher' exact test, given the total number of SNP analysed and the total number of significant SNP, the result of which is represented by the colour of that position in the matrix. **b:** Each matrix shows the pair wise comparison of the numbers of exon/gene with significant associations and the total number of exon/gene detected with significant associations for each analysis shown in parentheses. In panel **b**, the numbers were either exon numbers for sQTLs (splicing quantitative trait loci) and eeQTLs (exon expression quantitative trait loci) or gene numbers for geQTLs (gene expression quantitative trait loci. **c:** Between tissue genetic correlations of either the inclusion ratio of the exons, the expression of the exons or the expression of the genes that had significant sharing of expression QTLs in panel **a**. Dot size and transparency were negatively correlated with p value of the significance of the genetic correlation being different from 0. The error bars of the genetic correlation were shown in vertical lines of each dot. Some genes of interests were highlighted.

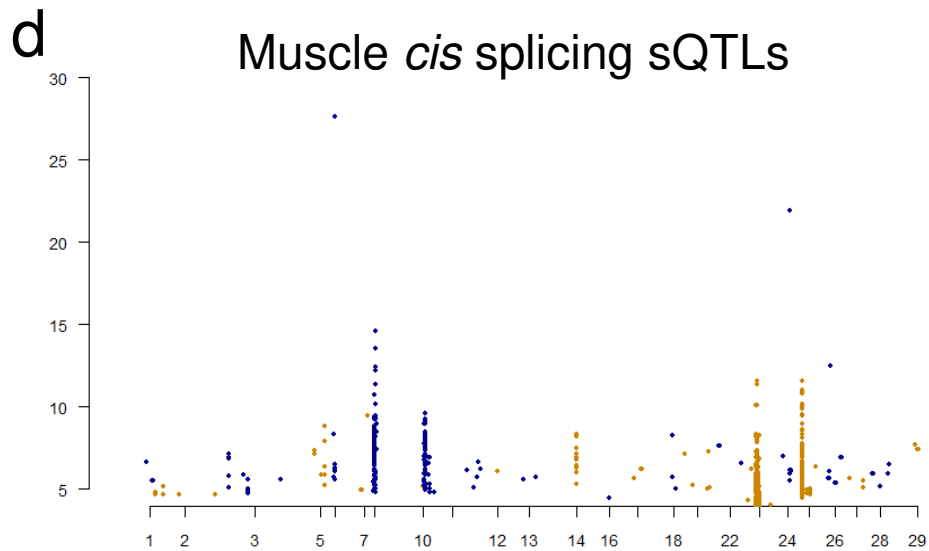
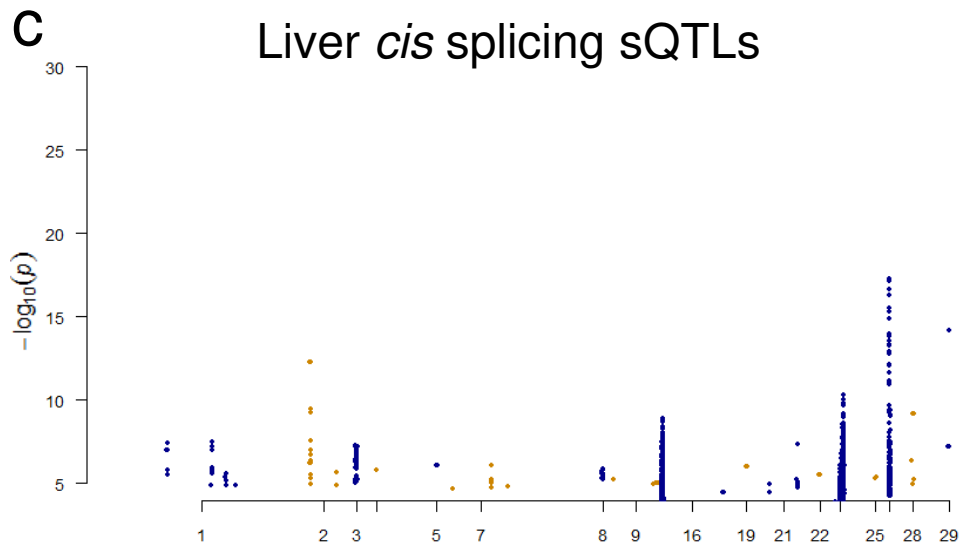
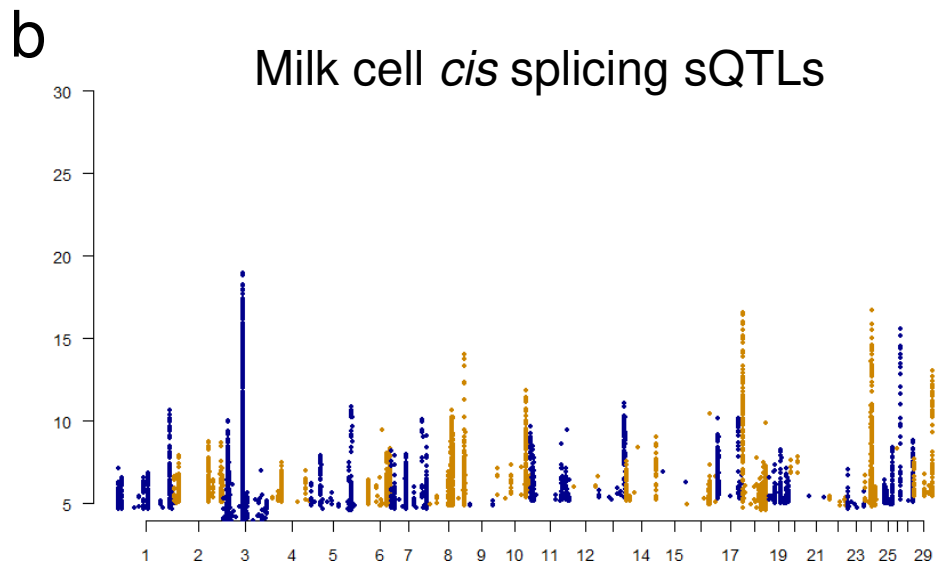
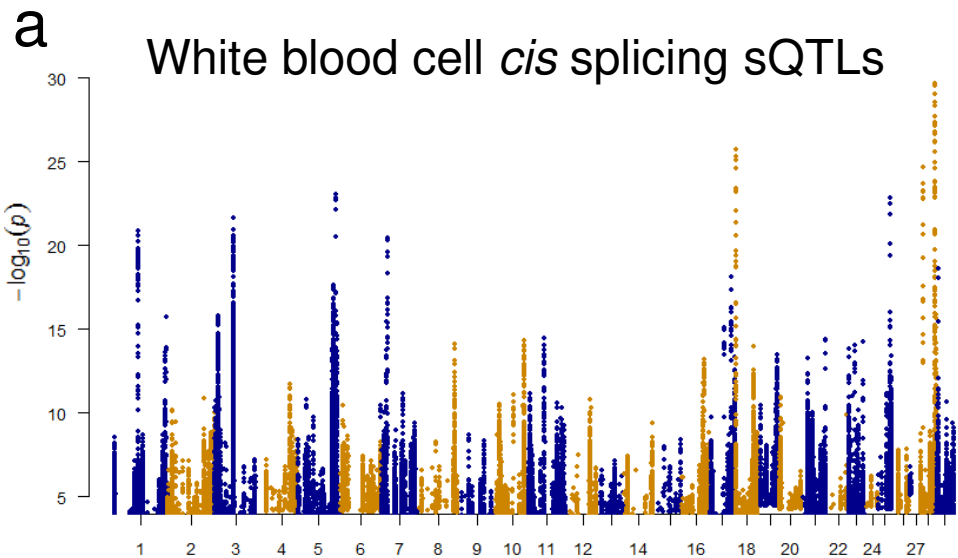
**Figure 6.** Multi-transcriptome meta-analysis (blood, milk and muscle) for cis splicing sQTLs (**a**), exon expression eeQTLs (**b**) and gene expression geQTLs (**c**). In each panel, the significance of multi-transcriptome effects were tested against a  $\chi^2$  with 1 degree of freedom for combined expression QTLs effects (dots in blue and orange). These multi-transcriptome effects were shown together with the single-transcriptome effects in liver of the same expression QTLs (dots in green).

**Figure 7.** Significance of the overlap, based on the Fisher's exact test, between pleiotropic QTL for a range of traits in cattle for dairy(**a**) and beef (**b**) and cis splicing quantitative trait loci (sQTLs), exon expression QTL (eeQTLs) and gene expression QTL (geQTLs) in all tissues, where the colour represents the significance of the overlap. Where blood refers to white blood cells and milk refers to milk cells. Significance of the overlap was based on the Fisher's exact test. Only chromosomes containing overlapping SNPs are shown. **c**) An example of *MGST1* showing the relationship between QTL effects on exon expression in milk cells and their effects on dairy cattle milk fat yield.

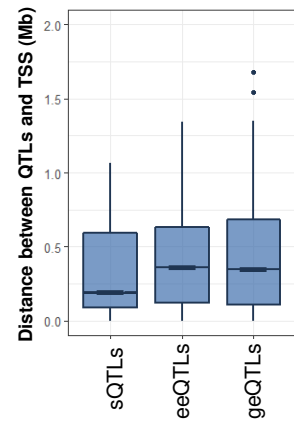
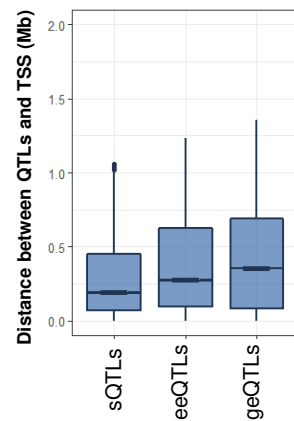
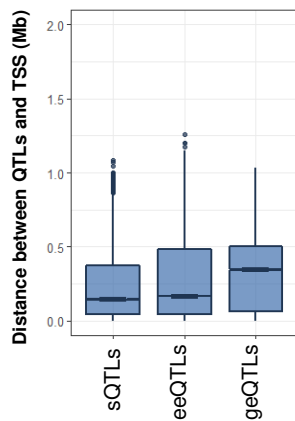
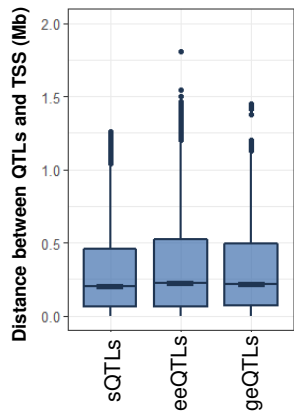


**b** *Kappa casein* in bovine milk cells





a



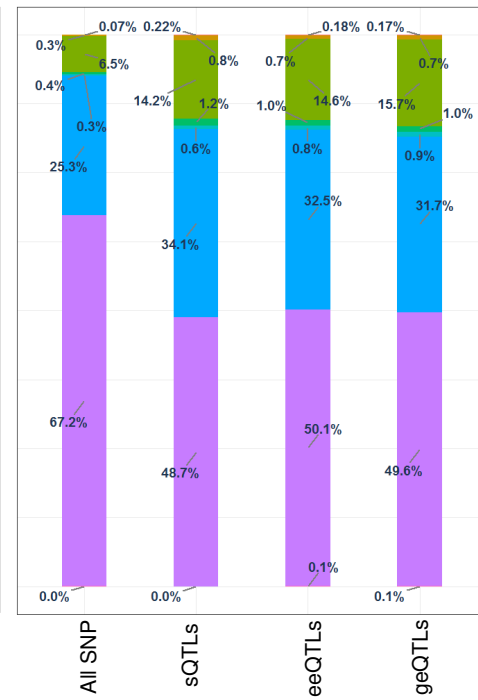
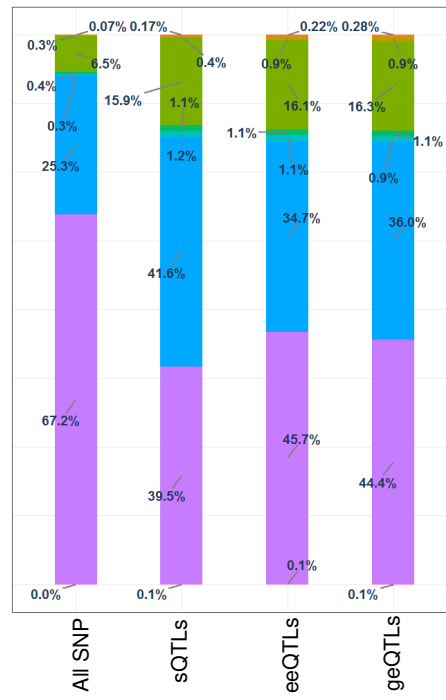
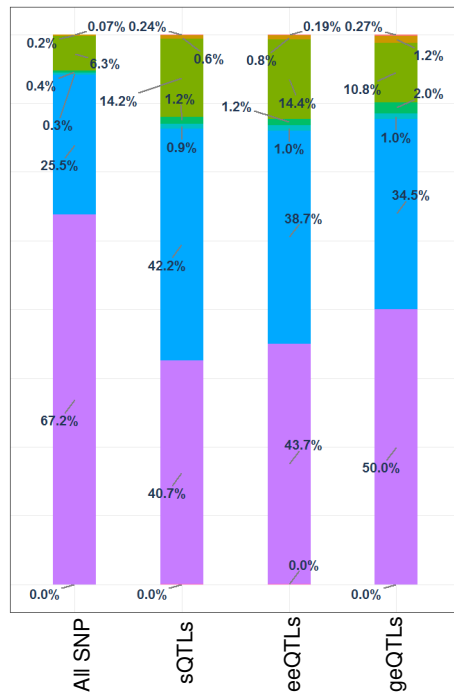
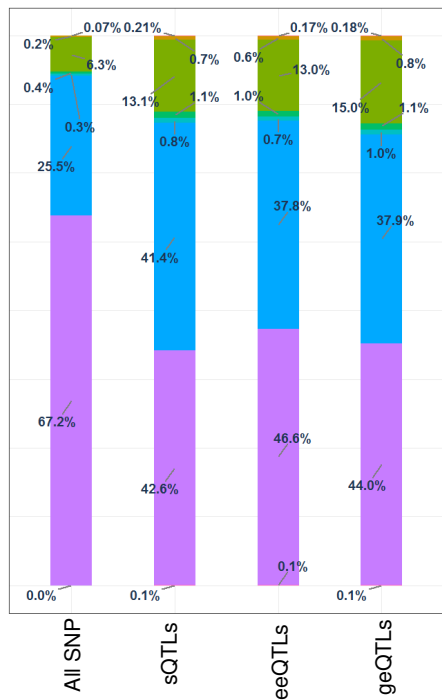
White blood cells

Milk cells

Liver

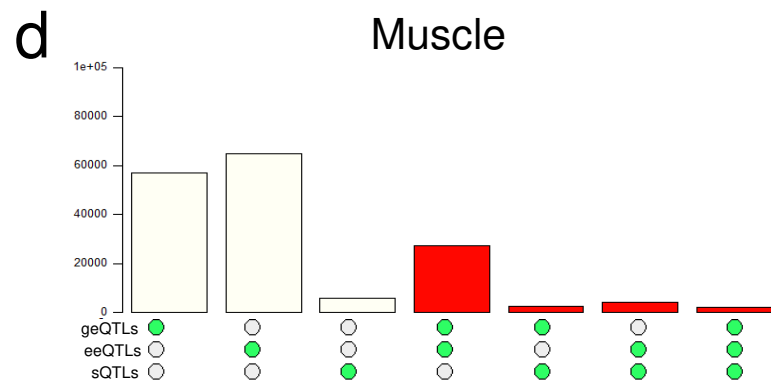
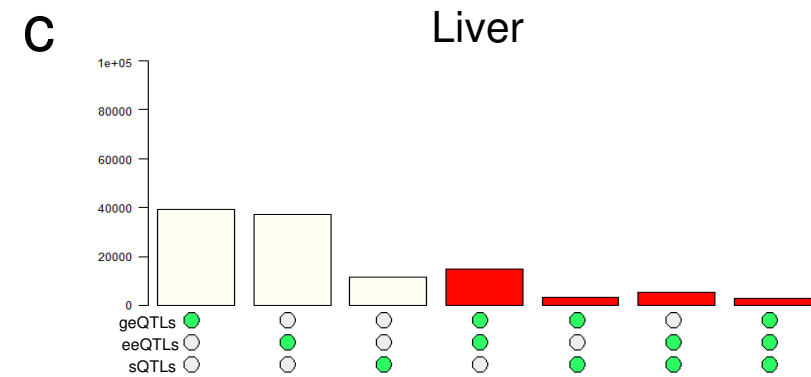
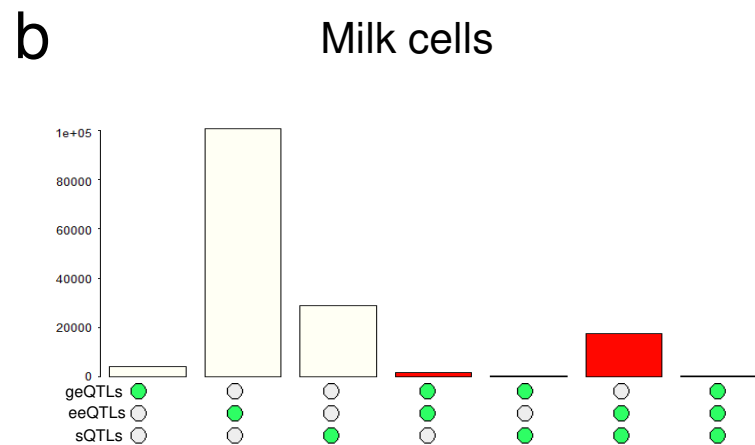
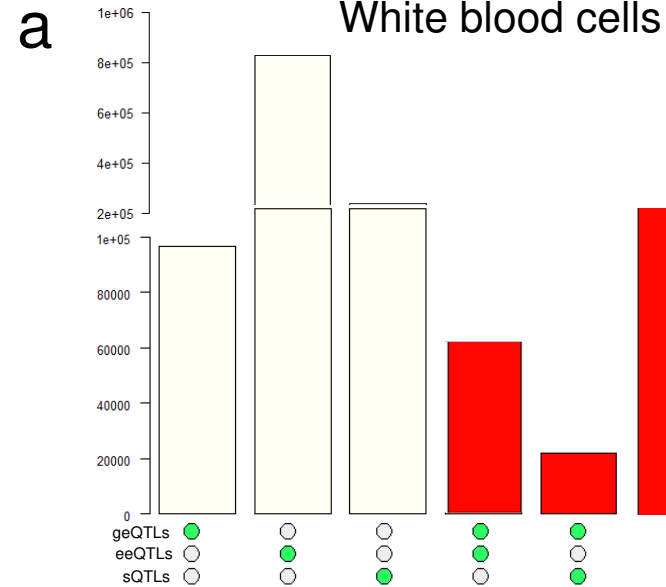
Muscle

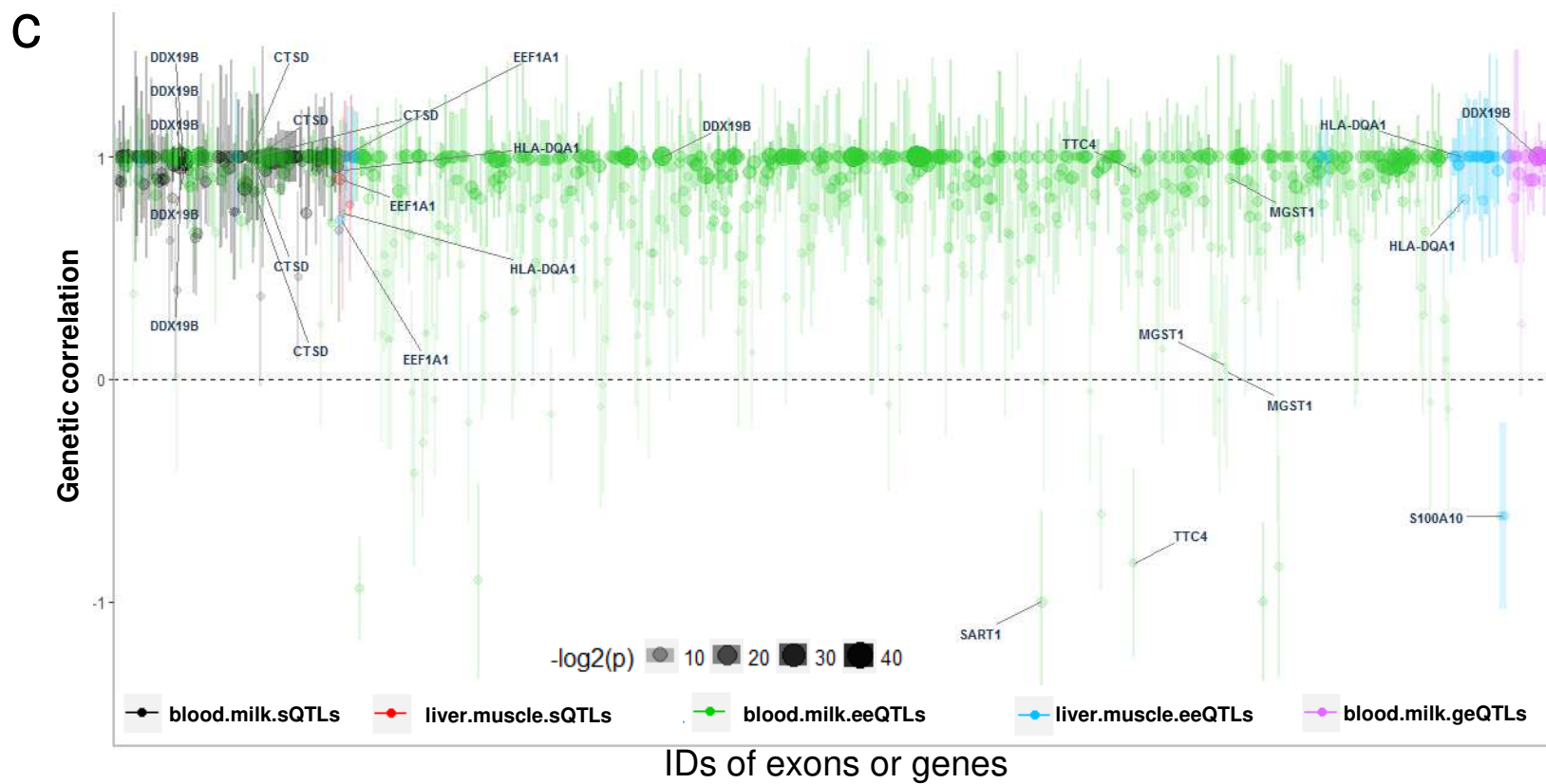
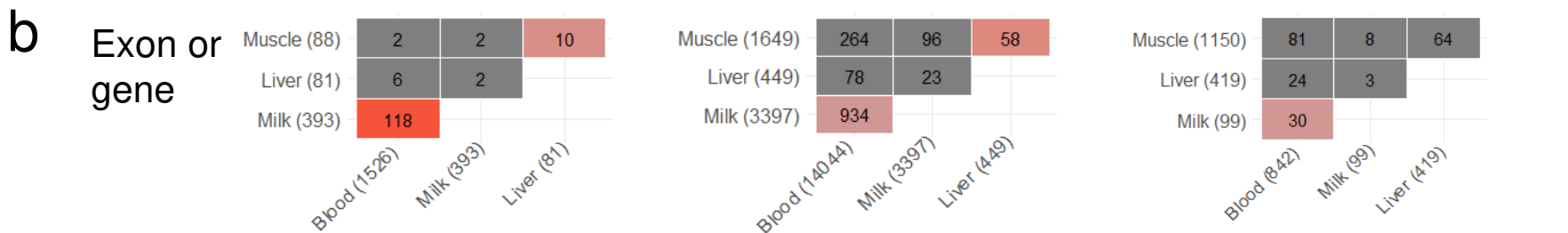
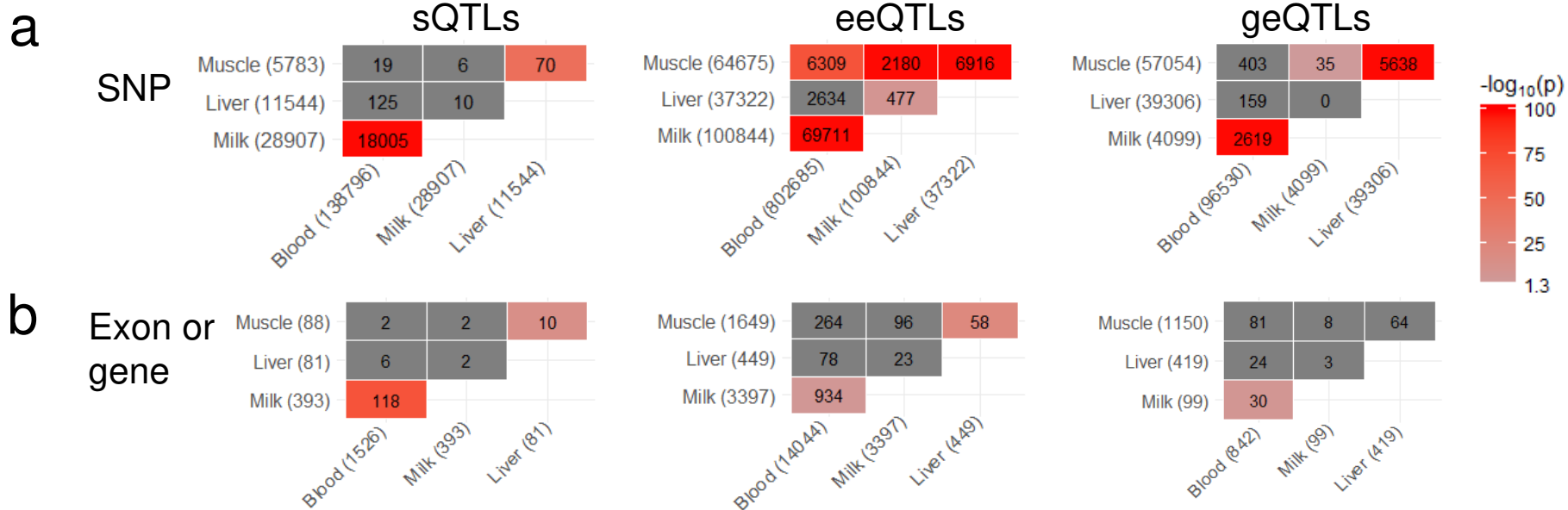
b



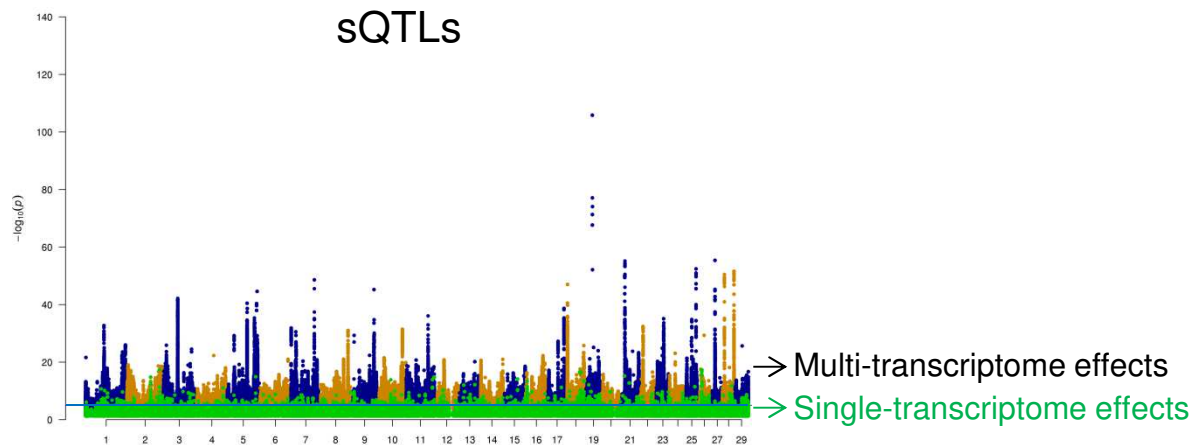
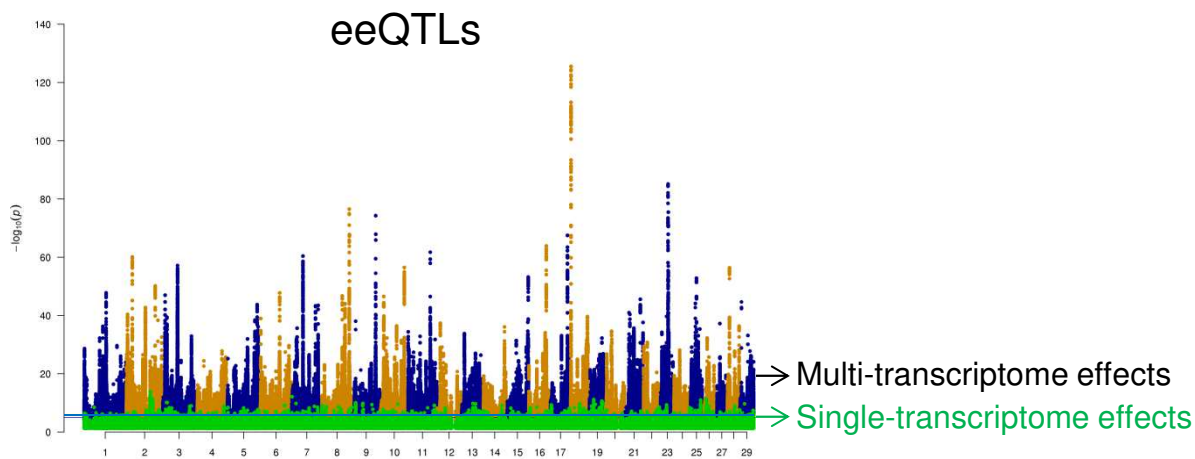
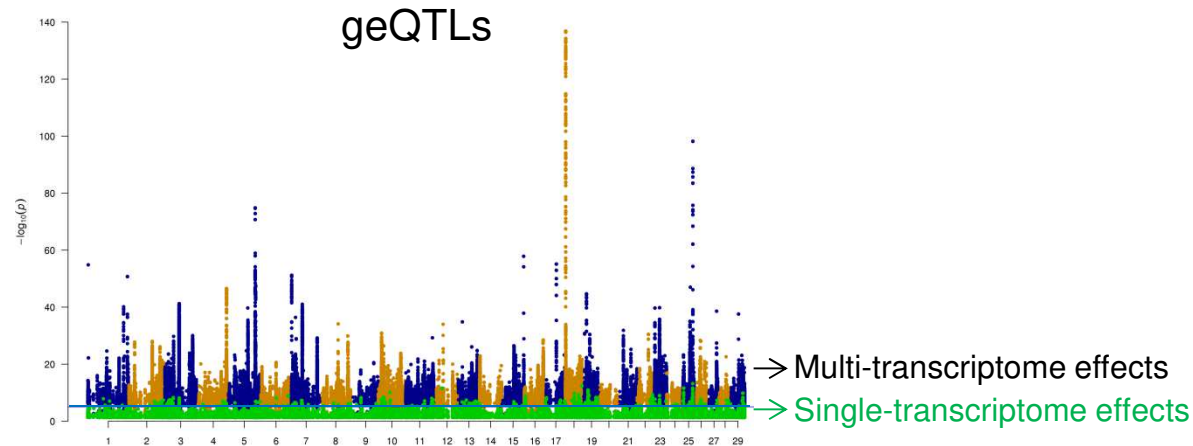
SNP annotation



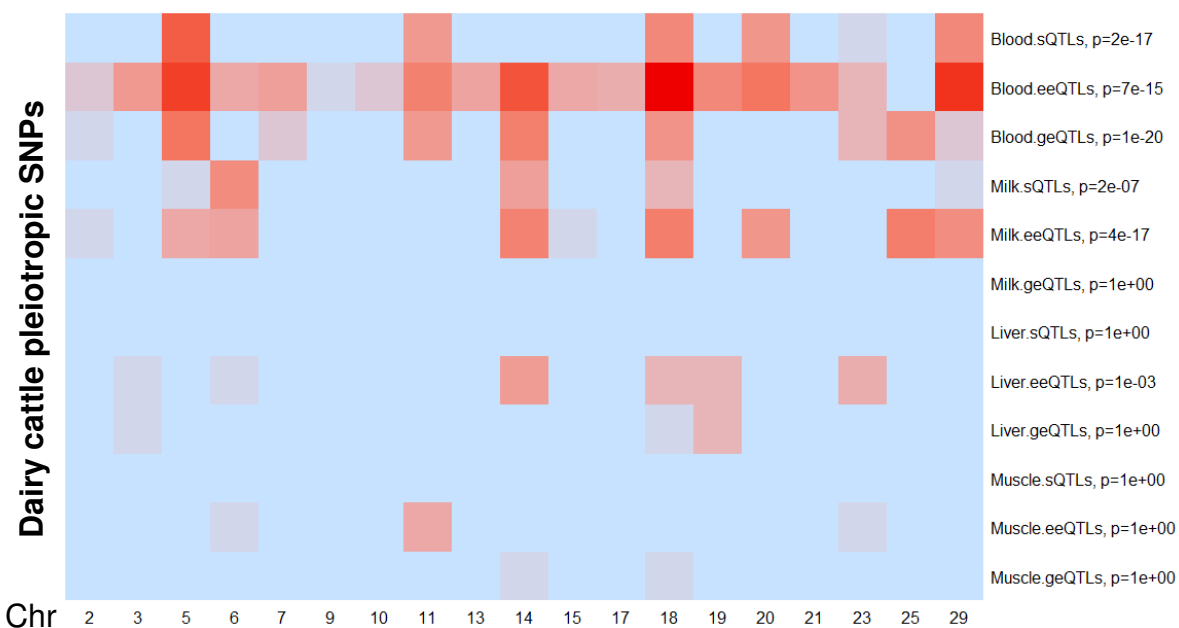




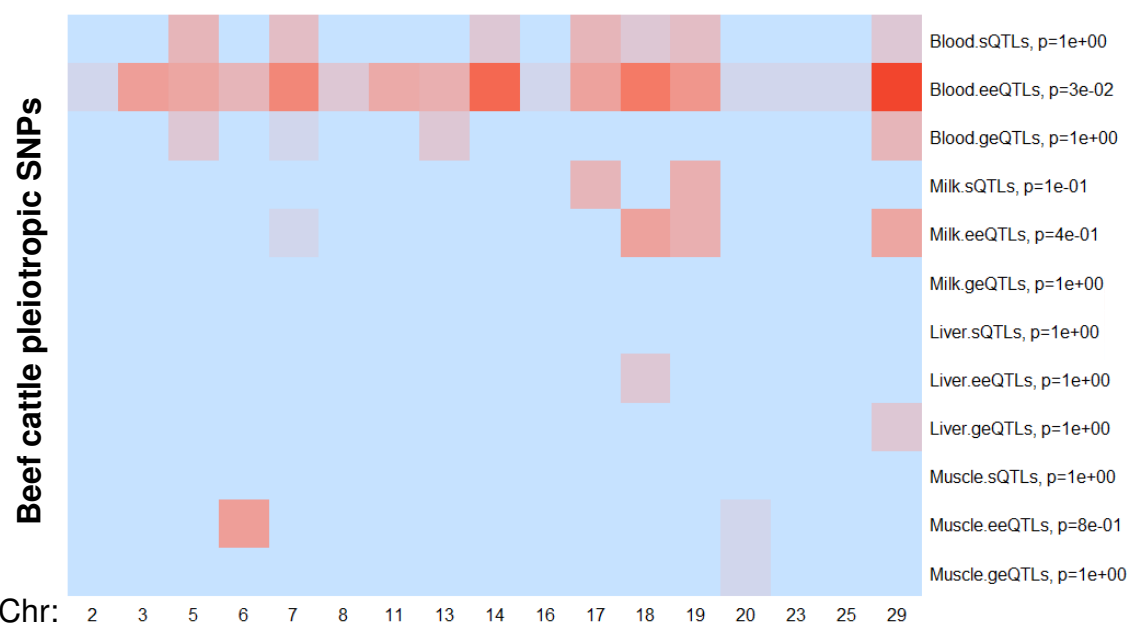


**a****b****c**

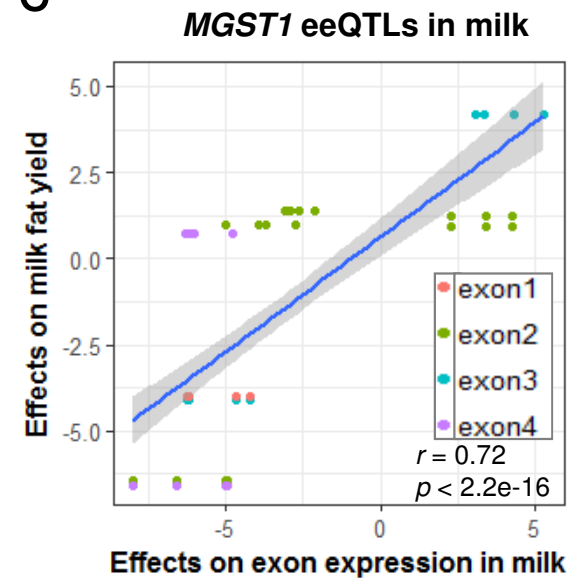
a



b



c



log2 SNP count

