


RESEARCH ARTICLE

Open Access



Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits

Aline S. M. Cesar^{1,2}, Luciana C. A. Regitano³, James M. Reecy², Mirele D. Poleti¹, Priscila S. N. Oliveira³, Gabriella B. de Oliveira¹, Gabriel C. M. Moreira¹, Maurício A. Mudadu⁴, Polyana C. Tizioto¹, James E. Koltes², Elyn Fritz-Waters², Luke Kramer², Dorian Garrick⁵, Hamid Beiki², Ludwig Geistlinger³, Gerson B. Mourão¹, Adhemar Zerlotini⁴ and Luiz L. Coutinho^{1*} 

Abstract

Background: Integration of high throughput DNA genotyping and RNA-sequencing data allows for the identification of genomic regions that control gene expression, known as expression quantitative trait loci (eQTL), on a whole genome scale. Intramuscular fat (IMF) content and carcass composition play important roles in metabolic and physiological processes in mammals because they influence insulin sensitivity and consequently prevalence of metabolic diseases such as obesity and type 2 diabetes. However, limited information is available on the genetic variants and mechanisms associated with IMF deposition in mammals. Thus, our hypothesis was that eQTL analyses could identify putative regulatory regions and transcription factors (TFs) associated with intramuscular fat (IMF) content traits.

Results: We performed an integrative eQTL study in skeletal muscle to identify putative regulatory regions and factors associated with intramuscular fat content traits. Data obtained from skeletal muscle samples of 192 animals was used for association analysis between 461,466 SNPs and the transcription level of 11,808 genes. This yielded 1268 *cis*- and 10,334 *trans*-eQTLs, among which we identified nine hotspot regions that each affected the expression of > 119 genes. These putative regulatory regions overlapped with previously identified QTLs for IMF content. Three of the hotspots respectively harbored the transcription factors *USF1*, *EGR4* and *RUNX1T1*, which are known to play important roles in lipid metabolism. From co-expression network analysis, we further identified modules significantly correlated with IMF content and associated with relevant processes such as fatty acid metabolism, carbohydrate metabolism and lipid metabolism.

Conclusion: This study provides novel insights into the link between genotype and IMF content as evident from the expression level. It thereby identifies genomic regions of particular importance and associated regulatory factors. These new findings provide new knowledge about the biological processes associated with genetic variants and mechanisms associated with IMF deposition in mammals.

Keywords: eQTL, Metabolic diseases, Fatty acids, Mammals, Gene expression

* Correspondence: llcoutinho@usp.br

¹Department of Animal Science, University of São Paulo, Piracicaba, SP 13418-900, Brazil

Full list of author information is available at the end of the article



Background

High-throughput genotyping and gene expression analysis combined with increased computational capacity and robust statistical methods allows for the identification of expression quantitative trait loci (eQTL). This approach can identify genomic regions that are associated with gene expression level [1–3] and can help to elucidate molecular mechanisms whereby genomic variants exert their effects on phenotypes or disease incidence [4, 5].

Adipose tissue is the largest endocrine organ in the body [6]. Within skeletal muscle, adipose deposition can occur within (intramyocellular) and external to (extramyocellular) skeletal muscle fibers. Lipid stored in skeletal muscle plays important roles in metabolic processes such as energy homeostasis, expression and secretion of hormones and proinflammatory cytokines, and in signaling pathways [7].

In humans, excessive fat deposition in skeletal muscle has been associated with metabolic diseases such as obesity, diabetes and coronary heart disease [8]. In contrast, in bovine, swine, and sheep, intramuscular fat (IMF) is positively associated with meat quality and consumer satisfaction [9], and can affect the final product price. Furthermore, the quantity and the fatty acid profile of the IMF present in edible red meat have both been positively and negatively associated with human health [10–12].

It is important to gain additional knowledge about the biological processes associated with IMF deposition and composition because of the important role of IMF deposition in areas such as human health and meat quality. Previously, Uemoto et al. [13], Ishii et al. [14], and our group have reported the identification of several quantitative trait loci (QTL) and putative candidate genes associated with IMF using high density SNP chip data [15]. In addition, we reported the identification of differentially expressed genes and putative candidate regulatory genes associated with IMF from RNA sequencing data (RNA-Seq) [16]. Furthermore, we have reported the identification of differentially expressed genes and gene networks associated with fatty acid composition in skeletal muscle [17]. However, so far there is no information on how genetic variation can influence gene expression and phenotypic variation associated with IMF content traits and composition [18–20]. Additional information can be gleaned by integrating the data of gene expression and whole genome association. Thus, the aim of this study was to perform eQTL analysis to identify putative regulatory regions and transcription factors (TFs) associated with intramuscular fat (IMF) content traits.

Results

Phenotypic, genotypic and RNA-Seq data

After the application of quality control filters, 461,466 markers and 11,808 genes from 192 animals were used

in these analyses. The descriptive statistic of all phenotypes related to fat deposition and composition are shown in Additional file 1. The variance components and genomic heritability of these phenotypes for the whole population of this study have been estimated and previously reported by Cesar et al. [15] and Tizioto et al. [16]. The heritability of these phenotypes ranged from 0.09 to 0.46 (Table 1).

Identification of eQTLs and the genomic location of variants

In total, 10,334 trans-eQTLs (variants that were located more than 1 Mb from an associated gene) and 1268 cis-eQTL (variants located within 1 Mb of the associated gene) were identified herein (5% false discovery rate, FDR) (see Additional files 2 and 3) by MatrixeQTL R package. Figure 1 shows a scatter plot of gene location (Mb) relative to its significant corresponding eQTL position (Mb).

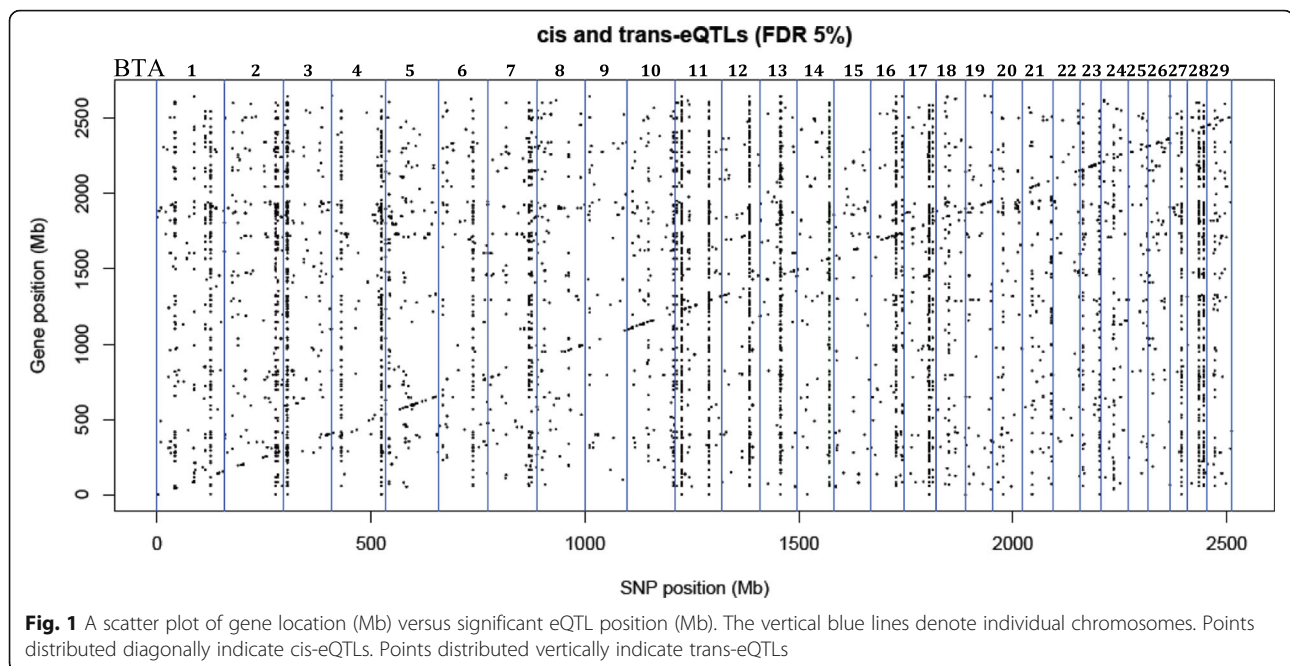
The cis-eQTLs were located in intronic (46%), intergenic (30%), upstream (11%), downstream (10%), and 3' UTR (1%) regions (see Additional file 4A). The trans-eQTLs were located in intergenic (61%), intronic (30%), upstream (4%) and downstream (4%) regions (see Additional file 4B). Analysis of the associated SNPs located in coding regions (less than 1 %) revealed that 66 and 34% of cis-eQTLs (see Additional file 5) and 52 and 48% of trans-eQTLs (see Additional file 6) were either synonymous (not causing a change in the protein sequence) or missense (causing a change in the protein sequence) variants, respectively.

Table 1 Posterior means of variance components for IMF deposition and composition in Nellore by Bayes B

Trait	Genetic variance	Residual variance	Total variance	Genomic heritability ^a
BFT (mm) ^b	0.78	2.87	3.65	0.21
IMF (%)	0.16	0.48	0.64	0.25
Myristic (%)	0.06	0.23	0.29	0.2
Myristoleic (%)	0.03	0.02	0.05	0.25
Palmitic (%)	2.42	5.47	7.89	0.31
Palmitoleic (%)	0.1	0.3	0.4	0.24
Stearic (%)	1.23	5.28	6.51	0.19
Oleic (%)	6.08	7.11	13.19	0.46
Linoleic (%)	0.00007	0.0003	0.00037	0.19
CLAc9t11 (%)	0.0003	0.002	0.0023	0.12
SFA (%)	1.36	15.54	16.9	0.08
MUFA (%)	1.67	15.18	16.85	0.1
PUFA (%)	0.09	0.54	0.63	0.14
Sum_n-3 (%)	0.004	0.013	0.017	0.25
Sum_n-6 (%)	0.0005	0.003	0.0035	0.15

^aCesar et al. [15]

^bTizioto et al. [16]



Overlap test between eQTLs and QTLs

Previous studies observed that many QTL are not located within coding regions [17, 21, 22], which might indicate that the causative mutations control gene expression level rather than altering gene function as would a non-synonymous exonic mutation. To determine if the eQTLs identified in this study overlap (see Additional file 7) with annotated QTL in Cattle QTLdb [23], a permutation test (p -value < 0.05) was performed using the regioneR package [23]. We observed that 2920 eQTLs (24%) overlapped QTLs in the Cattle QTLdb [23] with a total basewise overlap of 927.2 Mb, which corresponds to 61% of the genomic regions covered by eQTLs. In 47.4% of the cases, the eQTLs completely overlapped with QTL regions from Cattle QTLdb. A permutation test was conducted to evaluate if eQTL overlapped with each of the general QTL classes reported in Cattle QTLdb. In all cases, eQTL overlapped QTL in all general QTL classes to a significantly larger extent than expected by chance. The QTLs (see Additional file 8) associated with deposition and composition of intramuscular fat previously reported by our group [15] (Fig. 2) as well as QTLs associated with beef production and carcass and beef quality also had an overlap with eQTLs significantly larger than expected by chance.

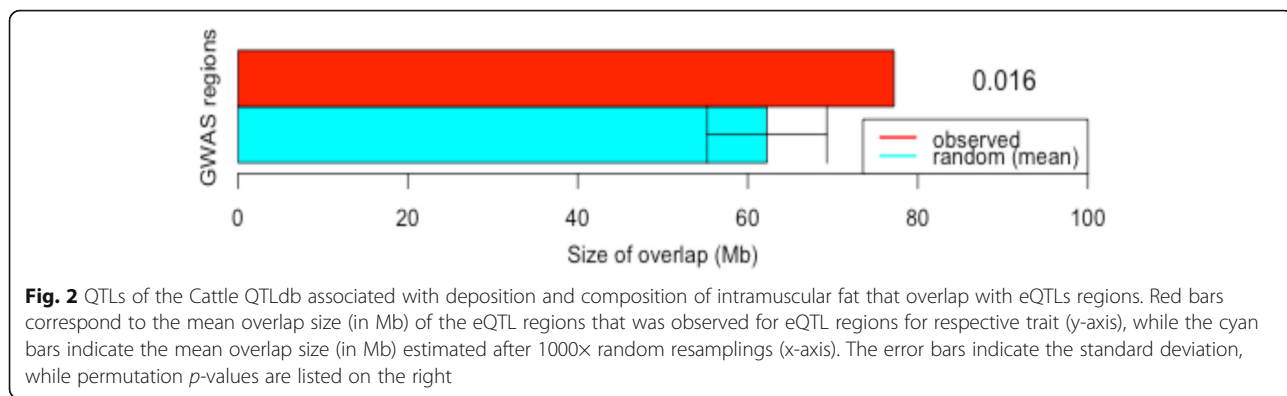
Identification of eQTL hotspots, their functional annotation and effect on IMF content traits

Mutations that influence the expression of several genes (eQTL hotspots) can modulate metabolic pathways and as a result can cause changes in phenotype [25]. Based on this definition of a hotspot, a permutation test was performed ($p < 0.05$), which indicated that a given variant needed to be

associated with the expression level of at least 119 genes. We identified 12 significant hotspots on BTA3, BTA4, BTA11, BTA14, BTA16, BTA17 and BTA28 (Fig. 3). Because some of these eQTL hotspots were located near each other, we examined the linkage disequilibrium (LD) among them and refined the eQTL hotspot regions to nine, located on BTA3 (8 Mb), BTA4 (108 Mb), BTA11 (11 Mb), BTA14 (73 Mb), BTA16 (59 Mb), BTA17 (55 Mb) and BTA28 (20 Mb, 32 Mb). The LD observed for hotspots on BTA11, BTA28, and also between each pair of the hotspots are shown in Additional files 9 and 10, respectively. The association test to verify the effect of all eQTL hotspots on the IMF content traits were performed using an ANOVA test (Table 2) (see Additional file 11) with correction for multiple tests by calculating the false discovery rate (FDR 5%).

Among the 244 genes identified within the hotspots (4 Mb windows), there were four transcription factors: *EGR4*, *RUNX1T1*, *NR1I3* and *USF1*; four miRNAs: *bta-mir-2294*, *bta-mir-1584*, *bta-mir-2322*, *bta-mir-584-3*; two nuclear receptors: *NCOR2* and *NRBF2*; three small nuclear RNAs: *SNORA70e*, *SNORA19* and *SNORA9*; and several important genes associated with cellular signaling and translation initiation such as: *CACYBP*, *CASP2*, *DIALLO*, *FSBP*, *GEM*, *RGS5* and *EIF2B1*.

In order to better understand the underlying biological processes associated with the hotspot eQTLs, PANTHER [26] was used to perform functional enrichment analysis, which uses the over-represented test (nominal p -value ≤ 0.05 , before correction for multiple tests of pathways). In this analysis, the list of all genes expressed in skeletal muscle (data from this study) was used as the background or reference gene list. Additional enrichment analysis was



conducted with the list of all annotated genes harbored within 4 Mb (2 Mb for each side) of cis (243 genes), trans (1453 genes) and hotspot (133 genes) eQTL regions (see Additional file 12A, B and C, respectively). Based on the results from functional enrichment, we suggest that the genes harbored within the eQTL regions are likely to be involved in modulation of transcription, translation and catalytic activity based on the GO terms associated with the genes in eQTL regions (see Additional file 12).

Transcription factor binding sites

Transcription factors modulate gene expression by binding to specific DNA regions [27]. To confirm that the identified

transcription factors found in the eQTL hotspot regions (*EGR4*, *RUNX1T1* and *USF1*) were modulating the differentially expressed genes, we determined the presence of TFBS in the promoter regions of the affected genes by LASAGNA (Length-Aware Site Alignment Guided by Nucleotide Association, algorithm) Search 2.0 [28]. Based on LASAGNA results, about 98% of the promoters of genes in regions associated with a hotspot region had a TFBS (*p*-value ≤ 0.05) for the respective TF (see Additional files 13, 14 and 15, respectively). The Circos plot in Fig. 4 shows the links between the eQTL hotspots that contained the transcription factors, *EGR4*, *USF1*, and *RUNX1T1*, and their associated target genes. The over-represented

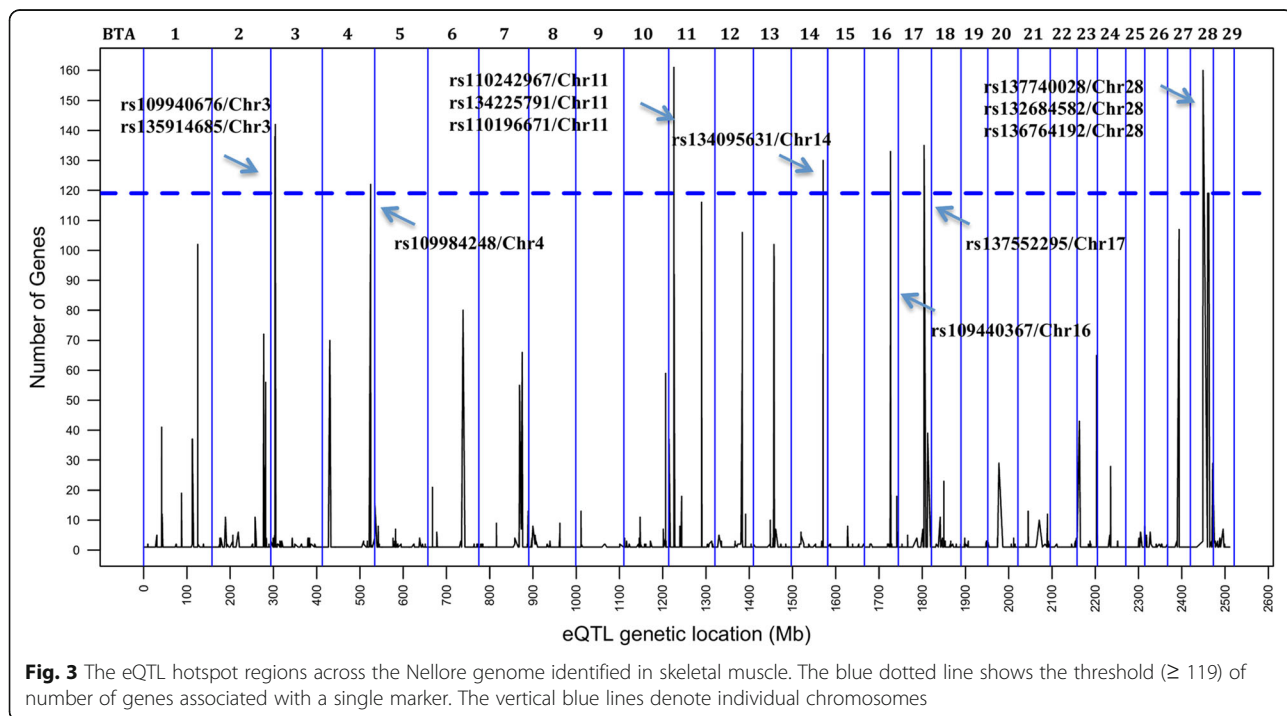


Table 2 Association test between the three eQTL hotspots that harbor TF in their region, and IMF content traits

eQTL hotspots Traits	rs135914685- hotspot 1(<i>USF1</i>)		rs110242967 - hotspot 2 (<i>EGR4</i>)		rs134095631 - hotspot 3 (<i>RUNX1T1</i>)	
	<i>p</i> -value	FDR ¹	<i>p</i> -value	FDR	<i>p</i> -value	FDR
BFT ^a	1.89E-06	7.47E-06 ⁱ	2.77E-06	8.30E-06 ⁱ	8.30E-06	2.07E-05 ⁱ
IMF ^b (%)	3.55E-02	6.61E-02	7.61E-02	8.15E-02	3.24E-02	3.74E-02 ⁱ
Myristic (%)	5.45E-04	2.30E-03 ⁱ	1.51E-05	2.84E-05 ⁱ	1.40E-05	2.63E-05 ⁱ
Myristoleic (%)	2.73E-05	1.14E-04 ⁱ	3.61E-04	5.41E-04 ⁱ	5.65E-04	7.71E-04 ⁱ
Palmitic (%)	5.37E-07	7.47E-06 ⁱ	5.34E-07	2.00E-06 ⁱ	5.55E-07	2.08E-06 ⁱ
Palmitoleic (%)	1.21E-05	5.45E-05 ⁱ	9.73E-06	2.09E-05 ⁱ	1.20E-05	2.58E-05 ⁱ
Stearic (%)	1.38E-01	1.89E-01	1.21E-01	1.21E-01	1.61E-01	1.61E-01
Oleic (%)	2.04E-14	5.17E-13 ⁱ	2.11E-14	1.06E-13 ⁱ	1.76E-14	8.82E-14 ⁱ
Linoleic (%)	2.74E-02	4.27E-02 ⁱ	3.65E-02	4.21E-02 ⁱ	4.22E-02	4.53E-02 ⁱ
CLAc9t11 ^c (%)	2.20E-16	3.30E-15 ⁱ	2.20E-16	3.30E-15 ⁱ	2.20E-16	3.30E-15 ⁱ
SFA ^d (%)	4.78E-06	3.75E-05 ⁱ	4.62E-06	1.16E-05 ⁱ	4.09E-06	1.23E-05 ⁱ
MUFA ^e (%)	1.72E-05	8.81E-05 ⁱ	1.66E-04	2.76E-04 ⁱ	1.41E-04	2.35E-04 ⁱ
PUFA ^f (%)	3.09E-04	1.26E-03 ⁱ	4.03E-04	5.49E-04 ⁱ	4.03E-04	6.04E-04 ⁱ
Sum_n3 ^g (%)	2.20E-16	3.30E-15 ⁱ	2.20E-16	1.65E-15 ⁱ	2.20E-16	1.65E-15 ⁱ
Sum_n6 ^h (%)	1.24E-02	3.80E-02 ⁱ	1.10E-02	1.38E-02 ⁱ	9.68E-03	1.21E-02 ⁱ

^aBackfat thickness^bintramuscular fat^cconjugated linoleic acid cis9 trans11^dsaturated fatty acid^emonounsaturated fatty acid^fpolyunsaturated fatty acid^gsum of omega-3, and^hsum of omega-6ⁱSignificant FDR 5%¹FDR – Adjusted *p*-value using the false discovery rate method by Benjamini and Hochberg 1995

TFBS motifs of *EGR4*, *RUNX1T1* and *USF1* observed in the promoters of genes within hotspot eQTL regions are shown in Additional file 16.

Co-expression networks and correlation with traits

We used weighted correlation network analysis (WGCNA) to further explore how the three eQTL hotspot regions that harbor the TFs could affect gene expression and phenotype. For that, in each hot spot region, we selected the SNP with lowest *p* value to compare alternative genotypes for the hotspot. Unfortunately, we could not compare alternative homozygous genotypes because there were only two or three BB animals. So, the comparison was made between the homozygous AA vs heterozygous AB genotypes. This approach allowed us to identify the pattern of co-expressed genes assigned to various co-expression modules for both AA and AB genotypes and correlate them to the different traits related to IMF content traits. This correlation can represent the set of co-expressed genes that are associated with biological processes involved in lipid metabolic process in skeletal muscle.

The eQTL hotspot region that harbored the *USF1* TF (rs135914685) presented 17 and 25 modules for AA (*n* = 175) and AB (*n* = 17) genotype, respectively (Fig. 5a and b).

The *EGR4* TF (rs110242967) presented 18 modules for the AA (*n* = 174) and 16 for AB (*n* = 18) genotype, respectively (Fig. 6a and b). The eQTL hotspot region that harbored the *RUNX1T1* TF (rs134095631) presented 16 and 25 modules for AA (*n* = 176) and AB (*n* = 16) genotype, respectively (Fig. 7a and b). After estimating the correlation values (*r*) between the modules and each trait of interest, modules that were correlated with at least three different traits were selected for functional enrichment analysis (*p*-value < 0.1). While the number of animals with a given hotspot genotype was very similar among the hotspot groups, the set of animals within a given hotspot genotype was different. It is important to point out that there is no selection for fat deposition or composition in this population.

Hotspot 1 - rs135914685, BTA3: 8117390-10117390, *USF1*

The AA-modules named “black”, “blue”, “green”, “midnightblue”, “pink”, “purple”, “salmon”, and “tan” were correlated (*p*-value < 0.10) with at least three different traits as shown in Fig. 5a. The “midnightblue” module presented higher correlation values (greater than 0.17) and was significantly correlated (*p*-value < 0.10) with eight of the 15 phenotypes studied.

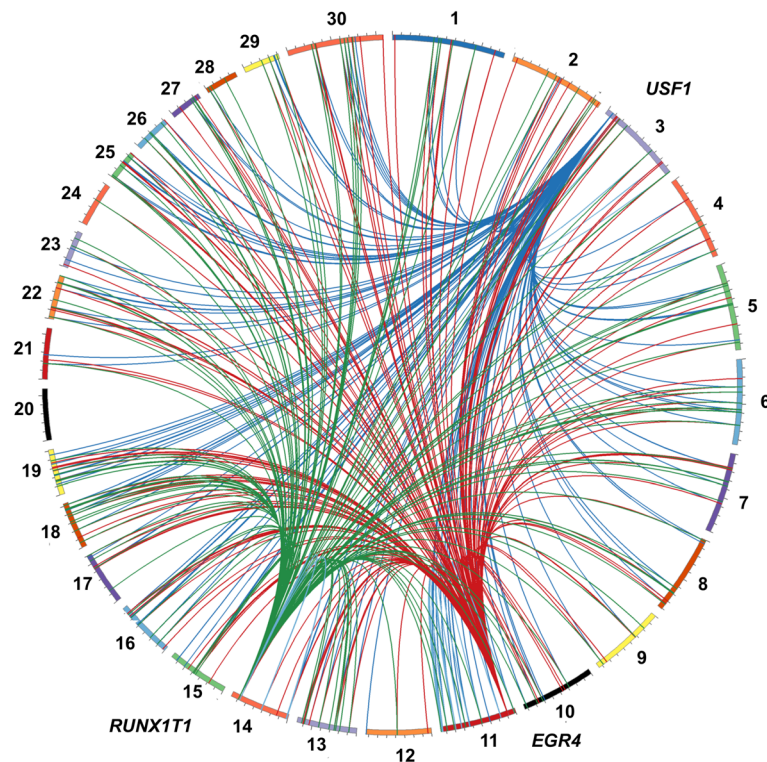


Fig. 4 Circos plot of eQTL hotspot regions that harbored the transcription factors (TF) *EGR4* (rs110242967), *RUNX1T1* (rs134095631) and *USF1* (rs135914685). Colored boxes represent each chromosome while the colored lines show the association between the regions that harbor the TFs and the genes that were associated with these hotspot regions. *USF1* region connections are shown in blue; *EGR4* region connections are shown in red; *RUNX1T1* region connections are shown in green

For the AB-modules the “black”, “brown”, “cyan”, “darkturquoise”, “lightyellow”, “pink”, “purple”, “royalblue”, “salmon”, “tan”, and “turquoise” were correlated (p -value < 0.10) with at least three traits (Fig. 5b). Differently than what was observed in co-expression analysis of AA group, the AB-modules showed higher correlation values (negative or positive) for IME, palmitoleic, oleic and linoleic acids.

Hotspot 2 - rs110242967, BTA11: 10540044-12540044, *EGR4*

The “magenta” and “purple” AA-modules were correlated (p -value < 0.10) with at least three different traits as shown in Fig. 6a. “Purple” presented the highest correlation values (greater than 0.17) as well as a significantly correlated (p -value < 0.10) with eight of the 15 phenotypes studied.

“Yellow”, “brown”, “pink” and “blue” AB-modules were respectively correlated (p -value < 0.10) with the following traits: BFT ($r = -0.52$), myristic ($r = 0.50$) and sum of n3 ($r = 0.41$); stearic ($r = -0.35$), oleic ($r = 0.45$), CLA cis9 tran11 ($r = 0.55$), SFA ($r = -0.38$); IME ($r = -0.39$), sum of n3 ($r = -0.36$) and sum of n6 ($r = -0.42$); BFT ($r = -0.47$), CLA cis9 tran11 ($r = 0.46$) and palmitoleic ($r = 0.37$) (Fig. 6b).

Hotspot 3 - rs134095631, BTA14: 79693309-74693309, *RUNX1T1*

“Greenyellow”, “black”, “blue”, “magenta”, “midnightblue”, “pink”, “purple”, and “salmon” AA-modules were correlated (p -value < 0.10) were correlated (Fig. 7a) with at least three traits. “Darkgreen”, “darkred”, “green” and “yellow” AB-modules were correlated (p -value < 0.1) with at least three different traits as shown in Fig. 7b. As observed for the other two hotspots, the AB-modules showed higher correlation values (negative or positive), however for different traits such as BFT, myristic, palmitoleic, oleic acids and PUFA.

Hotspot’s functional enrichment analysis, network construction and visualization

Functional enrichment analysis was performed from a list of genes within each module that were significantly correlated (p -value < 0.10) with at least three different traits according to the Cytoscape plugin BINGO [29]. Network construction and visualization for each eQTL hotspot genotype (see Additional file 17, sheets 1, 2, 5, 6, 9, 10) (Figs. 8, 9 and 10) were performed by Cytoscape 3.5.1 [30] connecting the top 20 hub genes [most connected genes, higher values of the module membership

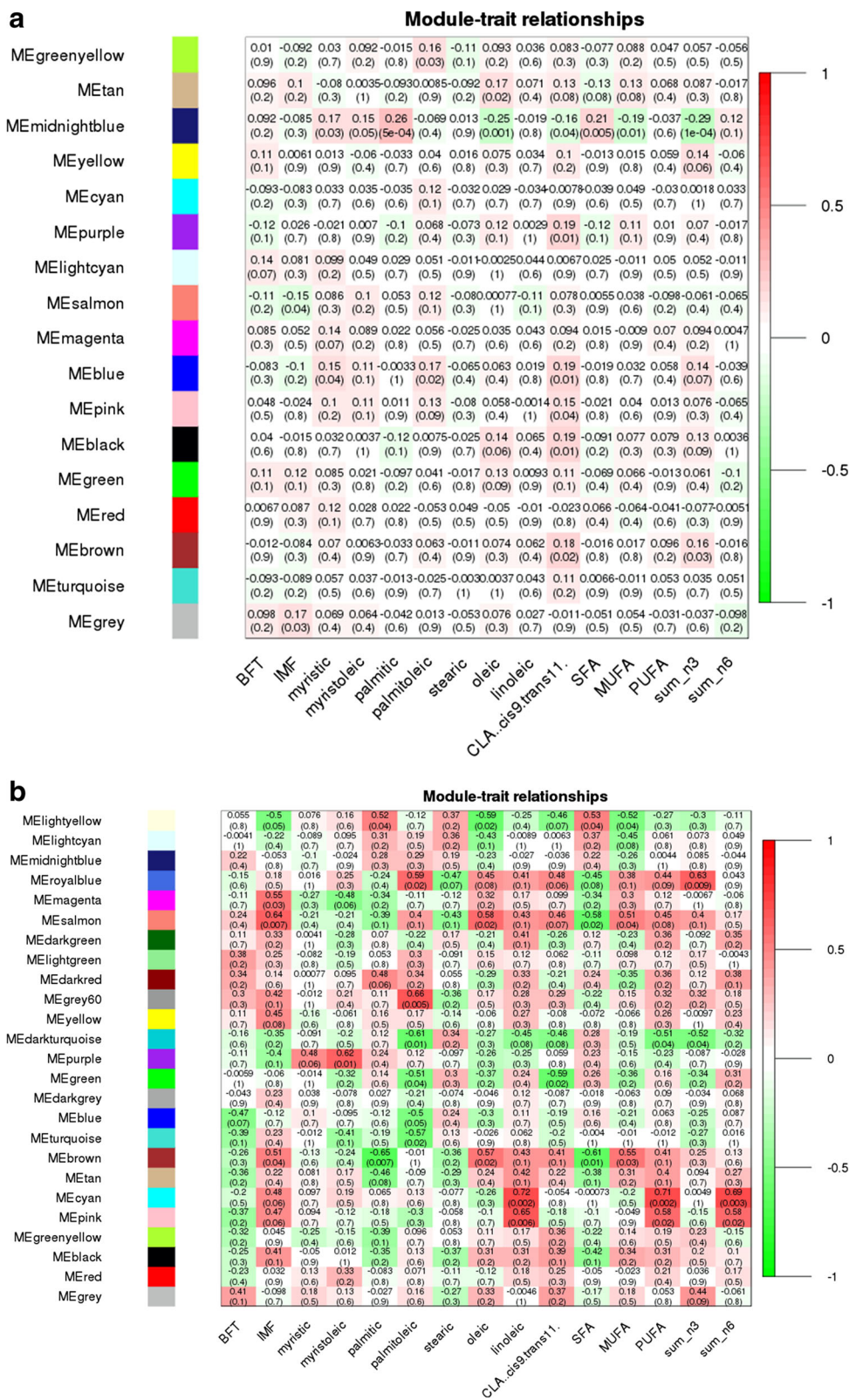


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 a Correlation between network module eigengene (ME) values and traits. Colors to the left represent the 17 gene expression modules identified from AA genotype (rs135914685) individuals. **b** Correlations between network module eigengene values and traits. Colors to the left represent the 25 modules in the AB genotype (rs135914685) network. For each module, the heatmap shows ME correlations to traits. Numbers in each cell report the correlation coefficients and student asymptotic p-value (parentheses) for significant ME-trait relationships. The scale bar, on the right, indicates the range of correlations from positive (red, 1) to negative (green, -1)

(MM)] of each module by the common significant (FDR 5%) biological process (BP) gene ontology (GO) terms.

Functional enrichment analysis for hotspot 1 identified the following BP GO term for the AA (Fig. 8a) genotype: immune system process, lipid metabolic process, cellular carbohydrate metabolic process, fatty acid metabolic process, electron transport chain, oxidation reduction, and mitochondrial ATP synthesis coupled electron transport. For the AB (Fig. 8b) genotype, the following BP GO terms were identified: translation, gene expression, regulation of transduction, regulation of cell communication, and cellular process (see Additional file 17, sheets 3 and 4).

For hotspot 2, the following BP GO terms were identified for AA (Fig. 9a) genotype: intracellular membrane-bounded organelle and cytoplasm, protein folding, response to temperature stimulus, muscle structure development, cell differentiation, striated muscle cell differentiation and muscle cell differentiation. BP were as follows for the AB (Fig. 9b) genotype: signal transduction, lipid metabolic process, carbohydrate metabolic process, cellular lipid metabolic process, fatty acid biosynthetic process, gene expression, lipid biosynthetic process, RNA metabolic process (see Additional file 17, sheets 7 and 8).

For hotspot 3, the following BP GO terms were identified for AA (Fig. 10a) genotype: electron transport chain, ATP metabolic process, protein binding, glucose metabolic process, and immune response. BP were as follows for the AB (Fig. 10b) genotype: and positive regulation of cell migration, T cell differentiation, protein binding, and lipid metabolic process (see Additional file 17, sheets 11 and 12).

Discussion

Identification of eQTL and their overlap with QTL regions

To our knowledge, this is the first eQTL analysis performed for complex phenotypes such as IMF content traits in bovine species. The association analysis between each of the 461,466 SNPs and the expression level of 11,808 genes expressed in muscle identified 1268 cis-eQTLs and 10,334 trans eQTLs that affected 243 and 1453 genes, respectively. These results agree with Wittkopp and Kalay [31] that described cis regulators that affected only a few genes and trans regulators having pleiotropic effect on many genes. The higher number of

trans-eQTLs than cis-eQTLs was also reported by Ramasamy et al. [32].

The overlapping test between eQTL and previous QTL regions support the hypothesis that many SNP-trait associations are mediated by changes in the expression level. Previous studies demonstrated that untranslated gene regions (UTR), such as 5' and 3' UTRs, introns and intergenic regions are involved in the regulation of expression [33] and that variation within these regions produce phenotypic variation [34–36].

In this study, although we used a panel of SNPs, in which most of the mutations are in introns and intergenic regions, we observed an enrichment of the intergenic regions for the identified trans-eQTLs. These findings corroborate the expected biologic function of the trans-eQTLs as potential distant regulators of gene expression [37].

Functional annotation of genes located near (2 Mb) cis and trans-eQTLs revealed biological processes GO terms such as transcription factor binding, protein binding, translation regulatory and transporter activities. Because of the presence of regulatory genes near eQTL regions, we suggest that these regions could be involved in modulation of gene expression and thus influencing quantitative traits [38, 39].

Hotspot eQTLs and transcription factors

Cis-eQTLs are in general considered more important than trans-eQTL because of their local activity [40, 41]; however, recent studies have demonstrated that both cis- and trans-eQTLs are important to better understand the expression variation in different species [42, 43]. Among the trans-eQTL, some were associated with the expression levels of many different genes [see Additional file 3], which were defined as eQTL hotspots as described previously [44, 45]. This pleiotropic effect can be explained by the presence of a TF in the hotspot eQTL region [46]. Even though the annotation of the bovine genome for TF is not complete, we identified three annotated (JASPAR database) [47] TFs near (2 Mb) the following eQTL hotspots: USF1 (BTA3: 8117390–10117390), *EGR4* (hotspot on BTA11: 10540044–12540044) and *RUNX1T1* (BTA14: 79693309–74693309). The involvement of these TF was supported by the presence of TFBS in the promoter region of differentially expressed genes. In another study, authors also demonstrated that many of

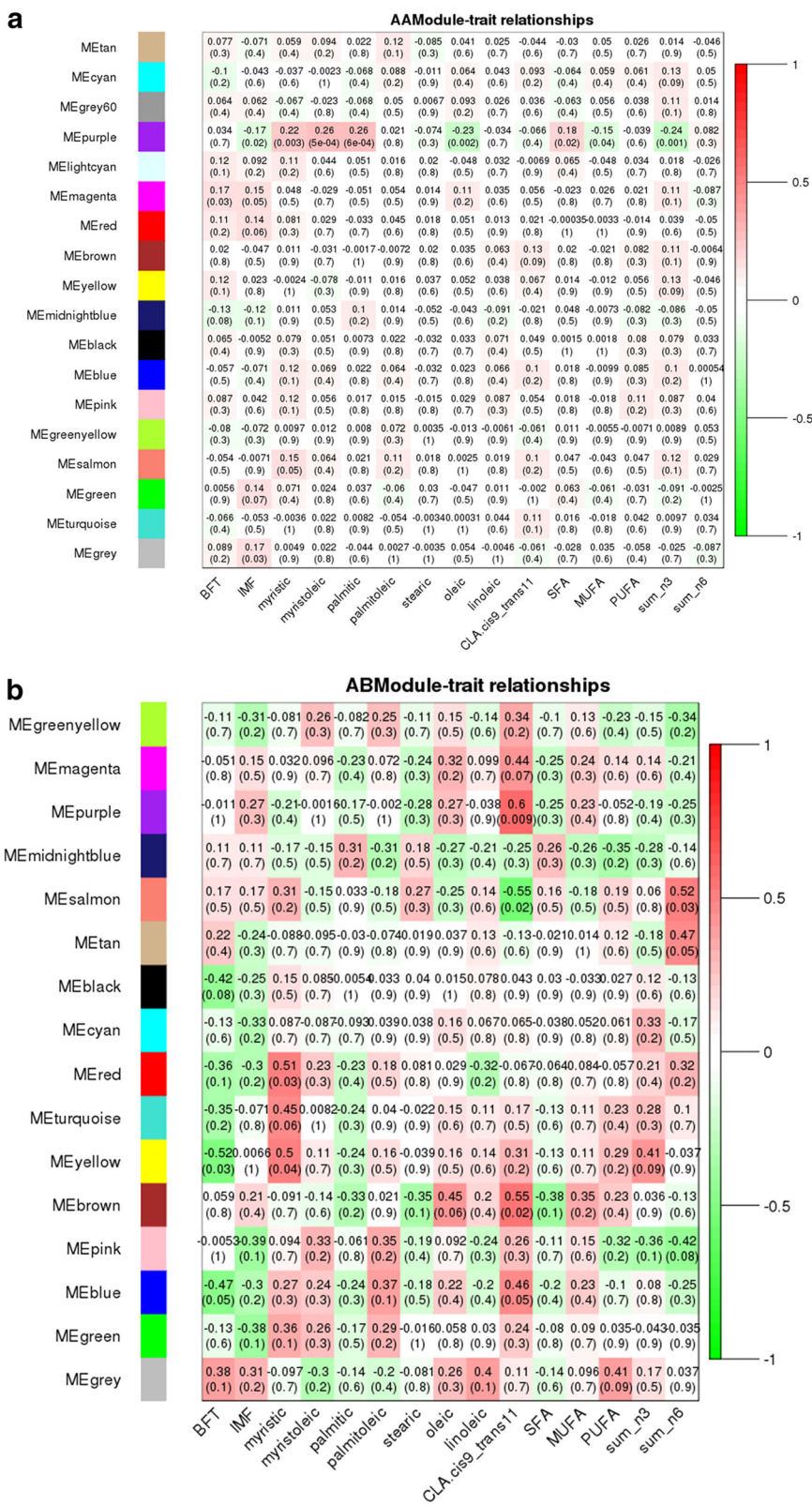


Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 a Correlation between network module eigengene values and traits. Colors to the left represent the 18 modules in the AA genotype (rs110242967) network. **b** Correlation between network module eigengene (ME) values and traits. Colors to the left represent the 16 modules in the AB genotype (rs110242967) network. For each module, the heatmap shows ME correlations to traits. Numbers in each cell report the correlation coefficients and student asymptotic p-value (parentheses) for significant ME-trait relationships. The scale bar, on the right, indicates the range of correlations from positive (red, 1) to negative (green, -1)

the harbored TFs identified within trans-eQTLs regions mediate the effect of inheritance of these loci on gene expression levels [46, 48]. In addition, the association test performed between these eQTL hotspots and the phenotypes confirmed the significant (FDR 5%) effect of these eQTLs on phenotypic variances of IMF content traits (Table 1).

The three annotated TFs identified near the eQTL hotspots were previously reported to be associated with lipid metabolism. *USF1* (Upstream Transcription Factor 1) encodes a member of the basic helix-loop-helix leucine zipper family and has been linked to familial combined hyperlipidemia (*FCHL*) [49]. A recent report pointed *USF1* as a new molecular link between lipid metabolism and energy expenditure, which is a potential therapeutic target for cardiometabolic disease in humans [50]. *EGR4* (Early Growth Response Protein 4) is one of the prototypes of a family of zinc-finger transcription factors, which activates the transcription of target genes whose products are required for processes such as mitogenesis and differentiation [49]. *EGR4* is an important TF in neuronal maturation and its expression is induced by cerebral ischemia and inflammation [51, 52]. Interaction of *EGR4* and fatty acids with *EGR1* and the PPAR pathway was associated with cardiovascular risk [53]. Finally, the TF *RUNX1T1* (Runt Related Transcription Factor 1 Translocation Partner 1) is a transcriptional co-repressor that acts as a negative regulator of adipogenesis [49]. A recent study reported that *RUNX1T1* is an inhibitor of brown adipogenesis (associated with a lean and healthy phenotype), which was associated with obesity and suggested that the miRNAs that downregulate this TF could be part of novel therapeutics to increase BAT (brown adipose tissue) in humans [54]. *RUNX1T1* has been also implicated in epigenetic regulation as demonstrated in genome-wide methylation study following prenatal and postnatal dietary omega-3 fatty acid supplementation in pigs, which was differentially methylated between the treatments [55].

Other important candidate regulators were also identified in the hotspot eQTLs regions, such as nuclear receptors (*NRII3*, *NCOR*, *NRBF2*), miRNAs (bta-mir-2294, bta-mir-1584, bta-mir-2322, and bta-mir-584-3) and small nuclear RNAs (*SNORA70e*, *SNORA19* and *SNORA9*).

The nuclear receptor subfamily 1 member 3 (*NRII3*) is an important regulator of xenobiotic, bile

acid, and cholesterol/HDL metabolism, energy homeostasis [56]. While, nuclear receptor binding factor 2 (*NRBF2*) is associated with the mTORC1 activation by lysosomal cholesterol, which is directly dependent of fatty acid content [57].

The transcriptional corepressor *NCOR*, interacts with nuclear receptors and mediates the silencing of retinoid and thyroid receptors [58]. LXRs demonstrate higher affinity for *NCOR* in biochemical assays, which when up-regulated increases the synthesis of long chain FAs (PUFA). These FAs, such as palmitoleic acid, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), are involved in anti-inflammatory activity [59].

Both small nuclear RNAs and miRNAs are non-coding regulatory RNAs, which utilize a similar set of processing enzymes. They are involved in several biological processes such as cell differentiation, cell proliferation, cell death, metabolic control, and transposon silencing [60]. It was interesting to observe that eQTL analysis also identified trans-eQTLs that act as post-transcriptional regulators.

Correlation between the eQTL hotspots and IMF content traits

It is known that DNA polymorphisms can alter complex phenotypic traits by modulating gene expression [4, 5]. Here we showed that genetic variation can influence gene expression, but the question of how alteration of gene expression could influence the phenotype remains unanswered. To further explore this question, we performed co-expression analysis that used the expression profile data between the two most frequently observed genotypes (AA and AB) of the three eQTL hotspot regions. Furthermore, we correlated the constructed modules with the phenotypes. These regions harbored the annotated TFs and are not in LD. WGCNA has been successfully used to investigate how gene expression changes are coordinated across transcripts and how these changes are associated with phenotype [61, 62].

Our results revealed significant differences in the pattern of the network construction (modules) for AA versus AB genotypes, as well as for correlation between network module eigengene (ME) values and phenotypic traits for all the three hotspots. The change in the number of modules present for each genotype may be attributed to the significant difference in co-activation of

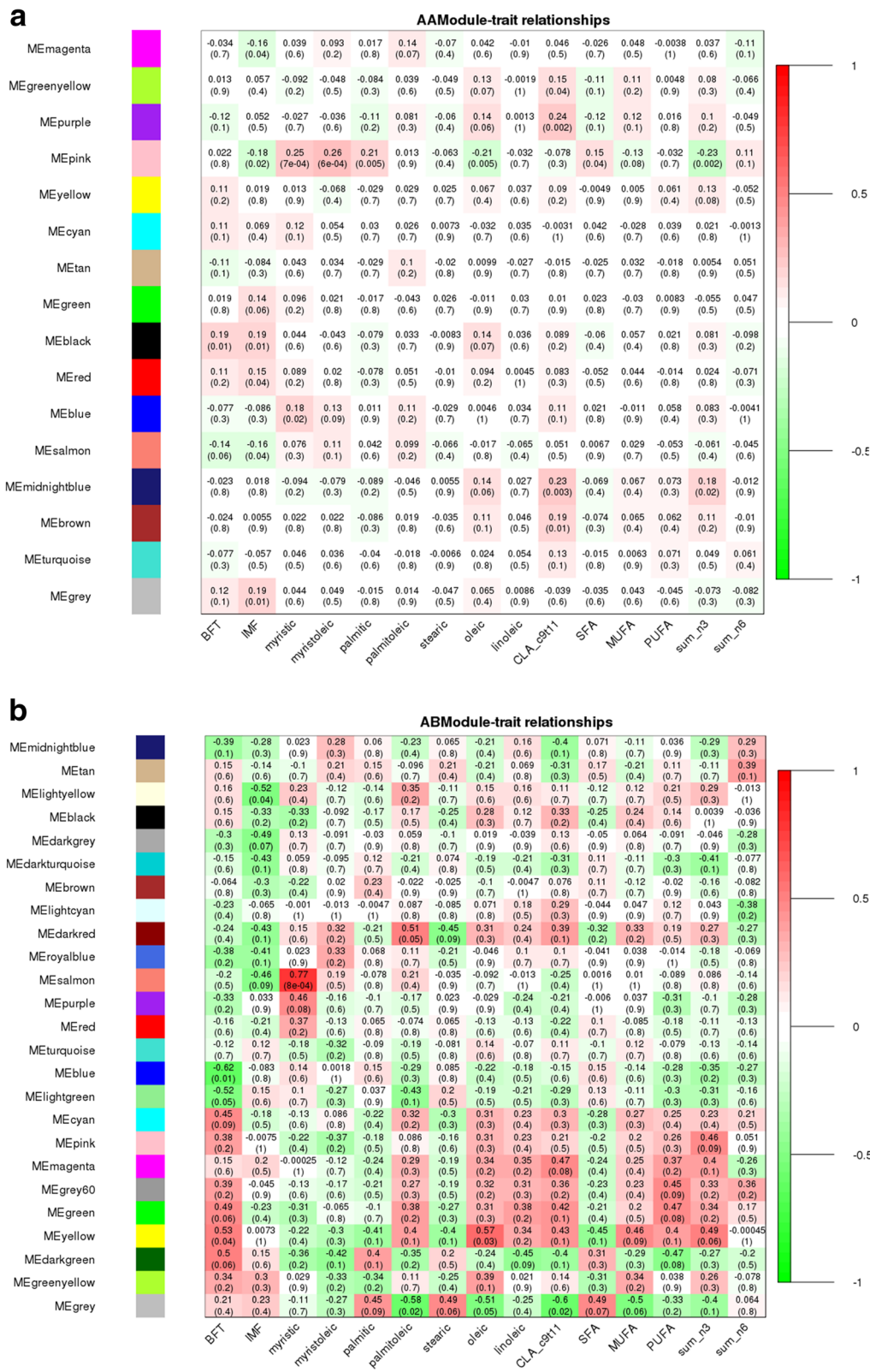


Fig. 7 (See legend on next page.)

(See figure on previous page.)

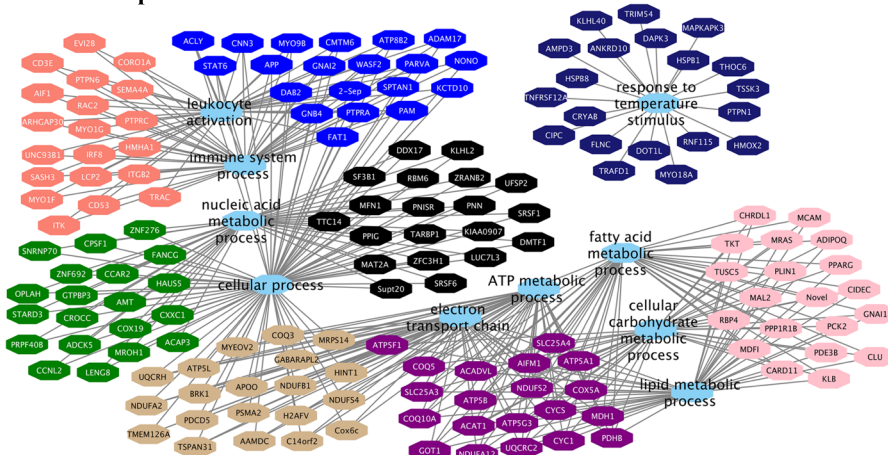
Fig. 7 a Correlation between network module eigengene values and traits. Colors to the left represent the 16 modules in the AA genotype (rs134095631) network. **b** Correlation between network module eigengene (ME) values and traits. Colors to the left represent the 25 modules in the AB genotype (rs134095631) network. For each module, the heatmap shows ME correlations to traits. Numbers in each cell report the correlation coefficients and student asymptotic p-value (parentheses) for significant ME-trait relationships. The scale bar, on the right, indicates the range of correlations from positive (red, 1) to negative (green, -1)

genes expressed in skeletal muscle, potentially due to regulation or mutation in these transcription factors. This would explain the differences in module content and GO enrichment results across genotypes. These findings also support the hypothesis that re-wiring of

gene expression modulates pathways that can influence the traits under study.

Recent studies showed that hub genes (highly connected) tend to play important roles in co-expressed networks (modules) acting as potential regulators [63, 64].

a - AA – Hotspot 1



b - AB – Hotspot 1

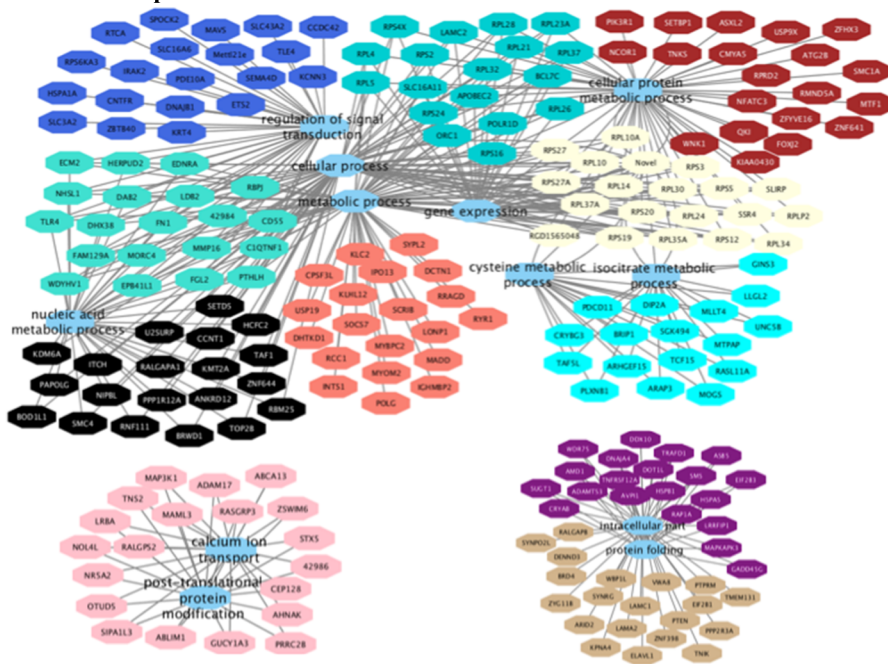


Fig. 8 Co-expression networks show the top 20 hub genes of AA (a) and AB-modules (b) of hotspot 1 correlated (p -value < 0.10) with at least three different traits associated with lipid deposition and composition in skeletal muscle of Nellore steers. Colored octagons represent the hub genes within each module, and blue octagons represent the biological processes associated (FDR 5%) with the genes

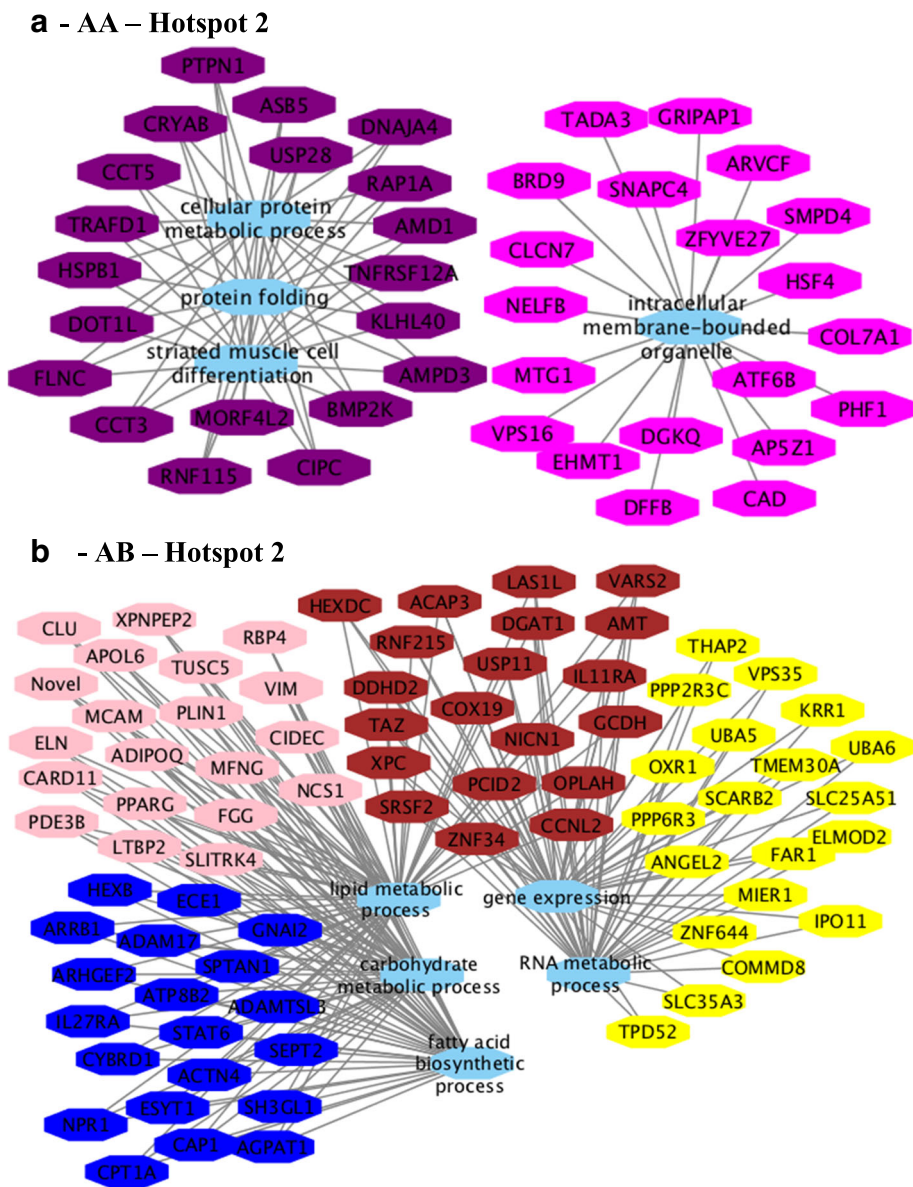
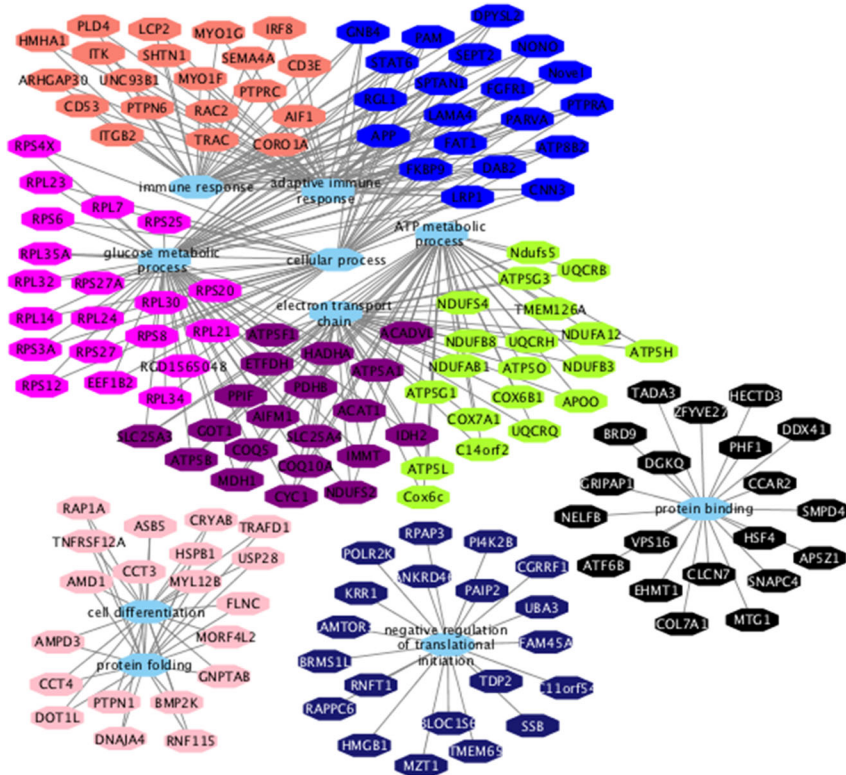


Fig. 9 Co-expression networks show the top 20 hub genes of AA (a) and AB-modules (b) of hotspot 2 correlated (p -value < 0.10) with at least three different traits associated with lipid deposition and composition in skeletal muscle of Nellore steers. Colored octagons represent the hub genes within each module, and blue octagons represent the biological processes associated (FDR 5%) with the genes

Many hub genes and candidate genes for phenotypic variation were identified in this study such as *TMEM39B*, *BCAR1*, *MED19*, *LIN37*, *TRIM54* and *PPAR* (see Additional file 17), were previously identified as differentially expressed genes between groups of animals with extreme values of IMF deposition and composition [17, 18]. The module-trait correlation analysis revealed connected modules and hub genes associated with BP such as immune response, leukocyte proliferation, lipid metabolism, and fatty acid biosynthesis for the different genotypes and hotspots. Some of these hub genes could be indicated as candidate biomarkers because they

probably share similar functions. Second, several hub genes (*RPL4*, *RPL5*, *BCL7C*, *KCNN3*, *DNAJB1*, *COX19*, *UBA6*, *UBA5*) were associated with BP such as gene expression, nucleic acid metabolic, and RNA metabolic processes suggesting they play an important role in gene expression regulation. Based on the genotype specific differences in number of modules, module-trait relationships pattern, co-expression networks constructed, and the functional enrichment between the two genotypes, some hub genes identified in this study are plausible biomarkers for fatty acid variability such as *TMEM39B* and *PPAR*.

a - AA – Hotspot 3



b - AB – Hotspot 3

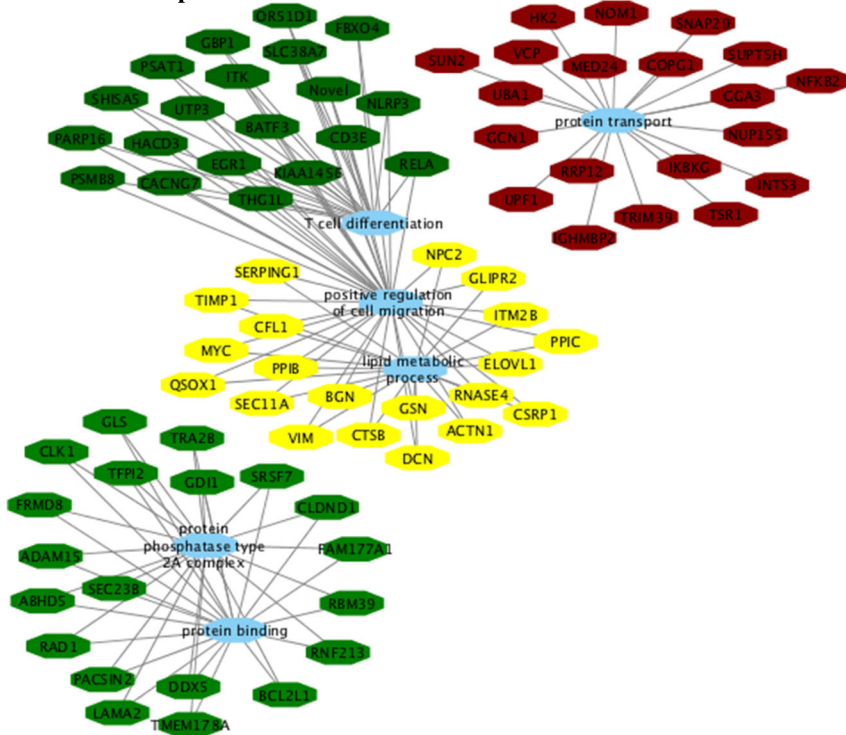


Fig. 10 Co-expression networks show the top 20 hub genes of AA (a) and AB-modules (b) of hotspot 3 correlated (p -value < 0.10) with at least three different traits associated with lipid deposition and composition in skeletal muscle of Nellore steers. Colored octagons represent the hub genes within each module, and blue octagons represent the biological processes associated (FDR 5%) with the genes

Genetic variation in the protein-coding region of a TF could modify the interaction of TF with TFBS. However, in our data we did not find any SNP in the coding regions of *USF1*, *EGR4* and *RUNX1T1* associated with the AA and AB genotypes, but one could not exclude the possibility that the observed effects result from linkage disequilibrium with a SNP not represented in the Illumina bovine chip. Alternatively, genetic variation in the promoter region of a TF could influence the abundance of the TF and thus alter expression of downstream genes. The low expression level of these TFs in our study could explain the lack of difference in TF gene expression between the AA and AB genotypes. Thus, further research is necessary to determine the causative mutation associated with the TF identified.

Conclusion

We identified several regions across the genome that affect gene expression level (expression quantitative trait loci, eQTL) and overlap with QTL regions associated with the deposition and composition of IMF. Some of these regions harbor TF and control the expression of several genes (hotspots). Results obtained supported the hypothesis that eQTL analysis can be used to identify putative regulatory regions and transcription factors associated with important phenotypic traits that are controlled by modulation of gene expression profile.

Methods

Animals, phenotype and genotype data

The animals ($n = 192$), phenotype and genotype data used in this study was comprised of Nellore steers sired by 34 unrelated sires, selected to represent the main genealogies used in Brazil according to the National Summary of Nellore produced by the Brazilian Association of Zebu Breeders (ABCZ) and National Research Center for Beef Cattle to ensure compliance with international guidelines for animal welfare as described previously by Cesar et al. [15]. A captive bolt pistol was used for stunning the animals prior to slaughter. SNPs with call rate $\leq 95\%$, minor allele frequency (MAF) $\leq 5\%$, those located on sex chromosomes or not mapped in the *Bos taurus* UMD 3.1 assembly were removed. The MAF threshold was chosen based on the sample size in order to minimize the number of false-positive and false-negative associations [65]. After filtering, a total of 461,643 SNP was utilized in eQTL mapping.

RNA extraction and sequencing

Total RNA was extracted from 100 mg of frozen LD muscle from 192 animals that were collected at slaughter using the TRIzol reagent (Life Technologies, Carlsbad, CA). RNA integrity was verified by Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Only samples

with RIN > 8 were used. A total of 2 μg of total RNA from each sample was used for library preparation according to the protocol described in the TruSeq RNA Sample Preparation kit v2 guide (Illumina, San Diego, CA). Average library sizes were estimated using the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and quantified using quantitative PCR with the KAPA Library Quantification kit (KAPA Biosystems, Foster City, CA, USA). Quantified samples were diluted and pooled (three pools of six samples each). Three lanes of a sequencing flowcell, using the TruSeq PE Cluster kit v3-cBot-HS kit (Illumina, San Diego, CA, USA), were clustered and sequenced using HiScanSQ equipment (Illumina, San Diego, CA, USA) with a TruSeq SBS Kit v3-HS (200 cycles), according to manufacturer's instructions. Sequencing analyses were performed at the Genomics Center at ESALQ, Piracicaba, São Paulo, Brazil.

Sequencing adaptors and low-complexity reads were removed in an initial data-filtering step. Quality control and read statistics were estimated with FASTQC version 0.10.1 software [<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>]. RNA-Seq by Expectation Maximization (RSEM) approach was performed to estimate the number of fragments originating from each gene in each replicate library, which is capable of handling reads that map ambiguously between isoforms and genes, and minimize the differences in total read counts across samples (normalization procedure) [66]. The UMD3.1 *Bos taurus* assembly available at Ensembl [http://www.ensembl.org/Bos_taurus/Info/Index/] was used as reference genome.

Identification of eQTL and hotspot regions

The Matrix eQTL R package [67] was used to identify associations between genetic variation from genotype (SNPchip) and gene expression (RNA-Seq) [68]. Contemporary group (including farm, year and slaughter date) and lane were included in the model as fixed effects and age as a covariate. Markers associated with variation in gene expression that were within 1 Mb of the gene were defined as cis-eQTLs (local variants), while markers more than 1 Mb from the gene were defined as trans-eQTLs (distant variants). Matrix eQTL tests the association between each marker (SNP) and each gene assuming the effect of genotype as additive, performs a separate test for each pair (marker and gene) and corrects for multiple tests by calculating false discovery rate (FDR) [69]. The estimated effect size (slope coefficient) and the genetic variance explained by the markers was also provided according the Matrix eQTL package [66]. eQTL hotspots (markers that affect the gene expression level for many genes) were identified by permuting the distribution of eQTLs across the genome after 1000 permutations. A hotspot threshold was

identified that corresponded to the 95th percentile of the value. Linkage disequilibrium (LD) analysis and visualization by PLINK v.1.07 [70] and Haploview [71], respectively, were used to check if the hotspots were in LD with each other and if so to select just one as the eQTL hotspot. The hotspot region was defined as a 4 Mb window around the hotspot eQTL, i.e. 2 Mb extended to each side of the hotspot eQTL.

Association test between eQTL hotspots and the phenotypes

The analysis of variance (ANOVA) model containing contemporary group (including farm, year and slaughter date) and lane as fixed effects and age as a covariate and was applied to test for association between a given SNP and a corresponding phenotype. Evidence of population stratification was not identified in this population based on previous results reported by our group [72]. Therefore, it was not included in the model to detect eQTL. The statistical test was performed by R software and applied to verify the effect of the eQTL hotspots identified in this study on the phenotypes of interest. The correction for multiple tests was applied by calculating false discovery rate (FDR 5%).

Overlap statistics (eQTLs / QTLdb)

Overlap analysis was carried out using the Bioconductor package *regioneR* [24]. The package implements a general framework for testing overlaps of genomic regions based on permutation sampling. We repeatedly sampled random regions ($N = 1000$ permutations) from the UMD_3.1 genome assembly matching size and chromosomal distribution of the detected eQTLs. This test was performed for QTL class, QTL associated with traits of production and quality of carcass and meat, and QTL previously reported by our group [15] associated with the traits of interest in this study. In every permutation, the overlap with the cattle QTLdb [23] was recomputed based on the total genomic size in Mb that was overlapped.

Annotation and functional annotation of the eQTLs

The eQTL annotations were performed using Ensembl Variant Effect Predictor, a free toolset for the analysis, annotation, and prioritization of genomic variants in coding and non-coding regions [73]. The reference genome assembly used was UMD3.1 *Bos Taurus* from Ensembl data bank [73]. With this set of tools, the location of an eQTL in relation to a gene can be defined as outside of the gene, in the coding sequence, or in untranslated regions (UTR). The functional impact was determined for those eQTLs that were localized in the coding sequences. Functional enrichment analyses were performed with Protein ANalysis THrough Evolutionary Relationships (PANTHER) [26] using the list of the

genes harbored in 4 Mb eQTL regions (hotspot, cis and trans). The statistical over-representation test by PANTHER was used to obtain the gene ontology association (biological processes and protein classes) from a given list of genes. That test was performed to compare a list of reference genes (background, all genes expressed in skeletal muscle identified in this study) to a list of genes harbored within 4 Mb eQTL regions, and determine if a particular class of gene ontology (GO) biological processes were over-represented or under-represented (nominal p -value ≤ 0.05).

Transcription factor binding site searching

Annotated transcription factors (TFs) by JASPAR CORE database [42] were searched within the eQTL hotspot region, and the transcription factor binding sites (TFBSs) of these TFs were searched using LASAGNA-Search 2.0 [28]. LASAGNA-Search 2.0 is an integrated web tool based on the algorithm Length-Aware Site Alignment Guided by Nucleotide Association, which allows the identification of TFBS from a list of target genes. To perform the LASAGNA-Search 2.0 program, the TFBSs and position-specific scoring matrix (PSSM) were collected from JASPAR CORE database; the name of the TFs were chosen based on the *Bos taurus* genome annotation; and the list of target genes was the list of gene affected by the eQTL hotspots identified herein. This method used by LASAGNA 2.0 can distinguish true binding sites from other non-functional sites with similar sequences by giving a weighted match to any given substring (combinations) of fixed length. The TFBS were searched in 1500 bp of length of promoter region obtained from Biomart tool of Ensembl website [<http://www.ensembl.org/biomart>] of those genes that were affected by eQTL hotspot for the specific TF.

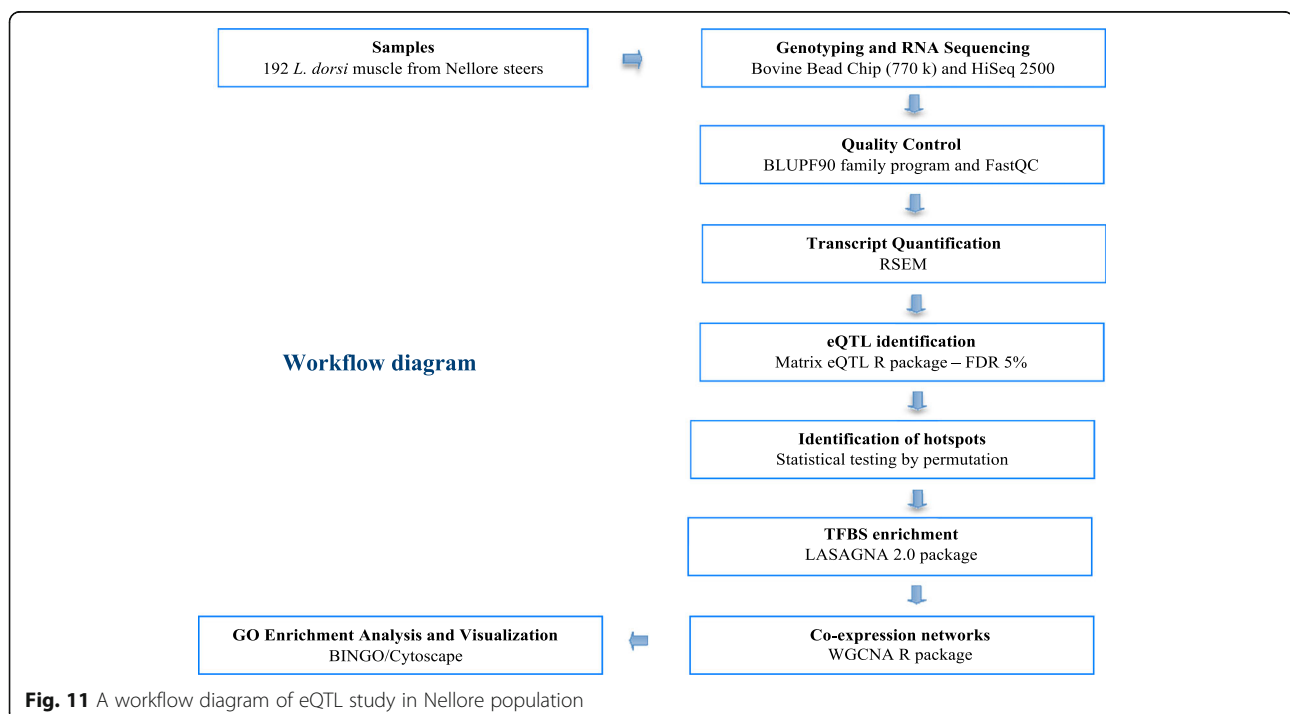
Association between eQTL hotspots and traits by co-expression network analysis

Hotspot eQTLs were chosen that had annotated TFs within them to associate the hotspot eQTLs with the traits of interest (IMF deposition and composition). WGCNA (Weighted Gene Correlation Network Analysis), which is a systems biology network method that describes the correlation patterns among all expressed genes across samples, was performed by WGCNA R package [74]. This approach was used to identify the differences in co-expression networks between hotspot eQTL genotypes (AA and AB, MAF > 0.05). Gene networks were constructed separately for each of the two most frequent genotypes, which were assigned an arbitrary color. For WGCNA analysis, the correlation matrix was built using the absolute value of the Pearson correlation coefficient between all gene pairs across all samples. The Pearson correlation matrix was subsequently

transformed into an adjacency matrix (A) using a power function based on scale-free topology criterion, as described [75]. A soft threshold power of 6 with scale free fitting index of 0.9 was applied to calculate the adjacency matrix. TOM-based dissimilarity (1-TOM) was used for module identification using Dynamic Tree Cut algorithm with cutreeDynamic function in WGCNA package [74] and defining the deep split = 2 and minimum module size = 30. To make the genotype networks comparable, for each eQTL hotspot, we scaled the TOM (Topology Overlap Matrix) connectivity's in genotype with the minimum number of data such that its 95th percentile equals the 95th percentile of the genotype with maximum number of data, as described by Langfelder and Horvath [76]. To quantify co-expression similarity of entire modules, we calculated their eigengene values using moduleEigengenes function in WGCNA package and clustered them based on their correlation using height cut of 0.25, which corresponded to a correlation of 0.75, to merge similar modules [76]. All other WGCNA parameters remained at their default settings. Grey-colored modules contained all genes that were not part of any module. The associations between individual genes with traits of interest (fat deposition and composition of intramuscular fat, adjusted phenotype as described by Cesar et al. [18] was quantified by the Gene Significance (GS) approach, which was defined as the correlation (the absolute value) between the gene and the trait of interest. The quantitative measure of module membership (MM) was

defined as the correlation of the module eigengene and the gene expression profile. With these assumptions, the similarity of all genes was quantified to every module. The p -value threshold applied in this correlation analysis was p -value < 0.10 based on previous studies that used the same approach [19, 77].

Gene Ontology (GO) annotation from a list of genes within of each module significantly correlated (p -value < 0.10) with at least three different traits by Cytoscape plugin BINGO [25] using the latest update of gene ontology annotation database (GOA) [78]. The statistical significance of GO term enrichment was measured by a hypergeometric test using the genes in the whole network as the background (all genes expressed in skeletal muscle). The Benjamini and Hochberg [69] correction (p-adjusted) was used to correct for multiple testing. Only GO terms that were significantly over-represented (p-adjusted ≤ 0.05) were reported. This functional enrichment analysis was followed by network construction using the hub genes to support the hypothesis that the expression pattern of the modules correlated to the phenotypes can influence the IMF content traits. The construction and visualization of the networks for each eQTL hotspot genotype were performed by Cytoscape 3.5.1 [30] connecting the top 20 hub genes [most connected genes, higher values of the module membership (MM)] of each module by the common significant (FDR 5%) BP. A workflow diagram of this study is shown in Fig. 11.



Additional files

Additional file 1: Descriptive statistics for backfat thickness (BFT) and IMF deposition and composition traits in Nellore steers. (DOCX 61 kb)

Additional file 2: QQ-plot of the distribution of *p*-values for cis-eQTLs (red line) and trans-eQTL (blue line) using mRNA sequencing and animal genotype data. The top horizontal grey line denotes a 5% false discovery rate significance threshold for trans-eQTLs and the bottom one for cis-eQTLs. (DOCX 221 kb)

Additional file 3: All information about the cis and trans-eQTLs identified from RNA-Seq and genotyping data of Nellore steers population. (XLSX 2154 kb)

Additional file 4: Variant annotation by VEP tool for cis-eQTLs (A) and trans-eQTLs (B) identified across the whole Nellore genome. (DOCX 244 kb)

Additional file 5: Variant effect annotation of cis-eQTLs by VEP tool. (XLSX 212 kb)

Additional file 6: Variant effect annotation of trans-eQTLs by VEP tool. (XLSX 262 kb)

Additional file 7: A. Shown are the number of overlaps (y-axis) between the detected eQTLs (set A) and known QTLs from Cattle QTLdb (set B) in four overlap categories: AinB denotes an eQTL encompassed by / completely contained in a QTL (and vice versa for BinA). On the other hand, AleftB and ArightB denote partial overlap of an eQTL and with the left or the right part of a QTL, respectively. B. QTLs classes that overlap with eQTLs regions from Cattle QTLdb. Red bars correspond to the mean overlap size (in Mb) of the eQTL regions that was observed for eQTL regions for respective trait (y-axis), while the cyan bars indicate the mean overlap size (in Mb) estimated after 1000x random resamplings (x-axis). The error bars indicates the standard deviation, while permutation *p*-values are listed on the right. C. QTLs associated with beef production, carcass and beef quality that overlap with eQTLs regions. Red bars correspond to the mean overlap size (in Mb) of the eQTL regions that was observed for eQTL regions for respective trait (y-axis), while the cyan bars indicate the mean overlap size (in Mb) estimated after 1000x random resamplings (x-axis). The error bars indicates the standard deviation, while permutation *p*-values are listed on the right. (DOCX 171 kb)

Additional file 8: eQTL regions that overlap with QTLdb regions identified from the same dataset. (XLSX 43 kb)

Additional file 9: Haploview visualization of linkage disequilibrium (LD) around two eQTL hotspots. The hotspots identified on chromosome 11 (A) and chromosome 28 (B) are marked in red, and the *D'* values estimated between the hotspots. *D'* value represent the percentage of the time that the both markers are co-inherited. *D'* prime values of 1.0 are not shown (the box is empty). The intensity of red color indicates the *D'* values estimated between the hotspot. (DOCX 954 kb)

Additional file 10: R-squared values of linkage disequilibrium between each pair of hotspot eQTLs. (XLSX 54 kb)

Additional file 11: Association test between hotspot eQTLs and IMF content-traits. (XLSX 57 kb)

Additional file 12: Biological processes identified using the list of genes harbored within 2 Mb cis (A), trans (B) and hotspot (C) eQTL regions as the input gene list file. The values shown in the pie chart are the percentage of genes classified to each GO term over the total number of genes in the list used. (DOCX 599 kb)

Additional file 13: Top ten transcription factor binding sites (TFBSs) over-represented in promoter sequence of genes that were associated with *EGR4*. (XLSX 49 kb)

Additional file 14: Top ten transcription factor binding sites (TFBSs) over-represented in promoter sequence of genes that were associated with *RUNX1T1*. (XLSX 54 kb)

Additional file 15: Top ten transcription factor binding sites (TFBSs) over-represented in promoter sequence of genes that were associated with *USF1*. (XLSX 91 kb)

Additional file 16: Over-represented TFBS motifs of: (A) *EGR4* (*NGFI-C* aliases), (B) *RUNX1* and (C) *USF1* in genes into hotspot eQTL regions. (DOCX 240 kb)

Additional file 17: Co-expression network results and functional enrichment. (XLSX 21960 kb)

Abbreviations

BFT: Backfat thickness; CLA: Conjugated linoleic acid; CLA-c9t11: Conjugated linoleic acid cis 9, trans 11; eQTL: Expression quantitative trait loci; FDR: False discovery rate; GO: Gene Ontology; LASAGNA: (Length-Aware Site Alignment Guided by Nucleotide Association); LD: Linkage disequilibrium; MUFA: Monounsaturated fatty acid; NGS: Next generation sequencing; PUFA: Polyunsaturated fatty acid; QTL: Quantitative trait loci; RNA-Seq: RNA sequencing; SFA: Saturated fatty acid; TF: Transcription factor; UTR: Untranslated region; WGCNA: Weighted gene correlation network analysis

Acknowledgements

This study was conducted with funding from EMBRAPA (Macroprograma 1, 01/2005) and FAPESP (processes number 2014/11871-5, 2014/22884-0 and 2012/23638-8). LCAR and LLC were granted CNPq fellowships. The authors would like to acknowledge the collaborative efforts among EMBRAPA, University of São Paulo, and Iowa State University.

Funding

This study was conducted with funding from EMBRAPA (Macroprograma 1, 01/2005) and FAPESP (processes number 2014/11871-5, 2014/22884-0 and 2012/23638-8). LCAR and LLC were granted CNPq fellowships.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the European Nucleotide Archive (ENA) repository (EMBL-EBI), under accession PRJEB13188, PRJEB10898, and PRJEB19421 [<https://www.ebi.ac.uk/ena/submit/sra>].

Authors' contributions

ASMC, LCAR, JMR and LLC conceived the idea of this research. ASMC and MDP participated in the experimental design and RNA-sequencing. ASMC, LK, DG, LG, HB, JEK, MAM, GBO, and EFW performed data analysis. ASMC drafted the manuscript. ASMC, LLC, MDP, JEK, JMR, GCMM, PCT, PSNO collaborated with interpretation, discussion and writing of the manuscript. LCAR, JMR, GBM, AZN, and LLC provided the experimental environment, phenotype and data analysis support. All authors have read and approved the final manuscript.

Ethics approval

All experimental procedures involving steers were approved by the Institutional Animal Care and Use Committee Guidelines from Brazilian Agricultural Research Corporation – EMBRAPA (process number: Macroprograma 1, 01/2005) and sanctioned by the president Dr. Rui Machado to ensure compliance with international guidelines for animal welfare.

Consent for publication

Not applicable.

Competing interests

Dr. James Reecy is a member of the editorial board (Associate Editor) of BMC Genetics journal.

Publisher's Note

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

Author details

¹Department of Animal Science, University of São Paulo, Piracicaba, SP 13418-900, Brazil. ²Department of Animal Science, Iowa State University, Ames, IA 50011, USA. ³Embrapa Pecuária Sudeste, São Carlos, SP 13560-970, Brazil. ⁴Embrapa Informática Agropecuária, Campinas, SP 13083-886, Brazil. ⁵School of Agriculture, Massey University, Ruakura, Hamilton, New Zealand.

Received: 16 October 2017 Accepted: 14 June 2018

Published online: 27 June 2018

References

1. Michaelson JJ, Loguercio S, Beyer A. Detection and interpretation of expression quantitative trait loci (eQTL). *Methods*. 2009;48:265–76. United States

2. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 2010;6(4):e1000888.
3. Innocenti F, Cooper GM, Stanaway IB, Gamazon ER, Smith JD, Mirkov S, Ramirez J, Liu W, Lin YS, Moloney C, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* 2011;7(5):e1002078.
4. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet.* 2013;14:125–38. England
5. Lowe WL Jr, Reddy TE. Genomic approaches for understanding the genetics of complex disease. *Genome Res.* 2015;25(10):1432–41.
6. Addison O, Marcus RL, Lastayo PC, Ryan AS. Intermuscular fat: a review of the consequences and causes. *Int J Endocrinol.* 2014;2014:309570.
7. Jung UJ, Choi MS. Obesity and its metabolic complications: the role of adipokines and the relationship between obesity, inflammation, insulin resistance, dyslipidemia and nonalcoholic fatty liver disease. *Int J Mol Sci.* 2014;15(4):6184–223.
8. Hausman GJ, Basu U, Du M, Fernyhough-Culver M, Dodson MV. Intermuscular and intramuscular adipose tissues: bad vs. good adipose tissues. *Adipocyte.* 2014;3(4):242–55.
9. Listrat A, Lebreton B, Louveau I, Astruc T, Bonnet M, Lefaucheur L, Picard B, Bugeon J. How muscle structure and composition influence meat and flesh quality. *ScientificWorldJournal.* 2016;2016:3182746.
10. Marin-Garcia J, Goldenthal MJ. Fatty acid metabolism in cardiac failure: biochemical, genetic and cellular analysis. *Cardiovasc Res.* 2002;54(5):516–27. England
11. Nakamura MT, Nara TY. Essential fatty acid synthesis and its regulation in mammals. *Prostaglandins Leukot Essent Fatty Acids.* 2003;68(2):145–50.
12. Furuhashi M, Hotamisligil GS. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nat Rev Drug Discov.* 2008;7(6):489–503.
13. Uemoto Y, Abe T, Tameoka N, Hasebe H, Inoue K, Nakajima H, Shoji N, Kobayashi M, Kobayashi E. Whole-genome association study for fatty acid composition of oleic acid in Japanese Black cattle. *Anim Genet.* 2010;42(2):141–8.
14. Ishii A, Yamaji K, Uemoto Y, Sasago N, Kobayashi E, Kobayashi N, Matsuhashi T, Maruyama S, Matsumoto H, Sasazaki S, Mannen H. Genome-wide association study for fatty acid composition in Japanese Black cattle. *Anim Sci J.* 2013;84(10):675–82.
15. Cesar AS, Regitano LC, Mourão GB, Tullio RR, Lanna DP, Nassu RT, Mudado MA, Oliveira PS, do Nascimento ML, Chaves AS, et al. Genome-wide association study for intramuscular fat deposition and composition in Nelore cattle. *BMC Genet.* 2014;15:39.
16. Tizioto PC, Decker JE, Taylor JP, Schnabel RD, Mudadu MA, Silva FL, Mourão GB, Coutinho LL, Tholon P, Sonstegard TS, Rosa AN, Alencar MM, Tullio RR, Medeiros SR, Nassu RT, Feijó GLD, Silva LOC, Torres RA, Siqueira F, Higa RH, Regitano LCA. A genome scan for meat quality traits in Nelore beef cattle. *Physiol Genomics.* 2013;45(21):1012–20.
17. Cesar AS, Regitano LC, Koltjes JE, Fritz-Waters ER, Lanna DP, Gasparin G, Mourão GB, Oliveira PS, Reecy JM, Coutinho LL. Putative regulatory factors associated with intramuscular fat content. *PLoS One.* 2015;10(6):e0128350.
18. Cesar AS, Regitano LC, Poleti MD, Andrade SC, Tizioto PC, Oliveira PS, Felício AM, do Nascimento ML, Chaves AS, Lanna DP, et al. Differences in the skeletal muscle transcriptome profile associated with extreme values of fatty acids content. *BMC Genomics.* 2016;17(1):961.
19. Sheng X, Ni H, Liu Y, Li J, Zhang L, Guo Y. RNA-seq analysis of bovine intramuscular, subcutaneous and perirenal adipose tissues. *Mol Biol Rep.* 2014 Mar;41(3):1631–7.
20. Fortes MR, Snelling WM, Reverter A, Nagaraj SH, Lehnert SA, Hawken RJ, DeAtley KL, Peters SO, Silver GA, Rincon G, et al. Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *J Anim Sci.* 2012 Sep;90(9):2894–906.
21. Saatchi M, Garrick DJ, Tait RG Jr, Mayes MS, Drewnoski M, Schoonmaker J, Diaz C, Beitz DC, Reecy JM. Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. *BMC Genomics.* 2013;14:730.
22. Lemos MV, Chiaia HL, Berton JP, Feitosa FL, Aboujaoud C, Camargo GM, Pereira AS, Albuquerque LG, Ferrinho AM, Mueller LF, et al. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nelore cattle using the single step procedure. *BMC Genomics.* 2016;17:213.
23. Hu ZL, Park CA, Reecy JM. Developmental progress and current status of the animal QTLdb. *Nucleic Acids Res.* 2016;44(D1):D827–33.
24. Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics.* 2016;32(2):289–91.
25. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015;16:197–212. England
26. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 2016;44(D1):D336–42.
27. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet.* 2012;13:613–26. England
28. Lee C, Huang CH. LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser. *Bioinformatics.* 2014;30:1923–5. England: The Author 2014. Published by Oxford University Press For Permissions, please e-mail: journals.permissions@oup.com
29. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics.* 2005;21:3448–9. England
30. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campillo I, Creech M, Gross B, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc.* 2007;2(10):2366–82.
31. Steige KA, Reimegard J, Koenig D, Scofield DG, Slotte T. Cis-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Mol Biol Evol.* 2015;32(10):2501–14.
32. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Coin L, de Silva R, Cookson MR, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci.* 2014;17(10):1418–28.
33. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci.* 2012;69(21):3613–34.
34. Mattick JS. Introns: evolution and function. *Curr Opin Genet Dev.* 1994;4(6):823–31.
35. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AJ, Haley CS, Buys N, Tally M, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature.* 2003;425:832–6. England
36. Piraino SW, Furney SJ. Identification of coding and non-coding mutational hotspots in cancer genomes. *BMC Genomics.* 2017;18(1):17.
37. Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, Munson PJ, Hill DE, Vidal M, Levy D. Dynamic role of trans regulation of gene expression in relation to complex traits. *Am J Hum Genet.* 2017 Jun 1;100(6):985–6.
38. Gibson S, Powell JE, Marigorta UM. Expression quantitative trait locus analysis for translational medicine. *Genome Med.* 2015;7(1):60.
39. Hasin-Brumshtein Y, Khan AH, Hormozdiari F, Pan C, Parks BW, Petyuk VA, Piehowski PD, Brümmer A, Pellegrini M, Xiao X, Eskin E, Smith RD, Lusis AJ, Smith DJ. Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. *Elife.* 2016;5:e15614. <https://doi.org/10.7554/eLife.15614>.
40. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics.* 2005;171(4):1813–22.
41. Schaeffe B, Emerson JJ, Wang TY, Lu MY, Hsieh LC, Li WH. Inheritance of gene expression level and selective constraints on trans- and cis-regulatory changes in yeast. *Mol Biol Evol.* 2013;30:2121–33. United States
42. Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res.* 2014;24(1):84–95.
43. Guerrero RF, Posto AL, Moyle LC, Hahn MW. Genome-wide patterns of regulatory divergence revealed by introgression lines. *Evolution.* 2016;70(3):696–706.
44. Joo JW, Sul JH, Han B, Ye C, Eskin E. Effectively identifying regulatory hotspots while capturing expression heterogeneity in gene expression studies. *Genome Biol.* 2014;15(4):r61.
45. Tian J, Keller MP, Oler AT, Rabaglia ME, Schueler KL, Stapleton DS, Broman AT, Zhao W, Kendziorski C, Yandell BS, et al. Identification of the bile acid transporter *Slco1a6* as a candidate gene that broadly affects gene expression in mouse pancreatic islets. *Genetics.* 2015;201(3):1253–62.
46. Lee E, Bussemaker HJ. Identifying the genetic determinants of transcription factor activity. *Mol Syst Biol.* 2010;6:412.

47. Mathelier A, Fornes O, Arenillas DJ, Chen C, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016;44(Database issue):D110–5.
48. Jiang G, Chakraborty A, Wang Z, Boustani M, Liu Y, Skaar T, Li L. New aQTL SNPs for the CYP2D6 identified by a novel mediation analysis of genome-wide SNP arrays, gene expression arrays, and CYP2D6 activity. *Biomed Res Int.* 2013;2013:493019.
49. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics.* 2016;54(1):1.30.1–1.30.33.
50. Laurila PP, Soronen J, Kooijman S, Forsstrom S, Boon MR, Surakka I, Kaiharju E, Coomans CP, Van Den Berg SA, Autio A, et al. USF1 deficiency activates brown adipose tissue and improves cardiometabolic health. *Sci Transl Med.* 2016;8:323ra313. United States: American Association for the Advancement of Science
51. Decker EL, Nehmann N, Kampen E, Eibel H, Zipfel PF, Skerka C. Early growth response proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. *Nucleic Acids Res.* 2003;31(3):911–21.
52. Mengozzi M, Cervellini I, Villa P, Erbayraktar Z, Gokmen N, Yilmaz O, Erbayraktar S, Manohasandra M, Van Hummelen P, Vandenabeele P, et al. Erythropoietin-induced changes in brain gene expression reveal induction of synaptic plasticity genes in experimental stroke. *Proc Natl Acad Sci U S A.* 2012;109(24):9617–22.
53. Fruchart JC. Peroxisome proliferator-activated receptor-alpha (PPARalpha): at the crossroads of obesity, diabetes and cardiovascular disease. *Atherosclerosis.* 2009;205:1–8. Ireland
54. Trajkovski M, Lodish H. MicroRNA networks regulate development of brown adipocytes. *Trends Endocrinol Metab.* 2013;24(9):442–50.
55. Boddicker RL, Koltes JE, Fritz-Waters ER, Koesterke L, Weeks N, Yin T, Mani V, Nettleton D, Reecy JM, Baumgard LH, et al. Genome-wide methylation profile following prenatal and postnatal dietary omega-3 fatty acid supplementation in pigs. *Anim Genet.* 2016;47(6):658–71.
56. Konno Y, Negishi M, Kodama S. The roles of nuclear receptors CAR and PXR in hepatic energy metabolism. *Drug Metab Pharmacokinet.* 2008;23:8–13.
57. Castellano BM, Thelen AM, Moldavski O, Feltes M, van der Welle RE, Mydock-McGrane L, Jiang X, van Eijkeren RJ, Davis OB, Louie SM, et al. Lysosomal cholesterol activates mTORC1 via an SLC38A9-Niemann-pick C1 signaling complex. *Science.* 2017;355(6331):1306–11.
58. Yoon HG, Wong J. The corepressor silencing mediator of retinoic and thyroid hormone receptor and nuclear receptor corepressor are involved in agonist- and antagonist-regulated transcription by androgen receptor. *Mol Endocrinol.* 2006 May;20(5):1048–60.
59. Yan Y, Jiang W, Spinetti T, Tardivel A, Castillo R, Bourquin C, Guarda G, Tian Z, Tschopp J, Zhou R. Omega-3 fatty acids prevent inflammation and metabolic disorder through inhibition of NLRP3 inflammasome activation. *Immunity.* 2013;38(6):1154–63.
60. Kim T, Reitmaier A. Non-coding RNAs: functional aspects and diagnostic utility in oncology. *Int J Mol Sci.* 2013;14(3):4934–68.
61. Jiang J, Sun X, Wu W, Li L, Wu H, Zhang L, Yu G, Li Y. Construction and application of a co-expression network in *Mycobacterium tuberculosis*. *Sci Rep.* 2016;6:28422.
62. Mina E, van Roon-Mom W, Hettne K, van Zwet E, Goeman J, Neri C, ActH P, Mons B, Roos M. Common disease signatures from gene expression analysis in Huntington's disease human blood and brain. *Orphanet J Rare Dis.* 2016;11(1):97.
63. Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, Aguilar-Salinas C, Pajukanta P. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* 2009;5(9):e1000642.
64. Ponsuksili S, Siengdee P, Du Y, Trakooljul N, Murani E, Schwerin M, Wimmers K. Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. *PLoS One.* 2015;10(4):e0123678.
65. Anderson CA, Petterson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5(9):1564–73.
66. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
67. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28(10):1353–8.
68. Xia W, Zhu XW, Mo XB, Wu LF, Wu J, Guo YF, Zeng KQ, Wang MJ, Lin X, Qiu YH, et al. Integrative multi-omics analysis revealed SNP-lncRNA-mRNA (SLM) networks in human peripheral blood mononuclear cells. *Hum Genet.* 2017; 136:451–62. Germany
69. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B (Methodological).* 1995;57:289–300.
70. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–75.
71. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc.* 2009;2009:pdb ip71. United States
72. Mudadu MA, Porto-Neto LR, Mokry FB, Tizioto PC, Oliveira PS, Tullio RR, Nassu RT, Niciura SC, Tholon P, Alencar MM, et al. Genomic structure and marker-derived gene networks for growth and meat quality traits of Brazilian Nelore beef cattle. *BMC Genomics.* 2016;17:235.
73. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, Garcia Girón C, Hourlier T, Howe K, Kähäri A, Kokocinski F, Martin FJ, Murphy DN, Nag R, Ruffier M, Schuster M, Tang YA, Vogel JH, White S, Zadissa A, Flicek P, Searle SM. The Ensembl gene annotation system. *Database (Oxford).* 2016;2016. <https://doi.org/10.1093/database/baw093>
74. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
75. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:Article17.
76. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54.
77. Vallejo R, Tilley DM, Cedeño DL, Kelley CA, DeMaegd M, Benyamin R. Genomics of the effect of spinal cord stimulation on an animal model of neuropathic pain. *Neuromodulation.* 2016 Aug;19(6):576–86.
78. Huntley RP, Sawford T, Mutowo-Muullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Res.* 2015;43(Database issue):D1057–63.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

