Systems biology reveals NR2F6 and TGFB1 as key regulators of feed efficiency in beef cattle

Pâmela A. Alexandre[1,2]; Marina Naval-Sanchez[2]; Laercio R. Porto-Neto[2]; José Bento S. Ferraz[1]; Antonio Reverter[2] and Heidge Fukumasu[1§]

[1]Department of Veterinary Medicine, College of Animal Sciences and Food Engineering, University of São Paulo, Av. Duque de Caxias Norte 225, Pirassununga, 13635-900, Brazil

[2]Commonwealth Scientific and Industrial Research Organisation (CSIRO), Agriculture & Food, 306 Carmody Road, St. Lucia, 4067, QLD, Australia

[§]corresponding author

Heidge Fukumasu, DVM, PhD, Associate Professor

Phone: 55-19-35656864

Fax: 55-19-3565

e-mail: fukumasu@usp.br

Abstract

Systems biology approaches are used as strategy to uncover tissue-specific perturbations and regulatory genes related to complex phenotypes. We applied this approach to study feed efficiency (FE) in beef cattle, an important trait both economically and environmentally. Poly-A selected RNA of five tissues (adrenal gland, hypothalamus, liver, skeletal muscle and pituitary) of eighteen young bulls, selected for high and low FE, were sequenced (100bp, pared-end). From the 17,354 expressed genes, 1,317 were prioritized by five selection categories (differentially expressed, harbouring SNPs associated with FE, tissue-specific, secreted in plasma and key regulators) and used for network construction. NR2F6 and TGFB were identified and validated by motif discovery as key regulators of hepatic inflammatory response and muscle tissue development, respectively, two biological processes demonstrated to be associated to FE. Moreover, we indicated potential biomarkers of FE which are related to hormonal control of metabolism and sexual maturity. By using robust methodologies and validation strategies, we confirmed main biological processes related to FE in *Bos indicus* and indicated candidate genes as regulators or biomarkers of superior animals.

Keywords: *Bos indicus*/co-expression/motif discovery/inflammation/residual feed intake

Introduction

Since the domestication of the first species, animal selection aims to meet human needs and their changes over time. The current main selection goals in livestock production are increase of productivity, reduction of the environmental impact and reduction of competition for grains with human nutrition (Hayes *et al*, 2013). Thus, feed efficiency (FE) has become a relevant trait of study, as animals considered of high feed efficiency are those presenting reduced feed intake and lower production of methane and manure without compromising animal's weight gain (Gerber *et al*, 2013). However, the incorporation of FE as selection criteria in animal breeding programs is costly and time consuming. Daily feed intake and weight gain for a large number of animals need to be recorded for at least 70 days to obtain accurate estimates of FE (Archer *et al*, 1997).

In the past years, several studies have been carried out with the aim to identify molecular markers associated with FE to enable the faster and cost-effectively identification of superior animals (de Oliveira *et al*, 2014; Rolf *et al*, 2011; Santana *et al*, 2014; Seabury *et al*, 2017). However, for each population, different biological processes seem to be identified (de Oliveira *et al*, 2014; Rolf *et al*, 2011; Santana *et al*, 2014; Seabury *et al*, 2017). Probably, that is because FE is a multifactorial trait and many different biological mechanisms seems to be involved in its regulation (Herd *et al*, 2004; Herd & Arthur, 2009). It has been indicated that high FE animals present increased mitochondrial function (Lancaster *et al*, 2014; Connor *et al*, 2010), less oxygen consumption (Gonano *et al*, 2014) and delayed puberty (Randel & Welsh, 2013; Shaffer *et al*, 2011; Fontoura *et al*, 2016). On the other hand, low FE animals have increased physical activity, ingestion frequency and stress (Francisco *et al*, 2015; Cafe *et al*, 2011; Kelly *et al*, 2010; Chen *et al*, 2014), increased leptin and cholesterol levels (Alexandre *et al*, 2015; Foote *et al*, 2016; Nkrumah *et al*, 2007; Mota *et al*, 2017), higher subcutaneous and visceral fat (Santana *et al*, 2012; Gomes *et al*, 2012; Mader *et al*, 2009),

higher energy wastage as heat (Montanholi *et al*, 2010, 2009; Archer *et al*, 1999) and more hepatic lesions associated with inflammatory response (Alexandre *et al*, 2015; Paradis *et al*, 2015).

In the context of such a complex trait, we perform a multiple-tissue transcriptomic analyses of high and low FE Nellore cattle across tissues related to endocrine control of hanger/satiety, hydric and energy homeostasis, stress and immune response, physical and sexual activity, as is the case of hypothalamus-pituitary-adrenal axis and organs as liver and skeletal muscle. Based on gene co-expression across tissues and conditions we derived a regulatory network revealing NR2F6 and TGFB signalling as key regulators of hepatic inflammatory response and muscle tissue development, respectively. Next, we apply advanced motif discovery methods which i) validate that co-expressed genes are enriched for NR2F6 and TGFB signalling effector molecule SMAD3 binding sites in their 10KB upstream regions and ii) predict direct transcription factor (TF) – Target gene (TG) interactions at the sequence level. These binding interactions were experimentally validated with public TF ChIP-seq from ENCODE. Regulatory activity in the tissues of interest was also confirmed by performing an enrichment analysis on open chromatin tracks and histone chromatin marks across cell types and tissues in the human and cow genome. Moreover, we propose a hormonal control of differences in metabolism and sexual maturity between high and low FE animals, indicating potential biomarkers for further validation such as adrenomedullin, FSH, oxytocin, somatostatin and TSH.

Results

Multi-tissue transcriptomic data reveal differences between high and low feed efficient animals

Feed efficiency is a complex trait characterized by multiple distinct biological processes including metabolism, ingestion, digestion, physical activity and thermoregulation (Herd *et al*, 2004; Herd & Arthur, 2009). To study FE at transcriptional level we performed RNAseq of five tissues (i.e. adrenal gland, hypothalamus, liver, muscle and pituitary) from nine male bovines of high feed efficiency (HFE, characterized by low residual feed intake (RFI) (Koch *et al*, 1963)) and nine of low FE (LFE, characterized by high RFI). In total, we analysed 18 samples of liver, hypothalamus and pituitary; 17 of muscle and 15 of adrenal gland, yielding 13 million reads per sample on average (S1 Supporting Information). Gene expression was estimated for 24,616 genes present in the reference genome (UMD 3.1) and after quality control (refer to methods), 17,354 genes were identified as being expressed in at least one of the five tissues analysed.

Differential expression (DE) analysis between high and low FE animals resulted in 471 DE genes across tissues (P<0.001, S2 Supporting Information), namely, 111 in adrenal gland, 125 in hypothalamus, 91 in liver, 104 in muscle and 98 in pituitary (S3A-E Supporting Information). Although no significant functional enrichment was found for the 281 genes up-regulated in high feed efficiency, the 248 genes down-regulated presented a significant enrichment of GO terms such as response to hormone (Padj=5.43 x 10^{-6}), regulation of hormone levels (Padj=3.48 x 10^{-6}), cell communication (Padj=3.18 x 10^{-4}), regulation of signaling receptor activity (Padj=3.20 x 10^{-4}), hormone metabolic process (Padj=5.86 x 10^{-4}), response to corticosteroid (Padj=6.28 x 10^{-4}), regulation of secretion (Padj=7.2 x 10^{-4}), response to lipopolysaccharide (Padj=7.9 x 10^{-4}) and regulation of cell proliferation (Padj=1.86 x 10^{-3}). Refer to S4 Supporting Information to see all enriched terms.

Overlap between gene selection criteria prioritizes genes associated with feed efficiency

The genetic architecture behind complex traits involves a large variety of genes with coordinated expression pattern, which can be represented by gene regulatory networks as a blueprint to study their relationships and to identify central regulatory genes (Swami, 2009). Therefore, it is important to select relevant genes and gene families according to the phenotype of interest to be used for network analysis. We defined five categories of genes (see methods for further information) for inclusion in co-expression analysis: 1 - differentially expressed (DE), 2 - genes harbouring SNPs previously associated with FE (harbouring SNP), 3 - tissue specific (TS), 4 - genes coding proteins secreted in plasma by any of the five tissues analysed (secreted) and 5 - key regulators.

As reported before, we have identified 471 DE genes between high and low FE animals (Figure 1A, S5A Supporting Information). In addition, 267 genes were selected for harbouring SNPs previously associated with FE, as not only differences in expression levels can influence the phenotype but also polymorphism in the DNA sequence that can alter the translated protein behaviour (S5B Supporting Information). Moreover, 396 were selected for being tissue specific (refer to methods for definition); 22 in adrenal gland, 32 in hypothalamus, 215 in liver, 218 in muscle and 9 in pituitary (S5C Supporting Information). A total of 244 genes coding proteins secreted in plasma were selected because of their potential as biomarkers of FE (S5D Supporting Information). From those, 135 had liver as the tissue of maximum expression and were functionally enriched for GO terms such as complement activation (Padj=1.82 x 10^{-19}), regulation of acute inflammatory response (Padj=1.89 x 10^{-14}), innate immune response (Padj=9.71 x 10^{-12}), negative regulation of endopeptidase activity (Padj=2.35 x 10^{-10}), platelet degranulation (Padj=1.08 x 10^{-10}), regulation of coagulation (Padj=3.39 x 10^{-9}), triglyceride homeostasis (Padj=1.23 x 10^{-6}), cholesterol efflux (Padj=1.03E^{-5}) (S6 Supporting Information). Finally, from 1570 potential regulators in public available TFdb, 78 were identified as key regulators of the genes selected by all the other

144 categories, i.e. 78 genes presented a coordinated expression level with many of the genes in
145 the network reflecting a tight control of expression pattern across tissues (S5E Supporting
146 Information).

147 Considering all the inclusion criteria, 1,317 genes were selected to be included in co-
148 expression network analysis (Figure 1B, S7 Supporting Information), some of them selected
149 by more than one category (Figure 1C). Regarding DE genes, six of them were also reported
150 before as harbouring SNPs associated with the phenotype (*LUZP2*, *MAOB*, *SFRS5*,
151 *SLC24A2*, *SOCS3* and *WIF1*) and 13 of them were key regulators (*HOPX*, *PITX1*, *CRYM*,
152 *PLCD1*, *ND6*, *cytb*, *ND1*, *MT-ND4L*, *ND5*, *ATP8*, *ND4*, *ENSBTAG00000046711* and
153 *ENSBTAG00000048135*). Many of the genes that are both DE and regulators are involved in
154 respiratory chain (*ND6*, *cytb*, *ND1*, *MT-ND4L*, *ND5*, *ATP8* and *ND4*) and were all up-
155 regulated in high FE group.

156 Considering both DE and secreted genes, 18 were identified (*NOV*, *SPP1*, *CTGF*,
157 *OXT*, *PTX3*, *VGF*, *CCL21*, *COL1A2*, *PGF*, *SOD3*, *SERPINE1*, *PRL*, *PON1*, *SST*, *JCHAIN*,
158 *PCOLCE*, *IGFBP6* and *SCG2*). In addition, four genes were DE, secreted and tissue specific,
159 two from liver (*CXCL3* and *IGFBP1*) and two from pituitary (*NPY* and *CYP17A1*). Genes
160 *RARRES2* and *PENK* (proenkephalin) were DE, secreted and had been previously reported as
161 harbouring SNP associated with FE [30, AnimalQTLdb]. Other DE genes worthy to
162 highlight, due to their well-known role in metabolic processes, are *AMH* (anti-mullerian
163 hormone), *TSHB* (thyroid stimulating hormone beta), *FGF21* (Fibroblast growth factor 21)
164 and *FST* (follistatin), up-regulated in high FE group, and *PMCH* (pro-melanin concentrating
165 hormone), *ADM* (adrenomedullin) and *FSHB* (follicle stimulating hormone beta), up-
166 regulated in low FE group.

167

168

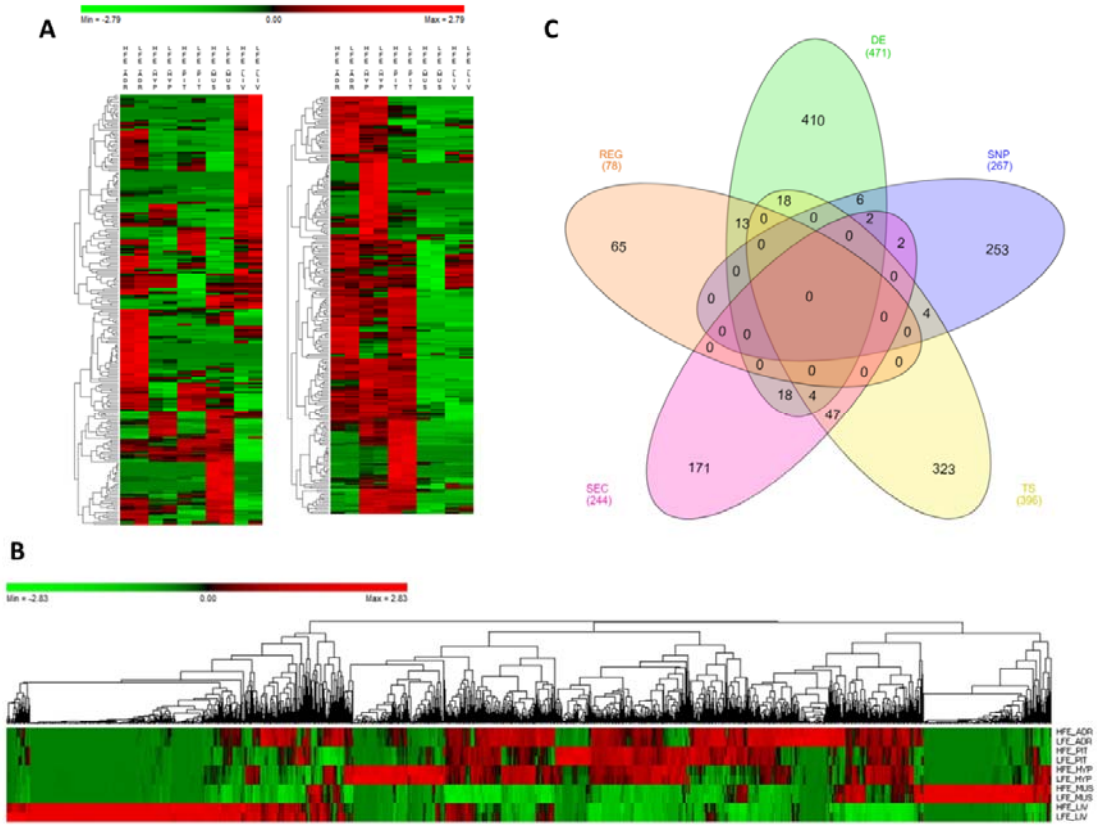


Figure 1. Genes selected for co-expression network construction. A) Heatmap of normalized mean expression (NME) of 471 differentially expressed (DE) genes between high (HFE) and low (LFE) feed efficient animals in adrenal gland (ADR), hypothalamus (HYP), liver (LIV), muscle (MUS) and pituitary (PIT). Genes (rows) and samples (columns) are organized by hierarchical clustering based on Euclidean distances. **B)** NME heatmap of all 1,317 genes selected for network construction. Genes (columns) and samples (rows) are organized by hierarchical clustering based on Euclidean distances. **C)** Venn diagram of 1,317 genes selected for network construction. The inclusion criteria for selecting genes were divided in five categories: differentially expressed genes (DE), tissue specific genes (TS), genes harbouring SNPs reported by literature as being associated with feed efficiency in beef cattle (SNP), genes encoding proteins secreted by at least one of the tissues in plasma (SEC) and key regulators (REG). Numbers between brackets indicate the total number of genes in each category.

Co-expression network reveals regulatory genes and biological processes related to feed efficiency

The co-expression network (Figure 2) was composed by 1,317 genes and 91,932 connections, with a mean of 70 connections per gene. Most of the connections (51%) involved a DE gene and 23% of those were between two DE genes. Tissue specific (TS) gene

were involved in 49% of the connections with 119 connections per gene in average, which was higher than the overall network mean and reflects the close relationship between genes involved in tissue specific functions. Key regulators was the least represented category in the network (only 78 genes) but accounted for 11% of the connections in the network with the highest value of mean connections per gene, 131 connections, which is in accordance with their regulatory role. Regarding the connections within tissues, when we ranked all the genes in the network by the number of connections and looked at the top 50 genes, 29 were from liver, 15 were from muscle and 3, 2 and 1 were from pituitary, adrenal gland and hypothalamus, respectively. This result indicates a very well-coordinated expression pattern in liver and muscle that could be a reflex of the number of TS genes in those tissues and the presence of central regulatory genes coordinating the expression of many other genes.

In the network (Figure 2), genes grouped together by tissue which was mostly driven by TS genes. As mentioned before, most of the secreted proteins coding genes were locate in the liver. Most of the key regulators were located peripherally in relation to the clusters which could be reflecting their regulatory nature independent of tissue specificity. Despite that, some regulators draw attention because of their high number of connections.

The top five most connected regulators were *EPC1*, *NR2F6*, *MED21*, *ENSBTAG00000031687* and *CTBP1*, varying from 317 to 284 connections. They were all first neighbours of each other and were connected mainly to genes with higher expression in liver and essentially enriched for acute inflammatory response (Padj=4.5 x 10^{-13}, S8 Supporting Information). The next most connected regulator is *TGFB1* with 217 connections. It is mainly connected to genes from muscle that are primarily enriched for muscle organ development (Padj=6.87 x 10^{-5}) and striated muscle contraction (Padj=1.39 x 10^{-5}, S9 Supporting Information). Besides indicating main regulator genes, network approach can be useful to access the role of specific genes. For instance, gene *FGF21*, a hormone up regulated

in liver of high FE animals, is directly connected to genes enriched for plasma lipoprotein particle remodelling, regulation of lipoprotein oxidation and cholesterol efflux (Padj=5.64E-3, S10 Supporting Information). Indeed, according to the literature, this gene is associated to decrease in body weight, blood triglycerides and LDL-cholesterol (Cheung & Deng, 2014).

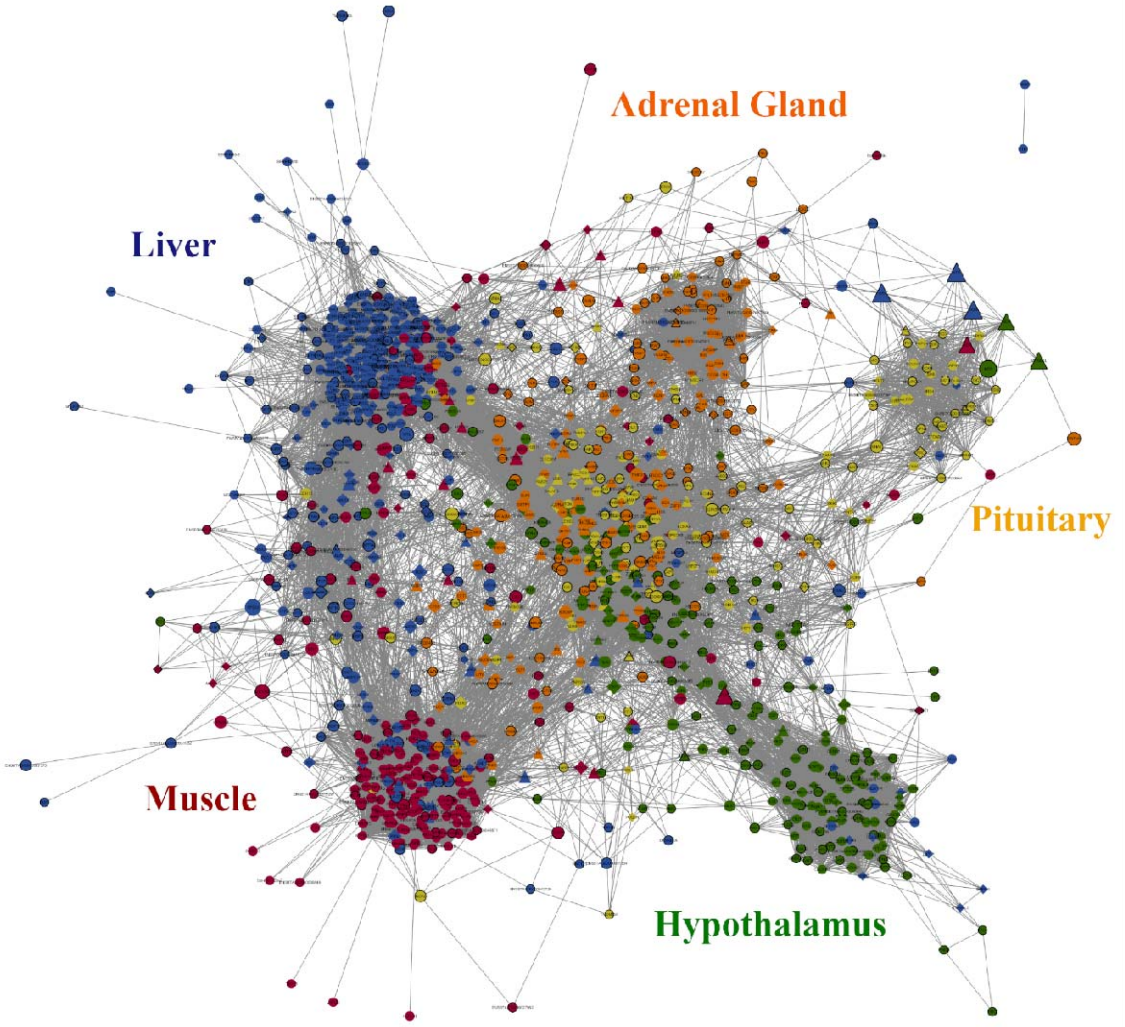


Figure 2. Gene co-expression network constructed using PCIT algorithm on 1,317 selected genes (see methods). Nodes with diamond shape correspond to secreted proteins coding genes and triangles correspond to key regulators; all the other genes are represented by ellipses. Nodes with black borders are differentially expressed between high and low feed efficiency. Colours are relative to the tissue of maximum expression: blue represent liver, red represent muscle, yellow represent pituitary, green represent hypothalamus and orange represent adrenal gland. The size of the nodules is relative to the normalized mean expression values in all samples. Only correlations above 0.9 and bellow -0.9 and its respective genes are shown in this figure.

Motif discovery confirms NR2F6 as a key regulator of liver transcriptional changes between high and low feed efficiency

By means of the power-law theory, co-expression networks present many nodules with few connections and few central nodules with many connections (de la Fuente, 2010), being the last ones indicated as central regulatory genes responsible for the transcriptional changes between the divergent phenotypes analysed. In our study, the most connected regulators were indicated, together with their target genes, i.e. their first neighbours in the network. Those genes are a mixture of direct and indirect regulator targets. In order to validate the regulatory role of the most connected regulators in the network and identify their core direct targets we performed motif discovery in their co-expressed target genes. It is noteworthy motif discovery should confirm the presence of DNA motifs of a TF in the regulatory regions of co-expressed genes. From the top five most connected regulators from our previous co-expression analysis, only NR2F6 has the ability to bind DNA. In contrast, the other four regulators act mainly as cofactors (corepressor, i.e. CTBP1; coactivator, i.e. MED21; or histones modifier, i.e. EPC1), that is co-binding through protein-protein interactions.

The analysis of 313 co-expressed genes with NR2F6 yield the Nuclear Factor motif HNF4-NR2F2 (transfac_pro-M01031) as the second motif most enriched out of 9732 PWMs with a Normalized Enrichment Score (NES) of 7.99 (Figure 3B). In addition, a total of 19 motifs associated with HNF4-NR2F2 were enriched in the dataset, associating HNF4-NR2F2 to 168 direct target genes (Figure 3C). Due to motif redundancy or highly similarity between a plethora of TFs, these motifs can be associated with multiple TFs from HNF4 (direct) to several nuclear factors such as NR2F6 (motif similarity score FDR 1.414E-5). However, our co-expression analysis strongly indicates NR2F6 is the key TF, since it was the TF with the

253 highest number of nodes in the co-expression network (Figure 3C) and neither HNF4 nor
254 NR2F2 were prioritized by any selection category to be included in the network.

255 Each of the NR2F6 inferred direct target genes contain one or more predicted
256 enhancers, i.e. regions with high-scoring motif binding sites for NR2F6 or TFs with highly
257 similar motifs. To validate the binding of these genomic regions by NR2F6 or TFs with
258 highly similar motifs to NR2F6 we performed a region enrichment analysis of our predicted
259 NR2F6 binding sequences against public TF ChiP-seq bound regions in human cell lines
260 from ENCODE the ENCODE consortium (1394 TF binding site tracks). This analysis,
261 confirms the experimental binding of TFs with similar binding as NR2F6 in HepG2 cells,
262 HNF4A (NES=8.57), HNF4G (NES=7.83), RXRA (NES=6.85), and NR2F2 (NES=4.45) as
263 the most enriched tracks (S11 Supporting Information). Recent NR2F6 ChIP-seq data in
264 HepG also confirms an enrichment for NR2F6 (Figure 3D), indicating predicted NR2F6
265 binding regions are experimentally bound by NR2F6 in hepatocyte cell lines (Figure 3D).

266 Next, to validate that the NR2F6 binding in those regions is functional in liver we
267 performed an enrichment analysis for open-chromatin (tracks=655) and histone modifications
268 (tracks =2450) related to active regulatory elements (S12 Supporting Information). This
269 analysis yielded FAIRE-seq on HepG2 cell lines and H3K9ac and H3K4me3 in adult liver
270 (E066 Roadmap Epigenomics Track) as the most enriched tracks respectively, strongly
271 indicating not only predicted target enhancers are bound by NR2F6 in Hepatocyte cell lines
272 but these regulatory regions are functionally active in hepatocytes and human liver (Figure
273 3D).
274

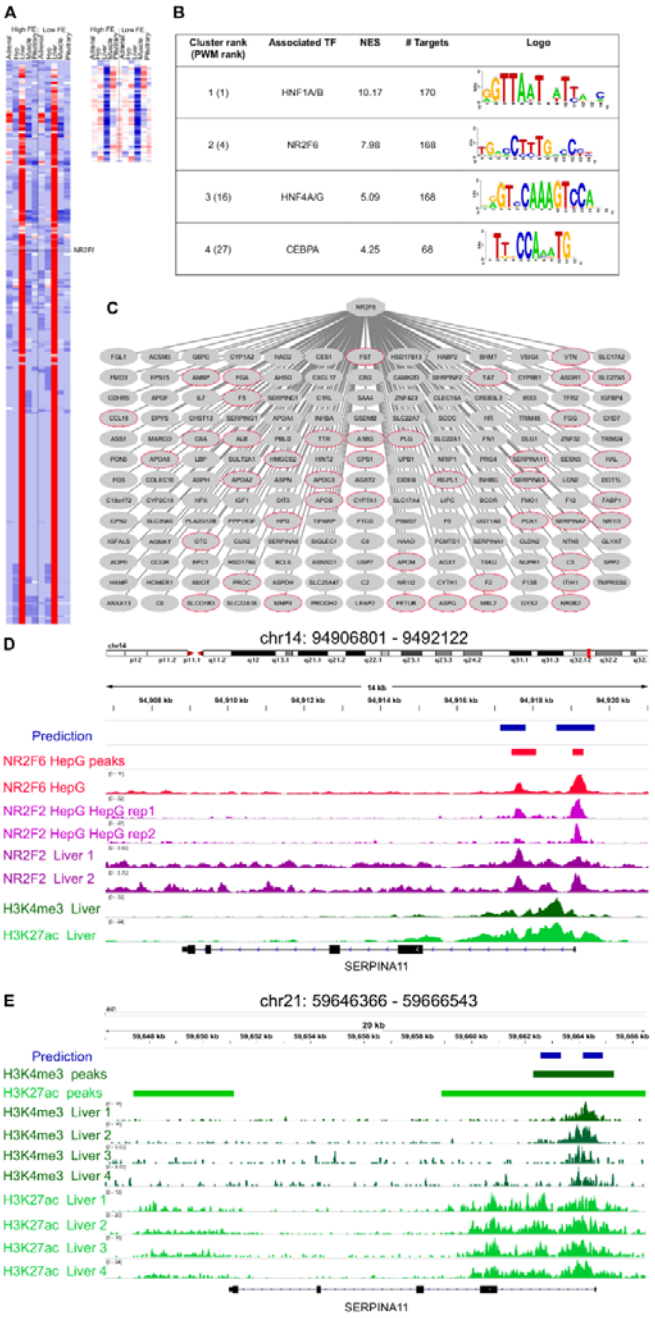


Figure 3. Mapping of NR2F6 direct targets. A) Heatmap of the 313 genes coexpressed with NR2F6 across all samples (derived from the co-expression analysis), B) i-cisTarget motif discovery results on the genes shown in (A), C) Predicted NR2F6 targetome. A red node indicates genes known to be targeted my NR2F6 in human Hepatocytes. D) Example of predicted NR2F6 target regions for SERPINA1 gene. The predicted enhancer overlaps the exact position for NR2F6 and NR2F2 binding in HepG sites from ENCODE dataset as well as histone chromatin marks related with active regulatory regions, namely H3K27ac, and promoters, H3K4me3 in human primary tissue from RoadMap Epigenetics E) The enhancer prediction in cow coordinates (bosTau6) overlaps a region marked with H3K4me3 in cow liver (Villar *et al*, 2015).

Regarding the cow genome, a recent open-chromatin study (Villar *et al*, 2015) has delineated the map active promoters and enhancers by H3K4me3 and H3K27ac ChIP-seq in cow liver resulting in 13796 promoter and 45786 enhancers (S13 Supporting Information). We performed an enrichment analysis of predicted NR2F6 enhancers converted to cow coordinates (n=779) resulting in 446 regions being identified as functional regulatory regions in cow liver. This number is significantly higher compared to only 43 regions are expected to overlap by random (1000 permutation tests) (Figure 3E).

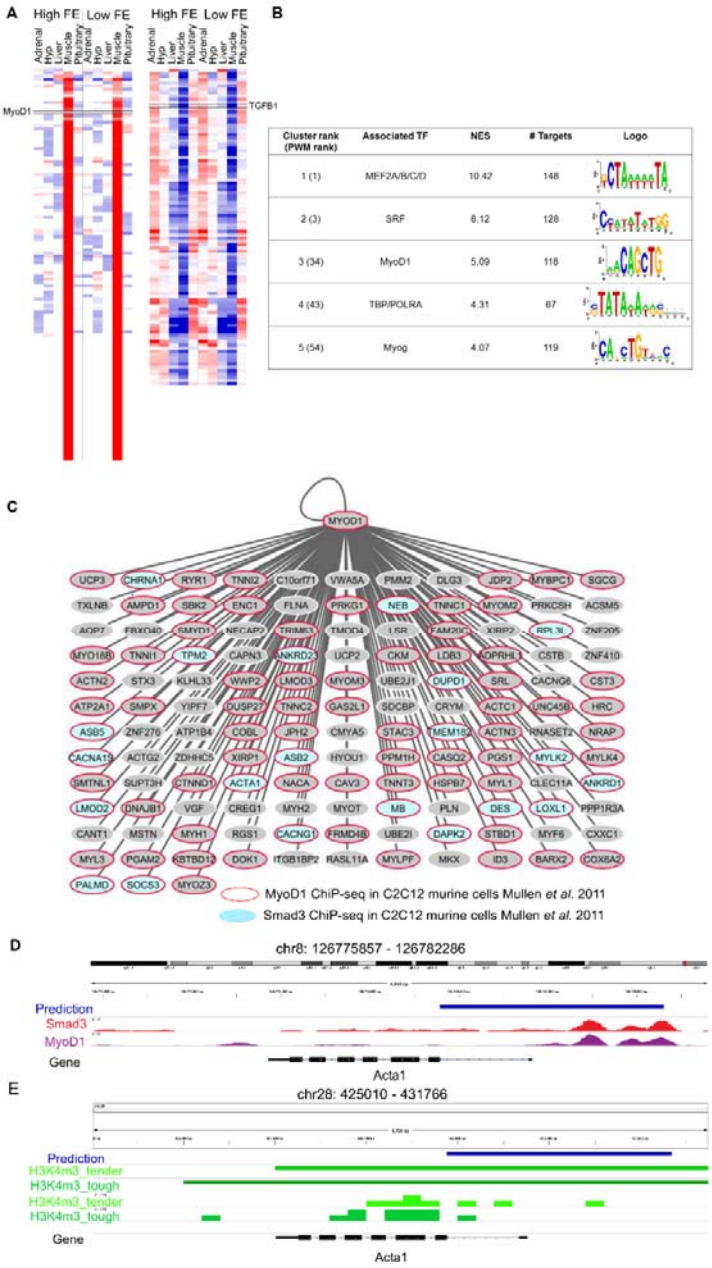
Finally, in addition to NR2F6 motif, HNF1A motif was found as a potential co-regulator in liver, in particular swissregulon-HNF1A.p2 with a NES =10.17 and in total 20 enriched motifs and 170 direct targets were associated to HNF1A (Figure 3B). HNF1 is a master regulator of liver gene expression (Tronche & Yaniv, 1992), thus making its finding justified.

Motif discovery validates TGF-beta signalling through Smad3/MyoD1 binding as drivers of transcriptional differences in muscle of divergent feed efficient cattle

The analysis of the 217 genes co-expressed with *TGFB1* (Figure 4A) showed most target genes motifs were enriched for master regulators of muscle differentiation, namely, *MEF2* (NES=10.42) a MADS box Transcription factor with 148 target genes, and *MYOD1* (NES=8.12), a bHLH transcription factor (CANNTG) with 135 direct target genes (Figure 4B, S14 Supporting Information). To evaluate the precision of our predicted *MYOD1* (bHLH) target genes we assessed how many of these TF-TG relationships had been previously experimentally reported. Based on *MYOD1* ChIP-seq binding in mouse myotubules, 86 genes had already been associated with *MYOD1* resulting in 63% success rate (hypergeometric test 1.72E-22). *SMAD3*, the effector molecule of *TGFB1* signalling is known to recruit *MYOD1* to drive transcriptional changes during muscle differentiation (Mullen *et al*, 2011). Thus, we

312 evaluated whether predicted *MYOD1* target genes were enriched for known *SMAD3* target
313 genes resulting in 21 out of 135 *MYOD1* predicted target genes presented *SMAD3* ChIP-seq
314 binding in myotubes. Thus indicating there is a statistically significant association between
315 *MYOD1* target genes and *SMAD3* target genes in myotubes (hypergeometric test 1.98 E-6)
316 (Figure 4 C) (Mullen *et al*, 2011) In contrast, no significant association was found between
317 predicted *MYOD1* target genes in this study and *SMAD3* target genes in other cell lines, such
318 as pro-B and ES cell (hypergeometric test 0.056 and 0.076, respectively) (Mullen *et al*,
319 2011). That is in agreement that the effect of TGFB signalling driven by SMAD3 DNA
320 binding is tissue-specific (Liu *et al*, 2001). Our analysis predicted 621 potential *MYOD1*
321 binding sites, of which 114 (18%) and 153 (24.5%) present a *MYOD1* ChIP-seq signal in
322 mouse C2C12 myotubes cells (Mullen *et al*, 2011) and in primary myotubes (Cao *et al*,
323 2009), respectively.

324 Finally, we evaluate whether predicted MYOD1 binding regions were regulatory
325 regions active in muscle cells across different species, namely human (S15 Supporting
326 Information), mouse (S16 and S17 Supporting Information) and cow (S18 Supporting
327 Information). To tackle this issue we performed an enrichment analysis across 2113 open-
328 chromatin ENCODE tracks. This analysis resulted in a clear enrichment of our predicted
329 MYOD1 binding regions with H3K27ac (NES=15.98) and H3K9ac (NES=8.78) regions in
330 skeletal muscle (Figure 4D). Both chromatin marks are associated with active transcription,
331 H3K27ac related to active enhancers and H3K9ac related to active gene transcription (Shin *et*
332 *al*, 2012) thus, validating most of our enhancer predictions in human are active in skeletal
333 muscle. Once in cow, we assess the overlap of predicted MYOD1 enhancers and promoter
334 regions in cow muscle experimentally detected with H3K4me3 (Cao *et al*, 2009). This
335 resulted in 282 regions out of 671 (42 %) overlap when only 11 regions are expected to
336 overlap by random 1000 permutation test) (Figure 4E).



337

338 **Figure 4. Mapping the downstream network of TGFB signaling through**
339 **SMAD3/MyoD1 DNA binding.** A) Heatmap of the 217 genes coexpressed with TGFB1
340 (derived from the co-expression analysis). B) i-cisTarget motif discovery results on the genes
341 shown in (A), C) Predicted MyoD targetome. A red node indicate genes know to be targeted
342 my MyoD1 in murine myotubes (Mullen *et al*, 2011). Blue nodes indicate genes to be
343 targeted by SMAD3, the effector DNA binding molecular of TGFB signalling, in murine
344 myotubes (Mullen *et al*, 2011). D) Example of predicted MyoD1 target regions for Acta1
345 gene. The predicted enhancer overlaps the exact position for SMAD3 and MyoD1 ChIP-seq
346 binding in murine myotubes (Mullen *et al*, 2011). E) The enhancer prediction in cow
347 coordinates (bosTau6) overlaps a promoter region marked with H3K4me3 in muscle tissue in
348 cow (Cao *et al*, 2009).

Finally, we evaluate whether predicted MYOD1 binding regions were regulatory regions active in muscle cells across different species, namely human (S15 Supporting Information), mouse (S16 and S17 Supporting Information) and cow (S18 Supporting Information). To tackle this issue we performed an enrichment analysis across 2113 open-chromatin ENCODE tracks. This analysis resulted in a clear enrichment of our predicted MYOD1 binding regions with H3K27ac (NES=15.98) and H3K9ac (NES=8.78) regions in skeletal muscle (Figure 4D). Both chromatin marks are associated with active transcription, H3K27ac related to active enhancers and H3K9ac related to active gene transcription (Shin *et al*, 2012) thus, validating most of our enhancer predictions in human are active in skeletal muscle. Once in cow, we assess the overlap of predicted MYOD1 enhancers and promoter regions in cow muscle experimentally detected with H3K4me3 (Cao *et al*, 2009). This resulted in 282 regions out of 671 (42 %) overlap when only 11 regions are expected to overlap by random 1000 permutation test) (Figure 4E).

Differential co-expression

Although the general co-expression network give us important insights about regulatory genes and their behaviour, by creating specific networks for high and low FE and comparing the connectivity of the genes in each one, we can identify genes that change their behaviour depending on the situation, moving from highly connected to lowly connected and vice-versa. We were able to identify 87 differentially connected genes between high and low FE ($P<0.05$); 63 mainly expressed in liver, 19 in muscle and 3, 1 and 1 in hypothalamus, adrenal gland and pituitary, respectively (S19 Supporting Information). Those genes were enriched for terms such as regulation of blood coagulation (Padj=3.14 x 10^{-10}), fibrinolysis (Padj=7.71 x10^{-7}), platelet degranulation (Padj=7.49 x10^{-6}), regulation of peptidase activity (Padj=6.16 x 10^{-4}), antimicrobial humoral response (Padj=2.49 x10^{-3}), acute inflammatory

response (Padj=2.18 x10^{-4}) and induction of bacterial agglutination (Padj=3.58 x 10^{-2}) (S20 Supporting Information). It is important to highlight 20 of the differentially connected genes were also differentially expressed (Table 1) and three of them, i.e. *SST*, *JCHAIN* and *IGFBP1*, were secreted in plasma as well, which make them very promising potential biomarkers.

Table 1. Differentially connected and differentially expressed genes between high and low feed efficiency.

Gene name	Number of connections		Category*	Tissue of maximum expression	Tissue of differential expression
	Low feed efficiency	High feed effciency			
SST	0	45	DE, SEC	Hypothalamus	Hypothalamus
SNORA73	41	108	DE	Liver	Liver
ENSBTAG00000047700	56	111	DE	Liver	Liver
ENSBTAG00000047121	62	111	DE	Liver	Liver
ENSBTAG00000047816	53	96	DE	Liver	Liver
ENSBTAG00000039928	50	89	DE	Liver	Liver
ANXA13	115	63	DE	Liver	Liver
FST	113	56	DE	Liver	Liver
PBLD	115	55	DE	Liver	Liver
ENSBTAG00000021368	95	0	DE	Liver	Liver
JCHAIN	52	113	DE, SEC	Liver	Liver
IGFBP1	55	0	DE, TS, SEC	Liver	Liver
SBK2	0	70	DE	Muscle	Muscle
ACTC1	54	0	DE	Muscle	Muscle
MYH1	0	47	DE, TS	Muscle	Muscle
HR	119	50	DE	Pituitary	Muscle
TAGLN	83	31	DE	Adrenal	Muscle, Pituitary
SFRP2	41	91	DE	Hypothalamus	Pituitary
FN1	119	69	DE	Liver	Pituitary
CAV1	98	50	DE	Muscle	Pituitary

*Differentially expresses genes between high and low feed efficiency (DE), tissue specific genes (TS) and genes encoding proteins secreted in plasma (SEC).

Comparing the two networks there was no large difference regarding the number of genes and connections. While high FE network contained 1,074 genes in total and 28,018 connections, low FE network was composed by 1,098 genes and 30,705 connections. For all tissues, low FE networks showed more connections but the difference is slight, being the bigger difference of 44 versus 40 connections per gene in liver.

Discussion

Feed efficiency is a complex trait, regulated by several biological processes. Thus, the indication of genomic regions associated with this phenotype, as well as regulators genes and

394 biomarkers to select superior animals and to direct management decisions is still a great
395 challenge. In this work, multi-tissue transcriptomic data of high and low feed efficient
396 Nellore bulls were analysed through robust co-expression network methodologies in order to
397 uncover some of the biology that governs this traits and put forward candidate genes to be
398 focus of further research. In this sense, the validation of target genes of main transcription
399 factors (key regulators) in the network by motif search proves the efficacy of the
400 methodology for network construction and prioritizes some transcription factors as central
401 regulators (Aerts *et al*, 2010; Naval-Sańchez *et al*, 2013; Potier *et al*, 2014). Moreover, the
402 addition of a category of genes coding proteins secreted in plasma in the co-expression
403 analysis highlight genes with potential to be explored as biomarkers of feed efficiency. We
404 were able to identify genes related to main biological process associated with feed efficiency
405 and indicate key regulators, as can be seen on the following lines.

406 Firstly, it is worthy to mention the 98 animals used to select the high and low FE
407 groups in this study have been previously analysed regarding several phenotypic and
408 molecular measures (Alexandre *et al*, 2015; Mota *et al*, 2017; Novais *et al*, 2018). It was
409 observed high and low FE groups had similar body weight gain, carcass yield and loin eye
410 area but low FE animals had higher feed intake, greater fat deposition, higher serum
411 cholesterol levels, as well as hepatic inflammatory response, indicated by transcriptome
412 analysis of liver biopsy and proved by the higher number of periportal mononuclear infiltrate
413 (histopathology) and increased serum gamma-glutamyl-transferase (GGT, a biomarker of
414 liver injury) in this group (Alexandre *et al*, 2015). In the present study, the simultaneous
415 analysis of five distinct tissues revealed a prominence of the hepatic tissue. Liver presented
416 the most connected genes in the network, the higher number of differentially connected genes
417 and higher number of secreted genes, which although can be explained by its biological
418 function, are enriched mostly for terms related to lipid homeostasis and inflammatory

419 response. Moreover, the top five most connected regulators in the network are co-expressed
420 mainly with genes highly expressed in liver and also enriched for inflammatory response.

421 The relationship between FE and genes or pathways related to immune response and
422 lipid metabolism is becoming more evident, as recent studies also reported it in beef cattle
423 (Karisa *et al*, 2014; Weber *et al*, 2016; Paradis *et al*, 2015; Zarek *et al*, 2017; Mukiibi *et al*,
424 2018) and pigs (Gondret *et al*, 2017; Ramayo-Caldas *et al*, 2018). In our previous work
425 (Alexandre *et al*, 2015), we proposed increased liver lesions associated with higher
426 inflammatory response in liver of low FE animals could be due to increased lipogenesis
427 and/or higher bacterial infection in the liver. While further evidence is needed to test these
428 hypotheses, the enrichment of terms such as induction of bacterial agglutination and response
429 to lipopolysaccharide makes bacterial infection a strong possibility. Indeed, pigs with low FE
430 were reported to have higher intestinal inflammation, neutrophil infiltration biomarkers and
431 increased serum endotoxin (lipopolysaccharide and other bacterial products) which could be
432 related to increased bacterial infection or to decreased capacity to neutralize endotoxins
433 (Mani *et al*, 2013). The authors hypothesized differences in bacterial population could
434 partially explain the increase in circulating endotoxins, which could also be true for cattle
435 given that differences in intestinal and ruminal bacterial population between high and low FE
436 animals have already been reported (Myer *et al*, 2015, 2016). Furthermore, the literature
437 reports lipopolysaccharides (LPS) may cause up-regulation of adrenomedullin (*ADM*)
438 hormone (Shindo *et al*, 1998), an up-regulated gene in low FE individuals as showed here. It
439 was also demonstrated in rats that intravenous infusion of LPS caused up-regulation of *ADM*
440 in ileum, liver, lung, aorta, skeletal muscle and blood vessels (Shoji *et al*, 1995) whereas in
441 our study, *ADM* presented differential expression in muscle, but not in liver.

442 Against pathogen invasion, a tightly regulated adaptive immune response must be
443 triggered in order to allow T lymphocytes to produce cytokines or chemokines and B cells to

444 differentiate and produce antibodies (Hermann-Kleiter & Baier, 2014). This regulation is
445 known to be strongly influenced by the expression level and transcriptional activity of several
446 nuclear receptors, including the NR2F-family, which consists of three orphan receptors:
447 NR2F1, NR2F2 and NR2F6 (Hermann-Kleiter & Baier, 2014). Those receptors present
448 highly conserved DNA and ligand binding domains among each other and across species
449 (Pereira *et al*, 2000), and all three are expressed in adaptive and immune cells (Hermann-
450 Kleiter & Baier, 2014). In our study, NR2F6 appeared as the second most connected
451 regulator gene in the network while the other family members, although present in our
452 expression data, were not selected by any of our inclusion criteria, thus indicating they might
453 not be so relevant in our conditions. Indeed, NR2F6 appears to be a critical regulatory factor
454 in the adaptive immune system by directly repressing the transcription of key cytokine genes
455 in T effector cells (Hermann-Kleiter *et al*, 2008; Klepsch *et al*, 2016). The role of NR2F6 as a
456 key regulator of inflammatory response in our network was validated at gene level by the
457 identification of the binding motif HNF4-NR2F2 (transfac_pro-M01031) as one of the most
458 enriched in NR2F6 target genes, due to the high similarity between NR2F2 and NR2F6
459 biding sites. Furthermore, using open chromatin data public available, we provided
460 experimental evidence of the binding of TFs with highly similar binding motifs as NR2F6 in
461 hepatocyte cells in humans and in cattle, thus, indicating predicted target enhancers are
462 functional in this tissue.

463 Another regulator prioritized in our analysis is *TGFB1*, the sixth most connected gene
464 in the co-expression network, and a potential driver of transcriptional changes between high
465 and low FE cattle in muscle. This gene has been previously pointed as a master regulator of
466 FE in beef cattle, using genomics and metabolomics data (Widmann *et al*, 2015). Moreover,
467 our motif discovery analysis showed *TGFB1* co-expressed genes are mostly enriched for
468 binding site of master regulators of muscle differentiation as *MEF2* and *MYOD*. Indeed,

public available data show many of *TGFB1* target genes were associated with *MYOD* (Mullen *et al*, 2011). It is known signalling pathways are an effective mechanism for cells to respond to environmental cues by regulation gene expression. *TGFB1* signalling triggers the phosphorylation of *SMAD2/3* transcription factors, which co-bind with cell-type master regulators at the nuclear level allowing/triggering/leading to cell-type specific transcriptional changes (Schmierer & Hill, 2007; Mullen *et al*, 2011). In skeletal muscle cells, myoblasts and myotubes, *SMAD3* co-binds with *MYOD1* (Mullen *et al*, 2011). The overlap between *MYOD1* and *SMAD3* target genes demonstrate the significant association between both genes in skeletal muscle, in agreement with the tissue-specificity *TGFB1* signalling response (Mullen *et al*, 2011). The overlap percentage between our predicted binding sites and *MYOD1* Chip-seq data (18 and 24.5%) confirms previous analysis in mice where they reported only 20% of experimental validated distal enhancers in mouse myotubes with a bHLH (MyoD1) binding were actually bound by *MYOD* ChIP-seq data (Blum et al. 2012). Thus, suggesting additional transcription-factors and/or histone modification have a key role in *MYOD1* binding. The *SMAD3/MYOD1* co-bound regions for known target genes are also captured, such as the promoter regions of *ACTA1* and *ANKRD1*, both genes involved in skeletal muscle differentiation (Figure 4C). We also demonstrated predicted *MYOD1* binding regions are enriched for muscle regulatory regions across species (human, mouse and cow).

Altogether, we showed co-expressed genes with *TGFB1* are enriched for *SMAD3/MYOD1* binding sites, which we validate at the gene and enhancer level by proving not only *MYOD1* and *SMAD3* binding, but also their accessibility, in human, mouse and cow. In pigs, it has been indicated increased feed efficiency is associated with stimulation of muscle growth through TGF-β signalling pathway (Jing *et al*, 2015). Finally, although not directly co-expressed with *TGFB1*, oxytocin (*OXT*) was DE in muscle and despite the lack of knowledge on its role in this tissue, previous work in cattle have shown a massive increase of

494 *OXT* expression in muscle of bovines chronically exposed to anabolic steroids (Jager *et al*,
495 2011). It is not known yet if oxytocin alone have an anabolic activity, but in a context where
496 muscle growth seems to be associated with high FE animals, this is a hormone that worth
497 further investigation.

498 From the 13 regulator genes that are DE between groups, six are involved in
499 respiratory chain and are up-regulated in high FE group. Genes *ND1, ND4, ND4L, ND5, ND6*
500 and *ND2*, which is DE but not identified as key regulator, are core subunits of the
501 mitochondrial membrane respiratory chain Complex I (CI) which functions in the transfer of
502 electrons from NADH to the respiratory chain, while *ATP8* is part of Complex V and
503 produces ATP from ADP in the presence of the proton gradient across the membrane.
504 Interestingly, greater quantity of mitochondrial CI protein were associated with high FE cattle
505 by Ramos and Kerley (2013) whereas Davis et al. (2016) found higher CI-CII and CI-CIII
506 concentration ratios for the same group. Other studies demonstrated high FE animals
507 consume less oxygen (Chaves *et al*, 2015) and present lower plasma CO2 concentrations,
508 which suggests a decreased oxidation process (Gonano *et al*, 2014). In general, the literature
509 suggests mitochondrial ADP has greater control of oxidative phosphorylation in high FE
510 individuals (Lancaster *et al*, 2014) and their increased mitochondrial function may contribute
511 to feed efficiency (Connor *et al*, 2010). In pigs, differences in mitochondrial function were
512 reported when analysing muscle (Vincent *et al*, 2015), blood (Liu *et al*, 2016) and adipose
513 tissue transcriptomes (Louveau *et al*, 2016). Differences in metabolic rate associated with FE
514 has long been discussed (Herd & Arthur, 2009) and here is corroborated by the up-regulation
515 of *TSHB* in high FE animals, which stimulates production of T3 and T4 in thyroid thus
516 increasing metabolism. It is inhibited by *SST*, a down-regulated hormone in this group which
517 was also found to be differentially connected between high and low FE.

518 Looking at the DE genes, many hormones can be identified. Hormones are signalling
519 proteins that are transported by the circulatory system to target distant organs in order to
520 regulate physiology. Regarding the relationship between FE and other production traits of
521 economic importance, *FSHB*, responsible for spermatozoa production by activating Sertoli
522 cells in the testicles (Walker & Cheng, 2005), is up-regulated in low FE group and is
523 inhibited by follistatin (*FST*), a gene found to be down-regulated in the same group.
524 Moreover, in rats, it was already demonstrated *FSH* secretion is stimulated by somatostatin
525 expression, which is up-regulated in low FE animals (Kitaoka *et al*, 1989). In this scenario,
526 one could argue that selection for high FE delay reproduction traits, something that could be
527 related to the lower fat deposition in this group, as previously observed (Alexandre *et al*,
528 2015; Santana *et al*, 2012; Gomes *et al*, 2012). Indeed, differences in body composition and
529 in intermediary metabolism can impact on reproductive traits (Shaffer *et al*, 2011) and it has
530 been observed before that feed efficient bulls present features of delayed sexual maturity, i.e.
531 decreased progressive motility of the sperm and higher abundance of tail abnormalities
532 (Montanholi *et al*, 2016; Fontoura *et al*, 2016). Moreover, high FE heifers presented less fat
533 deposition and later sexual maturity which results in calving later in the calving season than
534 their low FE counterparts (Randel & Welsh, 2013; Shaffer *et al*, 2011). It is important to
535 point that low FE animals also present down-regulation of *AMH* and the fall of this hormone
536 in serum was pointed as an excellent marker of Sertoli cells pubertal development (Rey *et al*,
537 1993).

538 Concerning the differences in lipid metabolism in divergent FE phenotypes, *FGF21*, a
539 hormone up-regulated in liver of high FE animals, is associated in humans to decrease in
540 body weight, blood triglycerides and LDL-cholesterol, with improvement in insulin
541 sensitivity (Cheung & Deng, 2014). It is an hepatokine released to the bloodstream and an
542 important regulator of lipid and glucose metabolism (Giralt *et al*, 2015). When we select its

543 first neighbours in the network and perform an enrichment analysis we indeed found terms
544 related to plasma lipoprotein particle remodelling, regulation of lipoprotein oxidation and
545 cholesterol efflux mostly due to *FGF21* co-expression with the apolipoproteins *APOA4*,
546 *APOC3* and *APOM*. In the same context, pro-melanin-concentrating hormone (*PMCH*)
547 encodes three neuropeptides: neuropeptide-glycine-glutamic acid, neuropeptide-glutamic
548 acid-isoleucine and melanin-concentrating hormone (*MCH*) being the last one the most
549 extensively studied (Helgeson & Schmutz, 2008). *MCH* up-regulation has been related to
550 obesity and insulin resistance, as well as increased appetite and reduced metabolism in
551 murine models (Ludwig *et al*, 2001; Ito *et al*, 2003). *PMCH* gene is up-regulated in low FE
552 animals and harbour SNPs found to be associated with higher carcass fat levels and marbling
553 score (Walter *et al*, 2014; Helgeson & Schmutz, 2008).

554 In this work, we were able to identify several biological processes known to be related
555 to feed efficiency, which together with the validation of the main transcription factors of the
556 network, demonstrate the quality of the data and the robustness of the analyses, giving us the
557 confidence to indicate candidate genes to be regulators or biomarkers of superior animals for
558 this trait. The transcription factors NR2F6 and TGFB1 play central roles in liver and muscle,
559 respectively, by regulating genes related to inflammatory response and muscle development
560 and growth, two main biological mechanisms associated to feed efficiency. Likewise,
561 hormones and other proteins secreted in plasma as oxytocin, adrenomedulin, TSH,
562 somatostatin, follistatin and AMH are interesting molecules to be explored as potential
563 biomarkers of feed efficiency.

564

565 **Material and methods**

566 **Phenotypic data and biological sample collection**

All animal protocols were approved by the Institutional Animal Care and Use Committee of Faculty of Food Engineering and Animal Sciences, University of São Paulo (FZEA-USP – protocol number 14.1.636.74.1). All procedures to collect phenotypes and biological samples were carried out at FZEA-USP, Pirassununga, State of São Paulo, Brazil. Ninety eight Nellore bulls (16 to 20 months old and 376 ± 29 kg BW) were evaluated in a feeding trial comprised of 21 days of adaptation to feedlot diet and place and a 70-day period of data collection. Total mixed ration was offered *ad libitum* and daily dry matter intake (DMI) was individually measured. Animals were weighted at the beginning, at the end and every 2 weeks during the experimental period. Feed efficiency was estimated by residual feed intake (RFI) which is the residual of the linear regression that estimates DMI based on average daily gain and mid-test metabolic body weight (Koch *et al*, 1963). Forty animals selected either as high feed efficiency (HFE) or low feed efficiency (LFE) groups were slaughtered on two days with a 6-day interval. Adrenal gland, hypothalamus, liver, muscle and pituitary samples were collected from each animal, rapidly frozen in liquid nitrogen and stored at -80 °C. Further information about management and phenotypic measures of the animals used in this study can be found in Alexandre et al. (2015).

RNAseq data generation

Samples of nine animals from each feed efficiency group (high and low) were selected for RNAseq using RFI measure. For hypothalamus and pituitary, the nitrogen frozen tissue was macerated with crucible and pistil and stored in aliquots at -80 °C. Then, RNA was extracted using AllPrep DNA/RNA/Protein Mini kit (QIAGEN, Crawley, UK). For liver, muscle and adrenal gland, a cut was made in the frozen tissue and the RNA was extracted using RNeasy Mini Kit (QIAGEN, Crawley, UK). RNA quality and quantity were assessed using automated capillary gel electrophoresis on a Bioanalyzer 2100 with RNA 6000 Nano

592 Labchips according to the manufacturer's instructions (Agilent Technologies Ireland, Dublin,
593 Ireland). Samples that presented an RNA integrity number (RIN) less than 8.0 were
594 discarded.

595 RNA libraries were constructed using the TruSeq™ Stranded mRNA LT Sample Prep
596 Protocol and sequenced on Illumina HiSeq 2500 equipment in a HiSeq Flow Cell v4 using
597 HiSeq SBS Kit v4 (2x100pb). Liver, pituitary and hypothalamus were sequenced on the same
598 run, each one in a different lane. Muscle and adrenal gland were sequenced in a second run,
599 in different lanes.

600

601 **Gene expression estimation**

602 The quality of the sequencing was evaluated using the software FastQC
603 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence alignment against the
604 bovine reference genome (UMD3.1) was performed using STAR (Dobin *et al*, 2013),
605 according to the standard parameters and including the annotation file (Ensembl release 89)
606 and secondary alignments, duplicated reads and reads failing vendor quality checks were
607 removed using Samtools (Li *et al*, 2009). Then, HTseq (Anders *et al*, 2014) was used to
608 generate gene read counts and expression values were estimated by reads per kilobase of gene
609 per million mapped reads (RPKM). Genes with average value lower than 0.2 FPKM across
610 all samples and tissues were discarded.

611 Gene expression normalization was performed using the following mixed effect
612 model (Reverter *et al*, 2005):

$$Y_{ijkl} = \mu + L_i + G_j + GT_{jk} + GP_{jl} + e_{ijkl}$$

613 where the log2-transformed FPKM value for i-th library (86 levels), j-th gene (17,354 levels),
614 k-th tissue (5 levels), l-th RFI phenotype (2 levels), corresponding to Y_{ijkl}, was modelled as a
615 function of the fixed effect of library (L_i) and the random effects of gene (G_j), gene by tissue

616 (GT_{jk}) and gene by RFI phenotype (GP_{jl}). Random residual (e_{ijkl}) was assumed to be
617 independent and identically distributed. Variance component estimates and solutions to the
618 model were obtained using VCE6 (Eildert Groeneveld, Milena Kovac and Norbert Mielenz,
619 ftp://ftp.tzv.fal.de/pub/vce6/doc/vce6-manual-3.1-A4.pdf). Normalized mean expression
620 (NME) values for each gene were defined as the linear combination of the solutions for
621 random effects.

622 The mixed model used to normalize the expression data explained 96% of the
623 variation in gene expression, of which the largest proportion (0.30) was due to tissue-
624 specificity. Contrariwise, differences between high and low FE represented no variation
625 (0.27E-11). For that reason, normalized mean expression (NME) were only used to identify
626 tissue specific genes and the raw FPKM values were used for differential expression and co-
627 expression analysis.

628

629 **Gene selection for network construction**

630 In order to select a set of relevant genes for network analysis, we defined five
631 categories based on the following inclusion criteria:

632 1. Differential expression (DE) - The mean expression value of each gene, for
633 each group (high and low FE) and each tissue was calculated and then the expression of low
634 FE group was subtracted from the expression in high FE group. Next, genes were ranked
635 according to their mean expression in all samples for each tissue and divided in five bins.
636 Genes were considered differentially expressed when the difference between the expression
637 in high and low FE groups were greater than 3.1 or smaller than -3.1 standard deviation from
638 the mean in each bin, corresponding to a t-test $P<0.001$.

639 2. Harbouring SNPs - Genes harbouring SNPs associated with feed efficiency,
640 mainly indicated by GWAS, were identified using PubMed database

641 (www.ncbi.nlm.nih.gov/pubmed/) and AnimalQTL database (www.animalgenome.org/cgi-
642 bin/QTLdb/index) and only bovine data were considered regardless of breed.

643 3. Tissue specific (TS) - A gene was considered as tissue specific when the
644 average NME in that tissue was greater than one standard deviation from the mean of all
645 genes AND the average NME in all the other four tissues was smaller than zero.

646 4. Secreted - The human secretome database
647 (www.proteinatlas.org/humanproteome/secretome) was used to select genes encoding
648 proteins secreted in plasma by any of the analysed tissues (adrenal gland, hypothalamus,
649 liver, muscle and pituitary).

650 5. Key regulators - In order to identify key regulatory genes to be included in the
651 co-expression network, a list of genes were obtained from the Animal Transcription Factor
652 Database (http://www.bioguo.org/AnimalTFDB/) and it was compared to a set of potential
653 target genes in each tissue, composed by the categories: TS, DE, harbouring SNPs and
654 secreted. The analysis was based on regulatory impact factor metrics (Reverter *et al*, 2010),
655 which comprises a set of two metrics designed to assign scores to regulator genes consistently
656 differentially co-expressed with target genes and to those with the most altered ability to
657 predict the abundance of target genes. Those scores deviating ±1.96 standard deviation from
658 the mean (corresponding to $P<0.05$) were considered significant. Genes presenting mean
659 expression value less than the mean of all genes expressed were not considered in this
660 analysis.

661 Some of the genes selected by the categories above were represented by more than
662 one ensemble ID. Those duplications were removed for further analysis, keeping only the
663 expression value of the most meaningful ensemble ID. Additionally, genes with mean
664 expression across the samples equal to zero were also removed from further analysis.

665

Co-expression network analysis

For gene network inference, genes selected by the five categories described previously were used as nodes and significant connections (edges) between them were identified using partial correlation and information theory (PCIT) algorithm (Reverter & Chan, 2008), considering all animals and all tissues. PCIT determinates the significance of the correlation between two nodes after accounting for all the other nodes in the network. The output of PCIT was visualized on Cytoscape (Shannon *et al*, 2003).

Network validation through transcription factor biding motifs analysis

Using the regulatory impact factor metric (RIF) we prioritize key regulator genes from gene expression data and predict target genes based on co-expression network (PCIT). In order to assess whether those target genes were enriched for motifs associated to the top most connected regulators in the network with a DNA biding domain (transcription factors - TF), we performed motif discovery analysis in the set of co-expressed target genes (first neighbours of the TF) using i-cistarget method (Herrmann *et al*, 2012) and i-Regulon, a Cytoscape plug-in (Janky *et al*, 2014). These tools use human (hg19) as the reference species, therefore only genes with human orthology are assessed. Then, to validate the binding of the identified genomic regions by the TFs, we performed a region enrichment analysis across experimentally available TF bound regions from ChiP-seq in cell lines from the ENCODE consortium (1,394 TF binding site tracks). Finally, we converted identified enhancer regions to cow coordinates and searched for regions of open-chromatin using data from a public available studies in cow tissues.

689 **Differential connectivity**

690 In order to explore differentially connected genes between high and low FE, two
691 networks were created, one for each condition, using the same methodology described before.
692 Then, the number of connections of each gene in each condition was computed and scaled so
693 that connectivity varied from 0 to 1, making possible to compare the same gene in the two
694 networks. The connectivity in high RFI group was subtracted from the connectivity in low
695 RFI group and results deviating ±1.96 standard deviation from the mean were considered
696 significant ($P<0.05$).

697

698 **Functional Enrichment**

699 Functional enrichment analysis was performed on the online platform GOrilla (Gene
700 Ontology enRIchment anaLysis and visuaLizAtion tool, http://cbl-gorilla.cs.technion.ac.il/),
701 using all genes that passed FPKM filter as background, hypergeometric test and multiple test
702 correction (FDR - false discovery rate). The human database was used to take advantage of a
703 more comprehensive knowledge regarding gene functions. GO terms were considered
704 significant when $Padj<0.05$. For genes in co-expression networks, visualized using Cytoscape
705 (Shannon *et al*, 2003), the functional enrichment was performed with BiNGO plug-in (Maere
706 *et al*, 2005) using the same background genes and statistical test.

707

708 **Acknowledgments**

709 The authors thank Mrs. Elisângela C. M. Oliveira for all technical support. This study
710 and PAA scholarships were funded by São Paulo Research Foundation (FAPESP) - Proc.
711 2014/07566-2, 2015/22276-3 and 2017/14707-0. M.N.S was funded by the CSIRO Science
712 Excellence Research Office.

713

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

HF was the overall project leader who conceived this study and supervised PAA in all data generation. TR performed the network analysis and supervised PAA in bioinformatics and data interpretation. MNS performed the motif discovery analysis. LRPN and JBSF provided informatics and statistical support. All authors contributed to and approved the final version of this manuscript.

Availability of supporting data

Datasets supporting the results of this article is public available in the European Nucleotide Archive (ENA) as part of FAANG consortium under de study ID PRJEB27337 and can be accessed following the link https://www.ebi.ac.uk/ena/data/view/PRJEB27337.

References

Aerts S, Quan XJ, Claeys A, Sanchez MN, Tate P, Yan J & Hassan BA (2010) Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in drosophila uncovers a regulatory basis for sensory specification. *PLoS Biol.* **8:**

Alexandre P a., Kogelman LJ a., Santana MH a., Passarelli D, Pulz LH, Fantinato-Neto P, Silva PL, Leme PR, Strefezzi RF, Coutinho LL, Ferraz JBS, Eler JP, Kadarmideen HN & Fukumasu H (2015) Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics* **16:** 1073 Available at: http://www.biomedcentral.com/1471-2164/16/1073

Anders S, Pyl PT & Huber W (2014) HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166–9 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4287950&tool=pmcentrez&rendertype=abstract [Accessed September 29, 2014]

Archer JA, Arthur PF, Herd RM, Parnell PF & Pitchford WS (1997) Optimum postweaning test for measurement of growth rate, feed intake, and feed efficiency in British breed cattle. *J. Anim. Sci.* **75:** 2024–32 Available at: http://www.ncbi.nlm.nih.gov/pubmed/9263047 [Accessed September 5, 2014]

Archer JA, Richardson EC, Herd RM & Arthur PF (1999) Potential for selection to improve

efficiency of feed use in beef cattle: a review. *Aust. J. Agric. Res.* **50:** 147–61 Available at: http://www.publish.csiro.au/paper/A98075.htm [Accessed February 29, 2016]

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM & Goddard ME (2011) Genome-wide association studies for feedlot and growth traits in cattle. *J. Anim. Sci.* **89:** 1684–97 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21239664 [Accessed March 21, 2012]

Cafe LM, Robinson DL, Ferguson DM, Geesink GH & Greenwood PL (2011) Temperament and hypothalamic-pituitary-adrenal axis function are related and combine to affect growth, efficiency, carcass, and meat quality traits in Brahman steers. *Domest. Anim. Endocrinol.* **40:** 230–40 Available at: http://www.sciencedirect.com/science/article/pii/S0739724011000063 [Accessed February 21, 2016]

Cao P, Hanai J-I, Tanksale P, Imamura S, Sukhatme VP & Lecker SH (2009) Statin-induced muscle damage and atrogin-1 induction is the result of a geranylgeranylation defect. *FASEB J.* **23:** 2844–2854 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19406843%5Cnhttp://www.fasebj.org/content/23/9/2844.full.pdf

Chaves AS, Nascimento ML, Tullio RR, Rosa AN, Alencar MM & Lanna DP (2015) Relationship of efficiency indices with performance, heart rate, oxygen consumption, blood parameters, and estimated heat production in Nellore steers. *J. Anim. Sci.* **93:** 5036–46 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26523596 [Accessed February 28, 2016]

Chen L, Mao F, Crews DH, Vinsky M & Li C (2014) Phenotypic and genetic relationships of feeding behavior with feed intake, growth performance, feed efficiency, and carcass merit traits in Angus and Charolais steers. *J. Anim. Sci.* **92:** 974–83 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24492561 [Accessed February 21, 2016]

Cheung BM & Deng HB (2014) Fibroblast growth factor 21: a promising therapeutic target in obesity-related diseases. *Expert Rev Cardiovasc Ther* **12:** 659–666

Connor EE, Kahl S, Elsasser TH, Parker JS, Li RW, Van Tassell CP, Baldwin RL & Barao SM (2010) Enhanced mitochondrial complex gene function and reduced liver size may mediate improved feed efficiency of beef cattle during compensatory growth. *Funct. Integr. Genomics* **10:** 39–51 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19777276 [Accessed February 28, 2016]

Davis MP, Brooks MA & Kerley MS (2016) Relationship between residual feed intake and lymphocyte mitochondrial complex protein concentration and ratio in crossbred steers. *J. Anim. Sci.* **94:** 1587–1591

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M & Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23104886 [Accessed February 17, 2017]

Fontoura ABP, Montanholi YR, Diel de Amorim M, Foster RA, Chenier T & Miller SP (2016) Associations between feed efficiency, sexual maturity and fertility-related measures in young beef bulls. *Animal* **10:** 96–105 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26351012 [Accessed March 2, 2016]

Foote AP, Tait RG, Keisler DH, Hales KE & Freetly HC (2016) Leptin concentrations in finishing beef steers and heifers and their association with dry matter intake, average daily gain, feed efficiency, and body composition. *Domest. Anim. Endocrinol.* **55:** 136–41 Available at: http://www.sciencedirect.com/science/article/pii/S0739724016000035 [Accessed April 15, 2016]

Francisco CL, Resende FD, Benatti JMB, Castilhos AM, Cooke RF & Jorge AM (2015)

798 Impacts of temperament on Nellore cattle: physiological responses, feedlot performance,
799 and carcass characteristics. *J. Anim. Sci.* **93:** 5419 Available at:
800 http://www.ncbi.nlm.nih.gov/pubmed/26641061 [Accessed February 21, 2016]
801 Gerber PJ, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkman J, Falcucci A & Tempio G
802 (2013) Tackling Climate Change Through Livestock - A Global Assessment of
803 Emissions and Mitigation Opportunities Rome Available at:
804 http://www.fao.org/docrep/018/i3437e/i3437e.pdf
805 Giralt M, Gavaldà-Navarro A & Villarroya F (2015) Fibroblast growth factor-21, energy
806 balance and obesity. *Mol. Cell. Endocrinol.* **418:** 66–73
807 Gomes RC, Sainz RD, Silva SL, César MC, Bonin MN & Leme PR (2012) Feedlot
808 performance, feed efficiency reranking, carcass traits, body composition, energy
809 requirements, meat quality and calpain system activity in Nellore steers with low and
810 high residual feed intake. *Livest. Sci.* **150:** 265–273 Available at:
811 http://www.sciencedirect.com/science/article/pii/S1871141312003587 [Accessed
812 February 19, 2014]
813 Gonano C, Montanholi Y, Schenkel F, Smith B, Cant J & Miller S (2014) The relationship
814 between feed efficiency and the circadian profile of blood plasma analytes measured in
815 beef heifers at different physiological stages. *Animal* **8:** 1684–1698 Available at:
816 http://www.ncbi.nlm.nih.gov/pubmed/24923431 [Accessed February 28, 2016]
817 Gondret F, Vincent A, Houée-Bigot M, Siegel A, Lagarrigue S, Causeur D, Gilbert H &
818 Louveau I (2017) A transcriptome multi-tissue analysis identifies biological pathways
819 and genes associated with variations in feed efficiency of growing pigs. *BMC Genomics*
820 **18:** 1–17
821 Hayes BJ, Lewin H a & Goddard ME (2013) The future of livestock breeding: genomic
822 selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* **29:**
823 206–14 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23261029 [Accessed
824 October 15, 2014]
825 Helgeson SC & Schmutz SM (2008) Genetic variation in the pro-melanin-concentrating
826 hormone gene affects carcass traits in Bos taurus cattle. *Anim. Genet.* **39:** 310–315
827 Herd RM & Arthur PF (2009) Physiological basis for residual feed intake. *J. Anim. Sci.* **87:**
828 E64-71 Available at: http://jas.fass.org/cgi/content/abstract/87/14_suppl/E64 [Accessed
829 March 28, 2012]
830 Herd RM, Oddy VH & Richardson EC (2004) Biological basis for variation in residual feed
831 intake in beef cattle. 1. Review of potential mechanisms. *Aust. J. Exp. Agric.* **44:** 423
832 Available at:
833 http://www.publish.csiro.au/view/journals/dsp_journal_fulltext.cfm?nid=72&f=EA0222
834 0 [Accessed April 18, 2012]
835 Hermann-Kleiter N & Baier G (2014) Orphan nuclear receptor NR2F6 acts as an essential
836 gatekeeper of Th17 CD4+ T cell effector functions. *Cell Commun. Signal.* **12:** 1–12
837 Hermann-Kleiter N, Gruber T, Lutz-Nicoladoni C, Thuille N, Fresser F, Labi V,
838 Schiefermeier N, Warnecke M, Huber L, Villunger A, Eichele G, Kaminski S & Baier G
839 (2008) The Nuclear Orphan Receptor NR2F6 Suppresses Lymphocyte Activation and T
840 Helper 17-Dependent Autoimmunity. *Immunity* **29:** 205–216
841 Herrmann C, Van De Sande B, Potier D & Aerts S (2012) i-cisTarget: An integrative
842 genomics method for the prediction of regulatory features and cis-regulatory modules.
843 *Nucleic Acids Res.* **40:**
844 Ito M, Gomori A, Ishihara A, Oda Z, Mashiko S, Matsushita H, Yumoto M, Ito M, Sano H,
845 Tokita S, Moriya M, Iwaasa H & Kanatani A (2003) Characterization of MCH-mediated
846 obesity in mice. *Am. J. Physiol. - Endocrinol. Metab.* **284:** E940–E945
847 Jager N De, Hudson NJ, Reverter A, Wang Y, Nagaraj SH, Cafe LM, Greenwood PL,

Barnard RT, Kongsuwan KP, Dalrymple BP, Pl G, Rt B, Kp K & Bp D (2011) Chronic exposure to anabolic steroids induces the muscle expression of oxytocin and a more than fiftyfold increase in circulating oxytocin in cattle. **:** 467–478

Janky R, Verfaillie A, Imrichová H, van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine JC & Aerts S (2014) iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput. Biol.* **10:**

Jing L, Hou Y, Wu H, Miao Y, Li X, Cao J, Brameld JM, Parr T & Zhao S (2015) Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Sci. Rep.* **5:** 11953 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4493709&tool=pmcentrez&rendertype=abstract [Accessed April 6, 2016]

Karisa B, Moore S & Plastow G (2014) Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. *Anim. Sci. J.* **85:** 374–87 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24373146 [Accessed February 1, 2015]

Kelly AK, McGee M, Crews DH, Fahey AG, Wylie AR & Kenny DA (2010) Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. *J. Anim. Sci.* **88:** 109–23 Available at: http://www.ncbi.nlm.nih.gov/pubmed/19820067 [Accessed February 21, 2016]

Kitaoka M, Takano K, Kojima I & Ogata E (1989) A stimulatory effect of somatostatin: Enhancement of activin A-mediated FSH secretion in rat pituitary cells. *Biochem. Biophys. Res. Commun.* **162:** 958–962

Klepsch V, Hermann-Kleiter N & Baier G (2016) Beyond CTLA-4 and PD-1: Orphan nuclear receptor NR2F6 as T cell signaling switch and emerging target in cancer immunotherapy. *Immunol. Lett.* **178:** 31–36

Koch RM, Swiger LA, Chambers D & Gregory KE (1963) Efficiency of Feed Use in Beef Cattle. *J. Anim. Sci.* **22:** 486–494

de la Fuente A (2010) From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26:** 326–33 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20570387 [Accessed July 18, 2014]

Lancaster PA, Carstens GE, Michal JJ, Brennan KM, Johnson KA & Davis ME (2014) Relationships between residual feed intake and hepatic mitochondrial function in growing beef cattle. *J. Anim. Sci.* **92:** 3134–41 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24894006 [Accessed February 29, 2016]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G & Durbin R (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–9 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723002&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014]

Liu D, Black BL & Derynck R (2001) TGF-beta inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. *Genes Dev.* **15:** 2950–2966

Liu H, Nguyen YT, Nettleton D, Dekkers JCM & Tuggle CK (2016) Post-weaning blood transcriptomic differences between Yorkshire pigs divergently selected for residual feed intake. *BMC Genomics* **17:** 73 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4724083&tool=pmcentrez&rendertype=abstract [Accessed February 1, 2016]

Louveau I, Vincent A, Tacher S, Gilbert H & Gondret F (2016) Increased expressions of

genes and proteins involved in mitochondrial oxidation and antioxidant pathway in adipose tissue of pigs selected for a low residual feed intake. *J. Anim. Sci.* **94:** 5042–5054

Ludwig DS, Tritos NA, Mastaitis JW, Kulkarni R, Kokkotou E, Elmquist J, Lowell B, Flier JS & Maratos-Flier E (2001) Melanin-concentrating hormone overexpression in transgenic mice leads to obesity and insulin resistance. *J. Clin. Invest.* **107:** 379–386

Mader CJ, Montanholi YR, Wang YJ, Miller SP, Mandell IB, McBride BW & Swanson KC (2009) Relationships among measures of growth performance and efficiency with carcass traits, visceral organ mass, and pancreatic digestive enzymes in feedlot cattle. *J. Anim. Sci.* **87:** 1548–57 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18952722 [Accessed January 30, 2015]

Maere S, Heymans K & Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **21:** 3448–3449

Mani V, Harris AJ, Keating AF, Weber TE, Dekkers JCM & Gabler NK (2013) Intestinal integrity, endotoxin transport and detoxification in pigs divergently selected for residual feed intake. *J. Anim. Sci.* **91:** 2141–50 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23463550 [Accessed March 1, 2016]

Montanholi YR, Fontoura ABP, Diel de Amorim M, Foster RA, Chenier T & Miller SP (2016) Seminal plasma protein concentrations vary with feed efficiency and fertility-related measures in young beef bulls. *Reprod. Biol.* **16:** 147–156

Montanholi YR, Swanson KC, Palme R, Schenkel FS, McBride BW, Lu D & Miller SP (2010) Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. *Animal* **4:** 692–701 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22444121 [Accessed February 29, 2016]

Montanholi YR, Swanson KC, Schenkel FS, McBride BW, Caldwell TR & Miller SP (2009) On the determination of residual feed intake and associations of infrared thermography with efficiency and ultrasound traits in beef bulls. *Livest. Sci.* **125:** 22–30 Available at: http://www.sciencedirect.com/science/article/pii/S1871141309000766 [Accessed February 29, 2016]

Mota LFM, Bonafé CM, Alexandre PA, Santana MH, Novais FJ, Toriyama E, Pires AV, da Luz Silva S, Leme PR, Ferraz JBS & Fukumasu H (2017) Circulating leptin and its muscle gene expression in Nellore cattle with divergent feed efficiency. *J. Anim. Sci. Biotechnol.* **8:**

Mukiibi R, Vinsky M, Keogh KA, Fitzsimmons C, Stothard P, Waters SM & Li C (2018) Transcriptome analyses reveal reduced hepatic lipid synthesis and accumulation in more feed efficient beef cattle. **:** 1–12

Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, Dekoter RP & Young RA (2011) Master transcription factors determine cell-type-specific responses to TGF-β signaling. *Cell* **147:** 565–576

Myer PR, Smith TPL, Wells JE, Kuehn LA & Freetly HC (2015) Rumen microbiome from steers differing in feed efficiency. *PLoS One* **10:** e0129174 Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4451142&tool=pmcentrez&rendertype=abstract [Accessed March 1, 2016]

Myer PR, Wells JE, Smith TPL, Kuehn LA & Freetly HC (2016) Microbial community profiles of the jejunum from steers differing in feed efficiency. *J. Anim. Sci.* **94:** 327 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26812338 [Accessed March 27, 2017]

Naval-Sańchez M, Potier D, Haagen L, Sánchez M, Munck S, Van De Sande B, Casares F, Christiaens V & Aerts S (2013) Comparative motif discovery combined with

948 comparative transcriptomics yields accurate targetome and enhancer predictions.
949 *Genome Res.* **23:** 74–88
950 Nkrumah JD, Keisler DH, Crews DH, Basarab JA, Wang Z, Li C, Price MA, Okine EK &
951 Moore SS (2007) Genetic and phenotypic relationships of serum leptin concentration
952 with performance, efficiency of gain, and carcass merit of feedlot cattle. *J. Anim. Sci.*
953 **85:** 2147–55 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17468416 [Accessed
954 February 23, 2016]
955 Novais FJ, Pires PRL, Alexandre PA, Dromms RA, Iglesias AH, Ferraz JBS, Styczynski MP
956 & Fukumasu H (2018) Serum metabolomics predicts biological processes associated to
957 feed efficiency in beef cattle [under review]. *Sci. Rep.*
958 de Oliveira PSN, Cesar ASM, do Nascimento ML, Chaves AS, Tizioto PC, Tullio RR, Lanna
959 DPD, Rosa AN, Sonstegard TS, Mourao GB, Reecy JM, Garrick DJ, Mudadu MA,
960 Coutinho LL & Regitano LCA (2014) Identification of genomic regions associated with
961 feed efficiency in Nelore cattle. *BMC Genet.* **15:** 100 Available at:
962 http://www.biomedcentral.com/1471-2156/15/100 [Accessed December 27, 2014]
963 Paradis F, Yue S, Grant JR, Stothard P, Basarab JA & Fitzsimmons C (2015) Transcriptomic
964 analysis by RNA sequencing reveals that hepatic interferon-induced genes may be
965 associated with feed efficiency in beef heifers. *J. Anim. Sci.* **93:** 3331 Available at:
966 http://www.ncbi.nlm.nih.gov/pubmed/26440002 [Accessed November 8, 2015]
967 Pereira FA, Tsai MJ & Tsai SY (2000) COUP-TF orphan nuclear receptors in development
968 and differentiation. *Cell Mol Life Sci* **57:** 1388–1398
969 Potier D, Davie K, Hulselmans G, NavalSanchez M, Haagen L, Huynh-Thu VA, Koldere D,
970 Celik A, Geurts P, Christiaens V & Aerts S (2014) Mapping Gene Regulatory Networks
971 in Drosophila Eye Development by Large-Scale Transcriptome Perturbations and Motif
972 Inference. *Cell Rep.* **9:** 2290–2303
973 Ramayo-Caldas Y, Ballester M, Sánchez JP, González-Rodríguez O, Revilla M, Reyer H,
974 Wimmers K, Torrallardona D & Quintanilla R (2018) Integrative approach using liver
975 and duodenum RNA-Seq data identifies candidate genes and pathways associated with
976 feed efficiency in pigs. *Sci. Rep.* **8:**
977 Ramos MH & Kerley MS (2013) Mitochondrial complex I protein differs among residual
978 feed intake phenotype in beef cattle. *J. Anim. Sci.* **91:** 3299–304 Available at:
979 http://www.ncbi.nlm.nih.gov/pubmed/23798519 [Accessed February 29, 2016]
980 Randel RD & Welsh TH (2013) Joint Alpharma-Beef Species Symposium: interactions of
981 feed efficiency with beef heifer reproductive development. *J. Anim. Sci.* **91:** 1323–8
982 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23048157 [Accessed March 2, 2016]
983 Reverter A, Barris W, McWilliam S, Byrne KA, Wang YH, Tan SH, Hudson N & Dalrymple
984 BP (2005) Validation of alternative methods of data normalization in gene co-expression
985 studies. *Bioinformatics* **21:** 1112–1120
986 Reverter A & Chan EKF (2008) Combining partial correlation and an information theory
987 approach to the reversed engineering of gene co-expression networks. *Bioinformatics*
988 **24:** 2491–2497 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18784117 [Accessed
989 June 11, 2017]
990 Reverter A, Hudson NJ, Nagaraj SH, Pérez-Enciso M & Dalrymple BP (2010) Regulatory
991 impact factors: Unraveling the transcriptional regulation of complex traits from
992 expression data. *Bioinformatics* **26:** 896–904
993 Rey R, Lordereau-Richard I, Carel JC, Barbet P, Cate RL, Roger M, Chaussain JL & Josso N
994 (1993) Anti-müllerian hormone and testosterone serum levels are inversely during
995 normal and precocious pubertal development. *J. Clin. Endocrinol. Metab.* **77:** 1220–
996 1226
997 Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL, Kerley MS &

Weaber RL (2011) Genome-wide association analysis for feed efficiency in Angus cattle. *Anim. Genet.* Available at: http://www.ncbi.nlm.nih.gov/pubmed/22497295 [Accessed April 24, 2012]

Santana MHA, Rossi P, Almeida R & Cucco DC (2012) Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. *Livest. Sci.* **145:** 252–257 Available at: http://linkinghub.elsevier.com/retrieve/pii/S1871141312000686 [Accessed January 13, 2015]

Santana MH, Utsunomiya YT, Neves HH, Gomes RC, Garcia JF, Fukumasu H, Silva SL, Oliveira Junior G a, Alexandre P a, Leme PR, Brassaloti R a, Coutinho LL, Lopes TG, Meirelles F V, Eler JP & Ferraz JB (2014) Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. *BMC Genet.* **15:** 21 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24517472 [Accessed February 19, 2014]

Schmierer B & Hill CS (2007) TGFβ-SMAD signal transduction: Molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* **8:** 970–982

Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freetly HC, Hansen SL, Yampara-Iquise H, Johnson KA, Kerley MS, Kim JW, Loy DD, Marques E, Neibergs HL, Schnabel RD, Shike DW, Spangler ML, et al (2017) Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* **18:**

Shaffer KS, Turk P, Wagner WR & Felton EED (2011) Residual feed intake, body composition, and fertility in yearling beef heifers. *J. Anim. Sci.* **89:** 1028–34 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21112981 [Accessed March 2, 2016]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13:** 2498–2504

Shin JH, Li RW, Gao Y, Baldwin VI R & Li CJ (2012) Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. *Funct. Integr. Genomics* **12:** 119–130

Shindo T, Kurihara H, Kurihara Y, Morita H & Yazaki Y (1998) Upregulation of endothelin-1 and adrenomedullin gene expression in the mouse endotoxin shock model. *J Cardiovasc Pharmacol* **31 Suppl 1:** S541-4

Shoji H, Minamino N, Kangawa K & Matsuo H (1995) Endotoxin markedly elevates plasma concentration and gene transcription of adrenomedullin in rat. *Biochem. Biophys. Res. Commun.* **215:** 531–537

Swami M (2009) Networking complex traits. *Nat. Rev. Genet.* **10:** 2566

Tronche F & Yaniv M (1992) HNF1, a homeoprotein member of the hepatic transcription regulatory network. *BioEssays* **14:** 579–587

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JMA, Bertelsen MF, Murchison EP, Flicek P & Odom DT (2015) Enhancer evolution across 20 mammalian species. *Cell* **160:** 554–566

Vincent A, Louveau I, Gondret F, Tréfeu C, Gilbert H & Lefaucheur L (2015) Divergent selection for residual feed intake affects the transcriptomic and proteomic profiles of pig skeletal muscle. *J. Anim. Sci.* **93:** 2745–58 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26115262 [Accessed February 29, 2016]

Walker WH & Cheng J (2005) FSH and testosterone signaling in Sertoli cells. *Reproduction* **130:** 15–28

Walter LJ, Gasch CA, Mcevers TJ, Hutcheson JP, Defoor P, Marquess FLS & Lawrence TE (2014) Association of pro-melanin concentrating hormone genotype with beef carcass quality and yield. *J. Anim. Sci.* **92:** 325–331

Weber KL, Welly BT, Van Eenennaam AL, Young AE, Port-Neto LR, Reverter A & Rincon

1048 G (2016) Identification of Gene networks for residual feed intake in Angus cattle using
1049 genomic prediction and RNA-seq. *PLoS One* **11:** 1–19
1050 Widmann P, Reverter A, Weikard R, Suhre K, Hammon HM, Albrecht E & Kuehn C (2015)
1051 Systems biology analysis merging phenotype, metabolomic and genomic data identifies
1052 Non-SMC Condensin I Complex, Subunit G (NCAPG) and cellular maintenance
1053 processes as major contributors to genetic variability in Bovine feed efficiency. *PLoS*
1054 *One* **10:** 1–22
1055 Zarek CM, Lindholm-Perry AK, Kuehn LA & Freetly HC (2017) Differential expression of
1056 genes related to gain and intake in the liver of beef cattle. *BMC Res. Notes* **10:** 1
1057 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28057050 [Accessed March 27,
1058 2017]
1059
1060
1061
1062

Supporting data

S1 Supporting Information. Information of reads mapping to bovine genome (UMD3.1) per sample.

S2 Supporting Information. Scatter plots showing differentially expressed genes between high and low feed efficiency (FE) from adrenal gland, hypothalamus, liver, muscle and pituitary. Dots represent the mean expression for a gene in high FE subtracted from the mean expression of the same gene in low FE (M) by the average expression value in both groups (A). Pink dots represent significant genes (P<0.001).

S3 Supporting Information. Differentially expressed genes between high and low feed efficiency in adrenal gland (A), hypothalamus (B), liver (C), muscle (D) and pituitary (E).

S4 Supporting Information. Functional enrichment of the 248 genes down-regulated in high feed efficiency considering the five tissues (adrenal gland, hypothalamus, liver, muscle and pituitary). Colour intensity increase with the significance of the term; white represents $P>10^{-3}$ and the darkest orange represents $P<10^{-9}$.

S5 Supporting Information. Genes selected for network construction for being differentially expressed between high and low feed efficiency (A), harbouring SNPs previourly associated with feed efficiency (B), tissue specific (C), coding proteins secreted in plasma (D) and key regulators (E).

S6 Supporting Information. Functional enrichment for the 135 genes coding proteins secreted in plasma which the tissue of maximum expression is liver. Colour intensity increase with the significance of the term; white represents $P>10^{-3}$ and the darkest orange represents $P<10^{-9}$.

S7 Supporting Information. Genes included in co-expression analysis.

S8 Supporting Information. Top five key regulator genes network and enrichment. A) Network of genes *EPC1, NR2F6, MED21, ENSBTAG00000031687* and *CTBP1* and their first neighbours. Nodes with diamond shape correspond to secreted proteins coding genes and triangles correspond to key regulators; all the other genes are represented by ellipses. Nodes with black borders are differentially expressed between high and low feed efficiency. Colours are relative to the tissue of maximum expression: blue represent liver, red represent muscle, yellow represent pituitary, green represent hypothalamus and orange represent adrenal gland. The size of the nodules is relative to the normalized mean expression values in all samples. Only correlations above 0.9 and bellow -0.9 and its respective genes are shown in this figure. **B) Functional enrichment of the 345 genes in the network (A).** Colour intensity increase with the significance of the term; white represents $P>5x10^{-3}$.

S9 Supporting Information. *TGFB1* network and enrichment. A) Network of *TGFB1* gene and its first neighbours. Nodes with diamond shape correspond to secreted proteins coding genes and triangles correspond to key regulators; all the other genes are represented by ellipses. Nodes with black borders are differentially expressed between high and low feed efficiency. Colours are relative to the tissue of maximum expression: blue represent liver, red represent muscle, yellow represent pituitary, green represent hypothalamus and orange

represent adrenal gland. The size of the nodules is relative to the normalized mean expression values in all samples. Only correlations above 0.9 and bellow -0.9 and its respective genes are shown in this figure. **B) Functional enrichment of the 157 genes in the network (A).** Colour intensity increase with the significance of the term; white represents $P>5x10^{-3}$.

S10 Supporting Information. *FGF21* network and enrichment. A) Network of *FGF21* gene and its first neighbours. Nodes with diamond shape correspond to secreted proteins coding genes and triangles correspond to key regulators; all the other genes are represented by ellipses. Nodes with black borders are differentially expressed between high and low feed efficiency. Colours are relative to the tissue of maximum expression: blue represent liver, red represent muscle, yellow represent pituitary, green represent hypothalamus and orange represent adrenal gland. The size of the nodules is relative to the normalized mean expression values in all samples. Only correlations above 0.9 and bellow -0.9 and its respective genes are shown in this figure. **B) Functional enrichment of the 98 genes in the network (A).** Colour intensity increase with the significance of the term; white represents $P>5x10^{-3}$.

S11 Supporting Information. NR2F6 i-cis Target results.

S12 Supporting Information. NR2F6 predicted regions binding hg19.

S13 Supporting Information. NR2F6 predcited regions bindg bosTau.

S14 Supporting Information. TGBF1 i-cis Target results.

S15 Supporting Information. MYOD predicted transcription factors binding sites in hg19.

S16 Supporting Information. MYOD predicted transcription factors binding sites in mm8.

S17 Supporting Information. MYOD predicted transcription factors binding sites in mm9.

S18 Supporting Information. MYOD predicted transcription factors binding sites in bosTau6.

S19 Supporting Information. Differentially connected genes between high and low feed efficiency.

S20 Supporting Information. Functional enrichment for the 87 differentially co-expressed genes between high and low feed efficiency. Colour intensity increase with the significance of the term; white represents $P>10^{-3}$ and the darkest orange represents $P<10^{-9}$.