Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

Identificação de sequências *motifs* e elementos regulatórios em ATAC-Seq *peaks* associados à expressão gênica e à deposição de gordura intramuscular em bovinos da raça Nelore

Natália Silva Morosini

Tese apresentada para obtenção do título de Doutora em Ciências. Área de concentração: Ciência Animal e Pastagens

Piracicaba 2020 Natália Silva Morosini Farmacêutica-Bioquímica

Identificação de sequências *motifs* e elementos regulatórios em ATAC-Seq *peaks* associados à expressão gênica e à deposição de gordura intramuscular em bovinos da raça Nelore versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador: Prof. Dr. LUIZ LEHMANN COUTINHO

Tese apresentada para obtenção do título Doutora em Ciências. Área de concentração: Ciência Animal e Pastagens

Piracicaba 2020

Dados Internacionais de Catalogação na Publicação DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Morosini, Natália Silva

Identificação de sequências *motifs* e elementos regulatórios em ATAC-Seq *peaks* associados à expressão gênica e à deposição de gordura intramuscular em bovinos da raça Nelore / Natalia Silva Morosini. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

61 p.

Tese (doutorado) - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Regulação gênica 2. ATAC-Seq 3. Gordura intramuscular 4. Bos indicus I. Título

DEDICATÓRIA

À minha irmã Júlia, meu tesouro e minha maior inspiração, Dedico.

AGRADECIMENTOS

À Deus por ser o responsável por eu chegar até aqui. Obrigada por sempre me acolher e amparar. Foram vários os momentos em que eu rezei baixinho e pedi em silêncio. Em todos o Senhor me atendeu. Deus, o Senhor sempre deu sinais pra me indicar o melhor caminho, me guiou com ternura e olhou por mim com muito zelo. Obrigada por nunca permitir que eu me sentisse sozinha. Obrigada por me fortalecer e me preparar pra todas as maravilhas que se iniciaram com o novo ano.

À Nossa Senhora Aparecida. Essa tese não estaria concluída se não fosse pela Senhora. Eu me lembro de quando precisei do seu amor e da sua benção. Senti sua presença. Obrigada por ter encoberto meus melhores passos pra que eles me guiassem, no momento certo. Obrigada por ter ouvido e atendido minhas preces.

Aos meus pais Olívia e Carlos pela minha família. Obrigada por serem minha base, meu porto seguro. Agradeço por me ensinarem sobre a vida. Vocês nos proporcionam felicidade em sua forma mais pura e genuína. Nossos momentos são únicos e ricos em união. Obrigada por estarem sempre de braços abertos. Sou grata por vocês viverem meus sonhos com tanto positivismo e alegria. Agradeço por terem me passado valores. Seguir com ética e princípios nos faz vencer. Obrigada por nos ensinarem sobre percorrer caminhos e desfrutar de tudo o que eles nos oferecem, como chance de amadurecimento. Obrigada pela resiliência. Todas as minhas conquistas são suas.

A minha irmã Júlia. Você é a autenticidade da excelência. Você desde sempre foi um exemplo a ser seguido. Você é a filha da Gloriosa, o orgulho de todos que te conhecem e que se inspiram no seu caráter e profissionalismo. Você é aquela pessoa que não tem como passar despercebida aos olhos da primazia. Obrigada por sempre me apoiar. Agradeço por caminharmos juntas. Obrigada pela amizade, pelo afeto e por ser quem você é.

Ao meu namorado Lucas Delfini. Obrigada pela cumplicidade, parceria e sonhos. Por ser meu ombro amigo, por crescermos juntos e por compartilharmos todas as conquistas, lado a lado. Agradeço pelo seu companheirismo, atenção e paciência.

Ao meu cunhado Fernando. Obrigada pelos ensinamentos, pelo exemplo e por todo o suporte ao longo da minha caminhada acadêmica.

A Helena, Lucca e Luiz. Obrigada pela confiança. Agradeço por termos nos reencontrado, em um outro momento, em uma nova fase. Obrigada pela referência, pelo Norte, por me darem voz, guiarem meus primeiros passos profissionais e me ensinarem tanto. Vocês são anjos em minha vida.

À Escola Superior de Agricultura "Luiz de Queiroz" pelo exemplo de profissionalismo, por agregar oportunidades e conhecimentos em minha história acadêmica e, principalmente, por ser Gloriosa em sua essência, representação plena de lealdade e compromisso aos ideais científicos educacionais.

Ao meu orientador Luiz Lehmann Coutinho pelo aprendizado, desafios e críticas construtivas.

A todos os Brasileiros que com o dinheiro suado do dia-a-dia financiam todas as pesquisas do país. Vocês são os melhores e maiores investidores de todas as melhorias pensadas, pesquisadas e executadas por nós, pesquisadores. Sem vocês, as inovações não teriam sentido e nem motivos para serem aplicadas. Obrigada.

À USP, minha Universidade amada. Encerro aqui minha trajetória acadêmica com a certeza de que minha vida profissional jamais teria sido a mesma se eu não tivesse tido a honra de ser USPiana. Agradeço pela excelência e majestosidade. Você será pra sempre a melhor da América Latina, o orgulho de quem teve a garra e perseverança de te conquistar. Prometo fazer jus à responsabilidade de te carregar no peito.

Á CAPES pelo auxílio financeiro e confiança em minha pesquisa.

"'Till we find our place On a path unwinding In the circle The circle of life."

(Circle of Life - The Lion King)

SUMÁRIO

RESUMO	9
ABSTRACT	10
1. REVISÃO DE LITERATURA	11
1.1 MATERIAL GENÉTICO	11
1.2. REMODELAGEM DA CROMATINA	11
1.2. Remodeladem da cromativa	12
	12
	15 16
1.5. IDENTIFICAÇÃO DE REGIÕES DE EUCROMATINA	10 16
	10
2.1. HIPOTESE	20
2.2. OBJETIVO	20
2.2.1. Objetivo geral	
2.2.2. Objetivos específicos	
2.2.3. MATERIAIS E METODOS	
2.2.3.1. Fase 1	
2.2.3.1.1. Amostras	21
2.2.3.1.2. Obtenção de regiões de eucromatina	21
2.2.3.1.3. Sequenciamento ATAC-DNA	22
2.2.3.2. Fase 2	22
2.2.3.2.1. Alinhamento contra o genoma de referência	22
2.2.3.2.2. Identificação dos ATAC-Seq peaks	22
2.2.3.2.3. Determinação da localização dos elementos regulatórios em ATA	IC-Seq
peaks	22 ~
2.2.3.2.1. <i>Clusterização</i> de genes com o mesmo perfil de expressão e detecç	ao de
associações de modulos ao fenotipo	23
2.2.3.2.2. Sobreposição entre ATAC-Seq <i>peaks</i> e genes expressos no mu esquelético	sculo 23
2 2 3 2 3 Identificação de seguências motifs em genes co-expressos	23
2 2 3 2 4 Descoberta de cis-actina elements (motifs)	24
2 2 3 2 5 Identificação de motifs em ATAC-Sea neaks	24
2 2 3 2 6 Ocorrância de seguência matifs em hancos de dados	25
2 2 3 2 7 Descrição dos fatores de transcrição identificados	25
	בר 23
	Z/
3.1. IDENTIFICAÇÃO DOS ATAC-SEQ PEAKS	27
3.2. DETERMINAÇÃO DA LOCALIZAÇÃO DOS ELEMENTOS REGULATORIOS EM ATAC-SEQ PEAKS	27
3.3. CLUSTERIZAÇÃO DE GENES COM O MESMO PERFIL DE EXPRESSÃO E ASSOCIAÇÃO DOS MÓDULOS G	ÉNICOS
AO FENÓTIPO DE INTERESSE	29
3.4. OVERLAP ENTRE ATAC-SEQ PEAKS E GENES EXPRESSOS NO MÚSCULO ESQUELÉTICO	31
3.5. IDENTIFICAÇÃO DE SEQUÊNCIAS <i>MOTIFS</i> EM GENES CO-EXPRESSOS	32
3.6. Ocorrência das sequências <i>motifs</i> em banco de dados	32
3.7. Identificação de sequências <i>motifs</i> em regiões de ATAC-Seq <i>peaks</i>	35
3.8. Módulos de interesse	40
3.9. Descrição dos fatores de transcrição identificados	40

4. DISCUSSÃO	43
5. CONCLUSÃO	49
REFERÊNCIAS	51

RESUMO

Identificação de sequências *motifs* e elementos regulatórios em ATAC-Seq *peaks* associados à expressão gênica e à deposição de gordura intramuscular em bovinos da raça Nelore

Gordura intramuscular (GIM) representa o conteúdo de gordura acumulada entre as fibras ou dentro das células musculares e vem sendo associada à maciez da carne. Compreender os mecanismos funcionais que regulam GIM tornou-se uma questão importante tanto para a indústria de carne quanto para a saúde humana. A regulação do genoma envolve todas as peculiaridades da expressão gênica, incluindo as modificações bioquímicas do DNA, o arranjo físico dos cromossomos e a ativação do mecanismo de transcrição. Neste estudo, a técnica Assay for Transposase-Accessible Chromatin (ATAC-Seq) foi realizada para caracterizar as regiões de acessibilidade da cromatina do músculo Longissimus dorsi de espécies bovinas (Bos indicus). Em um segundo momento, a análise de associação foi realizada entre essas regiões e os motivos candidatos e elementos reguladores associados à deposição e composição de GIM. O método ATAC-seq identificou 18.942 regiões acessíveis à cromatina e regiões reguladoras candidatas. Analisando os ATAC-Seq peaks nas regiões reguladoras de 10.117 genes co-expressos, observamos 506 motivos descritos em bancos de dados e associados à deposição intramuscular de gordura. Nosso estudo forneceu novas ideias sobre o processo de regulação do genoma que podem ajudar a entender melhor os mecanismos bioquímicos subjacentes a diferentes características fenotípicas relevantes para o mercado, para a sociedade e para a ciência.

Palavras-chave: Regulação gênica, ATAC-Seq, Gordura intramuscular, Bos indicus

ABSTRACT

Identification of sequence motifs and regulatory elements within ATAC-Seq peaks associated with gene expression and intramuscular fat deposition in Nellore cattle

Intramuscular fat (IMF) represents the amount of fat accumulated between muscle fibers or within muscle cells, which has been associated with beef tenderness. Understanding the functional mechanisms that regulate IMF content became is an important issue to meat industry and human health. Genome regulation involves all facets of gene expression, including the biochemical modifications of DNA, the physical arrangement of chromosomes and the transcription machinery activation. In this study, the Assay for Transposase-Accessible Chromatin (ATAC-seq) was performed to characterize the chromatin accessibility regions of Longissimus dorsi muscle from bovine (*Bos indicus*) species. In a second moment, the association analysis was performed among these regions and candidate motifs and regulatory elements associated with IMF deposition and composition. ATAC-seq method identified 18,942 chromatin accessible regions and candidate regulatory regions. By analyzing the ATAC-Seq peaks in the regulatory regions of 10,117 co-expressed genes, we observed 506 motifs described in databases and associated with intramuscular fat deposition. Our study provided novel insights into genome regulation process that can help to better understand the functional mechanisms underlying different phenotypic traits relevant to the meat industry, for society and for science.

Keywords: Gene expression, ATAC-Seq, Intramuscular fat, Bos indicus

1. REVISÃO DE LITERATURA

1.1. Material Genético

O genoma eucariótico é envolto por proteínas histonas formando elementos fundamentais conhecidos como nucleossomos. Esse enovelamento permite que o DNA se aloje fazendo com que, estruturalmente, o material genético receba o nome de cromatina^{1,2}. A unidade nucleossomal corresponde a aproximadamente 147 pares de bases circundadas por um octâmero proteico formado por um tetrâmero de histonas H3 e H4 e dois dímeros de histonas H2A e H2B^{3,4}.

O posicionamento dos nucleossomos, além de determinar a acessibilidade do DNA aos elementos regulatórios, define a ocorrência de um dos dois estados da cromatina: eucromatina e heterocromatina^{5,6,7}. Quando presentes, os nucleossomos favorecem a compactação do material genético impedindo a ligação dos fatores de transcrição a sítios específicos do DNA. Esse estado fechado da cromatina torna o DNA inacessível recebendo, portanto, o nome de heterocromatina⁶. No entanto, quando há anulação da barreira nucleossomal ocorre a abertura da cromatina e, consequentemente, a exposição de regiões genéticas susceptíveis a interação proteica. Esse estado, denominado eucromatina, viabiliza a ação da maquinaria de transcrição atuando positivamente sobre a expressão gênica^{8,9}.

1.2. Remodelagem da cromatina

A estrutura da cromatina, que regula a expressão gênica, é constantemente remodelada a fim de se obter a conformação genética adequada à demanda transcricional do organismo^{7,10}. O mecanismo de remodelagem de cromatina influencia a transcrição gênica uma vez que, nesse processo, a cromatina próxima a sequências regulatórias é dinamicamente alterada¹¹. Quando a ativação da expressão gênica é requerida, elementos proteicos remodeladores deslocam os nucleossomos para longe da região promotora enquanto que a exigência metabólica pela repressão gênica sinaliza para que os remodeladores deslizem os nucleossomos alterando novamente suas posições¹².

O dinamismo da cromatina é regulado por dois eventos que consistem, principalmente, em alterar a interação DNA-histona e modificar quimicamente a estrutura dos nucleossomos¹³. Em outras palavras, a remodelagem altera a densidade dos nucleossomos ou faz com que o octâmero de histonas seja movido ao longo da molécula de DNA, promovendo a mudança estrutural das regiões genéticas inacessíveis^{11,14}. A fim de se rearranjar os nucleossomos, os remodeladores de cromatina dependentes de adenosina trifosfato (ATP) utilizam a energia proveniente da hidrólise do ATP para romper e afrouxar o contato entre histona e DNA regulando, assim, o grau de compactação da cromatina. Esses remodeladores podem ser classificados em 4 famílias conhecidas como SWI/SNF, ISWI, CHD e INO80. O complexo nomeado NurD tem como peculiaridade atuar tanto alterando a composição química das histonas quanto como remodelador dependente de ATP¹⁴.

1.3. Epigenética

As histonas que formam o octâmero proteico possuem caudas aminoterminais que interagem com o fosfato presente na estrutura do DNA facilitando, assim, as interações intra- e inter- nucleossomal¹⁵. Os resíduos básicos presentes nas caudas das histonas são susceptíveis a alterações químicas (marcas epigenéticas) visto que se modificam covalentemente pela ligação de grupos acetil e metil sendo os primeiros enriquecidos em regiões de eucromatina e os segundos associados ao estado de heterocromatina^{16–19}. Resumidamente, a adição de grupos acetil nas histonas altera tanto a interação entre nucleossomo e DNA quanto entre nucleossomos adjacentes fazendo com que a mudança física do octâmero (deslizamento) resulte na acessibilidade da cromatina (cromatina aberta). O evento de acetilação agrupa o complexo SWI/SNF e ISWI responsáveis por perturbar a estrutura nucleossomal, aumentando a susceptibilidade do DNA à ação da nuclease, e por alterar a localização dos nucleossomos no DNA, respectivamente. Ambos promovem a formação de *nucleossome-free regions* (NFR), conhecidas como sítios de hipersensibilidade à nuclease que abrangem os elementos e as regiões regulatórias de genes transcricionalmente ativos²⁰.

A metilação do DNA, apesar de não modificar a estrutura das histonas, altera o suporte químico do DNA. Além disso, a metilação de DNA estimula a sinalização para que o complexo multiproteico de deacetilação de histonas atue. Esse processo é favorecido por uma enzima responsável por ligar grupos metil ao carbono 5 de um resíduo específico de citosina, conferindo um impedimento estérico à molécula de DNA^{21,22}. Em geral, os grupamentos metil são acrescentados às citosinas presentes nos dinucleotídios CG do organismo. Os dinucleotídios CG não metilados são encontrados em aglomerados situados próximos às regiões promotoras que recebem a denominação de ilhas CpG (onde P corresponde a uma ligação fosfodiéster)²³.

A acetilação de histonas relaciona-se ao estado de eucromatina enquanto o processo de metilação é associado ao estado de heterocromatina^{17–19}. No entanto, a tri-metilação de histonas

H3K4, H3K36, H3K79 sinaliza cromatina ativa e de histonas H3K9, H3K27, H3K56, cromatina inativa²⁴. A relação entre metilação de histona e inativação da cromatina pode ser explicada uma vez que histona H3 metilada em lisina 9 representa o local de ligação à maior componente proteica da heterocromatina, HP1²⁵. As regiões promotoras ativas geralmente são enriquecidas por H3K4me3 enquanto os *enhancers* ativos, por H3K4me1 e H3K27ac. A monometilação de lisina 4 da histona H3 promove a incorporação de nucleossomos contendo H2A – molécula responsável por favorecer a dinâmica estrutural da cromatina e, portanto, aumentar o acesso dos fatores de transcrição. Já a tri-metilação de lisina 4 da histona H3 é associado ao recrutamento dos domínios proteicos CHD1 e BPTF responsáveis por reprimir a ligação da proteína NurD propiciando a abertura da cromatina^{26,27}.

1.4. Evento transcricional e elementos regulatórios

A transcrição, primeira etapa do mecanismo da expressão gênica e da síntese proteica, é caracterizada pela cópia da informação genética do DNA em uma molécula linear denominada RNA. O processo transcricional é modulado por diversos elementos proteicos que permitem a ocorrência do evento *splicing* responsável por ampliar a possibilidade de produtos gênicos e, consequentemente, da síntese proteica²⁸. A principal enzima atuante nesse processo é a RNA polimerase (Pol) que, juntamente com os fatores de transcrição, reconhecem as sequências regulatórias do gene promovendo um ambiente bioquímico favorável para que a transcrição ocorra^{20,30}. Os organismos eucariotos requerem três classes de Pol - I, II e III - para efetivar a síntese de RNA de diferentes genes³¹. A Pol I foi associada à síntese da molécula de RNA ribossomal (rRNA) elemento primário dos ribossomos. A Pol II é responsável pela produção de RNA mensageiro (mRNA) indispensável à transferência das informações genéticas para o citoplasma. Pol III, por sua vez, sintetiza o RNA transportador (tRNA) e a subunidade 5S do rRNA, essenciais ao metabolismo proteico^{32–34}.

No entanto, as Pol, por apresentarem baixa afinidade pelas sequências promotoras dos genes, dependem da ação dos fatores proteicos de transcrição para sinalizar o início da transcrição³⁵. A Pol I auxilia na formação do complexo pré-iniciação (PIC) em torno do promotor local onde, em um segundo momento, a Pol II se encaixa para realizar o processo transcricional^{30,32}. De maneira mais detalhada, a Pol e os fatores de transcrição primeiramente reconhecem e se ligam à região regulatória *upstream* ao sítio de início da transcrição (TSS) para formar o PIC³⁰.

A maquinaria transcricional é constituída por duas classes de elementos regulatórios: *cis-acting* e *trans-acting*. Ambos se diferenciam devido a maneira como atuam no ambiente bioquímico^{36,37}. Os elementos *cis-acting* correspondem a sequências nucleotídicas sendo, portanto, regiões do próprio DNA, enquanto os *trans-acting* abrangem as proteínas que interagem umas com as outras ou com o material genético. Assim sendo, os elementos reguladores *cis-acting* contêm sítios de reconhecimento para fatores de transcrição *trans-acting* que promovem o aumento ou a repressão de evento transcricional³⁸. Os elementos *cis-acting* são representados por duas famílias: (a) promotor central e elementos reguladores proximais e (b) elementos regulatórios distais conhecidos como *enhancers*, silenciadores, isoladores ou regiões de controle locus (LCR)^{36,39}.

Os *trans-acting* podem ser classificados em três grupos: fatores gerais (ou básicos) de transcrição (GTFs), proteínas ativadoras específicas (ativadores) e coativadores. Nos GTFs há uma variedade de componentes auxiliares conhecidos como TFIIA, TFIIB, TFIID, TFIIE, TFIIF e TFIIH ⁴⁰. O TFIID é a proteína que identifica a região promotora. Sua subunidade TBP se liga à TATA box enquanto a outra subunidade TBP-*associated factors* (TAF), confere a especificidade à ligação. Este mecanismo promove a curvatura do DNA em torno do TBP favorecendo a ligação do TFIIB, que, por sua vez, posiciona o promotor na polimerase. TFIIH inclui helicases dependentes de ATP que liberam a região promotora para iniciar a transcrição^{41,42}.

Os GTFs se reúnem de maneira ordenada no sítio promotor para formar o PIC responsável por direcionar a Pol II para o local de TSS⁴³. A atividade transcricional é estimulada por ativadores que, além de interagirem com subunidades dos complexos proteicos, modificam a estrutura da cromatina permitindo que a Pol II e outras proteínas tenham acesso ao DNA. A região *upstream* (até algumas centenas de pares de bases) ao promotor principal (*core promoter*) contém múltiplos locais de ligação para ativadores. Há vários elementos regulatórios transcricionais e funcionais na região adjacente ao TSS. Os coativadores, por sua vez, embora não façam parte da maquinaria transcricional, facilitam a ativação gênica ao se associarem aos ativadores. O fato de os fatores de transcrição se ligarem a elementos regulatórios *cis-acting* (considerados sequências *motifs*) e interagirem com proteínas específicas aos sítios promotores, explica como a transcrição pode ser estimulada por sequências reguladoras distais ^{43,44}.

A transcrição basal pode ser potencializada pela ação dos elementos ativadores, ou pelo complexo ativador 'enhanceossomo', que se ligam a elementos *cis-acting upstream* e estimulam a formação do PIC⁴⁰. O efeito de um ativador pode ser afetado pelos elementos coativadores que, apesar de não se ligarem diretamente à molécula de DNA, interagem com outras proteínas modulando a montagem do PIC ou modificando a estrutura da cromatina^{45,46}. As sequências de DNA nas quais os ativadores se ligam, são chamadas de sítios de ligação de fatores de transcrição

(TFBSs) e apresentam, geralmente, comprimento entre 6 e 12 pares de base. Os TFBSs caracterizam-se por serem degenerados sendo, na maioria das vezes, descritos por uma sequência consenso^{36,47}.

As atividades dos elementos regulatórios distais são indispensáveis ao processo transcricional ³⁹. Os *enhancers*, frequentemente situados *upstream* a região promotora, são compostos por sítios TFBSs e regulam positivamente a transcrição gênica. Os *enhancers* recrutam fatores de transcrição que interagem com o promotor formando um DNA-*looping* que aproxima os elementos regulatórios *cis-acting* ^{39,48,49}. A ligação entre *enhancers* e fatores de transcrição pode ser discutida por dois modelos diferentes: *Enhanceosome* e *Billboard*. No primeiro modelo acredita-se que os *cis-acting* recrutem fatores de transcrição formando um complexo nucleoproteico ativador da transcrição. Nesse cenário há uma rede de interação proteica que só se rompe quando uma proteína é perdida ^{50,51}. O modelo *Billboard*, por sua vez, supõe que as ligações dos fatores de transcrição são totalmente independentes, levando-se em consideração apenas sua especificidade⁵².

Os *silencers*, localizados *donstream* à região promotora, são elementos que regulam negativamente a expressão de um gene. Se diferenciam dos *enhancers* pois são específicos a fatores de transcrição negativos conhecidos como repressores ^{53,54}. *Insulators* correspondem a sequências de aproximadamente 3 kb cuja função é limitar a ação dos elementos regulatórios inibindo, assim, tanto a interação entre *enhancer*-promotor quanto a permanência do estado estrutural de heterocromatina. Ambos efeitos impedem que genes próximos sejam controlados pela transcrição do gene vizinho ^{36,39,55}. Os LCRs, por sua vez, também encontrados em região *upstream* ao promotor, são compostos por vários outros elementos *cis-acting*, e situam-se próximos à DHSs, fornecendo um domínio de eucromatina para os genes aos quais se associam ⁵⁶.

Regiões *motifs* correspondem a curtas sequências nucleotídicas que apresentam importante atividade biológica sendo consideradas sítios específicos de ligação proteica ⁵⁷. Detectar e caracterizar *motifs* é indispensável para a elucidação de modelos bioquímicosmoleculares envolvidos na regulação da expressão gênica. A estratégia mais plausível para que *cise trans-acting elements* sejam identificados é primeiramente identificar *motifs* em genes co-expressos para o fenótipo de interesse. Ao predizer quais *motifs* estão super-representados em genes que carregam similaridade molecular de expressão, podemos apontar quais são os elementos regulatórios associados ao mecanismo transcricional de determinada característica. Quando obtidos, esses *motifs* podem ser pesquisados dentro das regiões de eucromatina a fim de entender como as vias metabólicas e rotas bioquímicas se associam à regulação das funções biológicas exercidas por essas sequências nucleotídicas. Em vista disso, o conceito "*Motif discovery*" vem sendo muito utilizado para identificar elementos regulatórios similares e super-representados ao longo das regiões *upstream* de genes funcionalmente relacionados ^{57,58}.

1.5. Identificação de regiões de eucromatina

Estudos recentes vêm utilizando diferentes metodologias para identificar regiões de eucromatina e componentes regulatórios. A técnica *Assay for Transposase-Accessible Chromatin* (ATAC-seq), que utiliza transposase bacteriana Tn5, isola apenas regiões de cromatina aberta devido a inviabilidade enzimática em acessar molécula de DNA vinculada a nucleossomos. Essa metodologia baseia-se em identificar regiões de eucromatina a partir do mecanismo de transposição realizado por uma enzima transposase bacteriana, Tn5. Uma vez que a transposase é incapaz de acessar molécula de DNA vinculada a nucleossomos, a Tn5 integra *transposans* em região de eucromatina. O método ATAC-seq consistem em incubar o DNA alvo com Tn5, favorecendo tanto a ligação de adaptadores (sequências específicas finais, OE) quanto a fragmentação de regiões de cromatina aberta⁵⁹.

Os *transposons* ou elementos transponíveis (TE) são os agentes causais de rearranjos genéticos e correspondem a elementos móveis com sequências específicas de DNA⁶⁰. Os DNA *transposons* correspondem aos elementos de Classe II e baseiam-se no mecanismo '*cut and paste*' no qual os *transposons* são excisados de um local genômico antes de ser inserido em outro. Os *retrotransposons*, por sua vez, referem-se aos elementos transponíveis Classe I mobilizados pelo mecanismo de '*copy and paste*' a partir do qual o RNA é reversamente transcrito e inserido em um novo local genômico⁶¹. Transposase (Tnp) é a enzima que se liga à extremidade de um *transposon* e catalisa o mecanismo '*cut and paste*' deste elemento genético para outra parte do genoma⁶². A Tnp facilita o movimento do *transposon* Tn5 em um processo de 5 etapas. A primeira consiste na ligação da enzima à sequência '*outside end*' (OE) do Tn5. Posteriormente ocorre a formação de um complexo nucleoproteico conhecido como sinapse seguido da clivagem e fragmentação da sequência específica final. As duas últimas etapas correspondem a associação do *transposon* ao DNA alvo e a transferência (inserção) genômica do elemento transponível⁶³.

1.6. Gordura intramuscular (GIM)

A GIM, também conhecida como marmoreio, é ocasionada pelo acúmulo de fosfolipídios e triglicerídeos no interior das células musculares ou entre as fibras do músculo⁶⁴. A

deposição de GIM não é influenciada por fatores isolados, mas por um conjunto de características tais como sexo, idade, nutrição e genética do animal^{65,66}.

A carne bovina tem grande importância para a nutrição humana uma vez que é constituída por nutrientes indispensáveis ao bom funcionamento do organismo. Além disso, a deposição de lipídios no músculo esquelético interfere nas propriedades organolépticas da carne sendo a GIM fortemente associada à suculência, maciez e ao sabor da carne^{67,68}. Apesar de o consumo de gordura ser relacionado a doenças cardíacas, obesidade e diabetes, estudos indicam que a alta concentração de ácidos graxos monossaturados presente na gordura da carne, favorece a redução de colesterol LDL na circulação sanguínea⁶⁵.

O conteúdo de GIM é considerado uma característica poligênica visto que muitos genes contribuem, direta ou indiretamente, para que o fenótipo de deposição de gordura seja expresso em determinado organismo. O entendimento da bioquímica-molecular envolvida no metabolismo lipídico nos permite esclarecer como que ocorre a regulação gênica no músculo esquelético e como este processo interfere na atividade das regiões promotoras, consideradas regulatórias.

2. INTRODUÇÃO

Os animais de corte brasileiros são, majoritariamente, da raça Nelore (origem *Bos indicus* - Zebu) considerada tolerante ao calor e a parasitas e mais bem adaptada às condições tropicais⁶⁹. A carne bovina é importante para a saúde humana uma vez que é composta por alto teor de proteína, ferro, zinco, vitaminas do complexo B e ácidos graxos poliinsaturados essenciais ⁶⁹. A gordura intramuscular (GIM), por sua vez, afeta o valor nutricional e interfere na maciez, na palatabilidade e em outras características sensoriais que favorecem a satisfação do consumidor, a competitividade comercial brasileira e a economia nacional^{70,71}. A deposição de GIM é considerada um fenótipo poligênico visto que é influenciada por diversos fatores genéticos que se associam e se comunicam entre si^{65,72–74}. Em vista disso, é importante elucidar os mecanismos genéticos que regulam a deposição de GIM no músculo esquelético. A seleção genética a favor do teor de GIM responderia questões biológicas interessantes à ciência da carne e à saúde do consumidor favorecendo, também, a economia nacional^{71,73}.

Cromatina corresponde a estrutura na qual o genoma dos organismos eucariotos é organizado dentro do núcleo celular^{75,76}. O enovelamento do material genético em torno de proteínas histonas origina a unidade funcional da cromatina, conhecida como nucleossomo, responsável por modular a acessibilidade do material genético e controlar o processo transcricional^{9,77}. Quando presentes na cromatina, os nucleossomos reprimem a transcrição uma vez que bloqueiam sítios de ligação proteica ao DNA, sendo este estado estrutural conhecido como heterocromatina⁷⁸. A anulação da barreira nucleossomal, ao exibir *nucleosome-free regions (NFR) e DNase I hypersensitive sites (DHSs)*, expõe sítios genéticos susceptíveis a interação com fatores de transcrição, esse estado é nomeado eucromatina⁷⁹. Os dois estados da cromatina ocorrem de acordo com a necessidade metabólica em requerer a expressão ou o silenciamento de determinado gene. Essa transição estrutural corresponde ao evento de remodelagem de cromatina, indispensável para o equilíbrio do organismo^{7,10}.

A transcrição, etapa inicial da expressão gênica, ocorre devido tanto à atividade da maquinaria transcricional quanto à organização e à orientação estrutural dos genes a serem transcritos⁴². Essa etapa é realizada pela enzima RNA polimerase II (Pol II) que, ao permitir o acesso de complexos proteicos ao DNA, sintetiza uma molécula de RNA. Os fatores de transcrição (FT) e as proteínas reguladoras de DNA são imprescindíveis na identificação dos genes a serem transcritos pela Pol II. O processo transcricional é um mecanismo que ocorre apenas quando o ambiente genético se encontra favorável para que toda a maquinaria funcione

adequadamente. A maquinaria transcricional utiliza elementos *cis-acting* (promotores e elementos regulatórios distais: *enhancers, silencers, insulators ou locus control regions* (LCR)) que representam locais específicos para ligação de fatores de transcrição *trans-acting*^{37,42,80,81}.

Regiões *motifs*, ao corresponderem a curtos padrões de sequências nucleotídicas evolutivamente conservadas, apresentam importante atividade biológica sendo, portanto, consideradas elementos *cis-acting* (sítios específicos de ligação proteica). Detectar e caracterizar *motifs* é indispensável para a elucidação de modelos bioquímicos-moleculares envolvidos na regulação da expressão gênica. Em vista disso, o conceito *"Motif discovery"* vem sendo muito utilizado para identificar elementos regulatórios similares e super-representados ao longo das regiões *upstream* de genes funcionalmente relacionados^{58,82}.

O objetivo principal deste projeto é identificar os elementos regulatórios e as sequências *motifs* presentes nas regiões de *ATAC-Seq peaks* obtidos de tecido muscular de bovinos Nelore, por meio da descrição de genes que compartilham semelhantes padrões de expressão no músculo esquelético. Espera-se, com esse trabalho, predizer o efeito regulatório das regiões de eucromatina na expressão dos genes associados a deposição de gordura intramuscular.

2.1. Hipótese

Regiões de eucromatina compreendem sequências *motifs* e elementos regulatórios indispensáveis à regulação de genes co-expressos.

2.2. Objetivo

2.2.1. Objetivo geral

Identificar elementos regulatórios e sequências *motifs* presentes nas regiões de *ATAC-Seq peaks* de tecido muscular de bovinos Nelore.

2.2.2. Objetivos específicos

- 1) Identificar elementos regulatórios nas regiões dos ATAC-Seq peaks;
- 2) Identificar sequências motifs nas regiões dos ATAC-Seq peaks;
- 3) Determinar os fatores de transcrição associados aos motifs previamente obtidos;
- 4) Atribuir funções biológicas aos motifs.

2.2.3. MATERIAIS E MÉTODOS

Esse projeto é continuação do projeto de mestrado⁸³ desenvolvido entre os anos 2016 e 2018 e financiado pela FAPESP (nº processo: 2016/13773-6). Em vista disso, as análises de bioinformática foram realizadas a partir das amostras obtidas no projeto citado anteriormente, as quais irei definir como fase 1. O alinhamento contra o genoma de referência precisou ser novamente analisado visto que entre os dois trabalhos houve alteração na versão do genoma. No projeto de mestrado foi utilizada a versão "*Bos taurus*" (versão UMD 3.1) e no presente projeto foi aplicada a versão mais recente, "*Bos taurus*" (versão ARS-UCD1.2). Devido a isso, todas as análises posteriores ao alinhamento contra o genoma de referência precisaram ser refeitas de acordo com os novos dados obtidos, as quais irei descrever como Fase 2.

2.2.3.1. Fase 1

Detalhes sobre a metodologia descrita na fase 1 podem ser encontrados descritos na seguinte referência: Morosini, N.S. (2018). Identificação e caracterização de regiões de eucromatina associadas à regulação da expressão gênica e à gordura intramuscular em bovinos da raça Nelore. Dissertação de Mestrado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/D.11.2018.tde-2906208-170903. Recuperado de www.teses.usp.br

2.2.3.1.1. Amostras

Foram obtidas amostras de tecido muscular de bovino Nelore macho, com idade de 24 meses, recém abatidos e entregues para açougue comercial. Para dar seguimento aos experimentos, foram utilizadas duas réplicas biológicas e quatro replicatas técnicas⁸³.

2.2.3.1.2. Obtenção de regiões de eucromatina

Para isolar e identificar as regiões de eucromatina (ATAC-Seq peaks)⁸³ foi aplicada a técnica 'Assay for Transposase-Accessible Chromatin' (ATAC-seq)⁸⁴.

2.2.3.1.3. Sequenciamento ATAC-DNA

Foram gerados aproximadamente 40 milhões de *reads* por biblioteca sequenciadas em HiSeq 2500. Foi utilizado o sequenciamento p*aired-end* 2 x 100⁸³.

2.2.3.2. Fase 2

2.2.3.2.1. Alinhamento contra o genoma de referência

O mapeamento de *reads* foi realizado contra o genoma de referência "Bos taurus" (versão ARS-UCD 1.2) por meio do software *Bontie2* (v. 2.3.5.1)⁸⁵. As *reads* mapeadas em regiões de DNA mitocondrial e classificadas como duplicatas de PCR foram removidas por meio dos *softwares Picard* (v. 2.20.5) e *Samtools* (v. 1.9)⁸⁶.

2.2.3.2.2. Identificação dos ATAC-Seq peaks

A identificação das regiões de eucromatina foi realizada pelo programa *Model-based Analysis of Chip*-Seq (MACS2 v. 2.1.2.1) que estima regiões (picos) potencialmente eucromáticas a partir de um valor esperado de distribuição, Λ , estipulado pela distribuição de *Poisson*⁸⁷. O controle de qualidade, tanto dos dados de mapeamento quanto dos gerados pelo MACS2, foi realizado pela plataforma *Qualimap* (v. 2.2.1)⁸⁸.

2.2.3.2.3. Determinação da localização dos elementos regulatórios em ATAC-Seq peaks

Para determinar quantos pares de bases *upstream e dountream* aos TSSs deveriam ser considerados para predizer os possíveis elementos regulatórios, utilizamos as regiões 5'UTR da lista de genes *Bos taurus*. De maneira mais detalhada, foi utilizado o arquivo contendo a anotação do genoma *Bos taurus* (versão ARS-UCD 1.2) para filtrar apenas os genes com informações 5'UTR descritas. Optamos por essa estratégia visto que as regiões 5'UTR correspondem às regiões TSS sendo, assim, possível utilizá-las para sobrepor com os dados *ATAC-Seq peaks* e determinar em quais localizações genéticas há enriquecimento de regiões de eucromatina (regulatórias). A sobreposição entre as regiões *ATAC-Seq peaks* e o genes selecionados foi

realizada pelo software *deepTools* (v. 3.1.2)⁸⁹ a partir do qual pudemos, também, obter a análise gráfica e estimar a distância entre os sítios TSS e os elementos regulatórios putativos.

2.2.3.2.1. *Clusterização* de genes com o mesmo perfil de expressão e detecção de associações de módulos ao fenótipo

Os clusters (módulos) de genes que compartilham o mesmo perfil de expressão, foram identificados por meio do *R software package Weighted Network Analysis*' (WGCNA)⁹⁰. Esse programa realiza o agrupamento dos genes co-expressos uma vez que calcula, aos pares, os coeficientes de *Pearson* de todos os genes a fim de predizer a similaridade de expressão existente entre eles. As correlações são estimadas e corrigidas por meio da matriz de similaridade e de adjacência⁹¹. O número mínimo de genes em um módulo foi definido como 30 e cada módulo representa genes com perfis de expressão semelhantes. No total 10.117 genes expressos em *Bos taurus* e identificados por Cesar et al., 2018⁹² foram utilizados. Destes, 9.775 genes expressos associados as características intramusculares de conteúdo de gordura foram submetidos à análise WGCNA pois foram utilizados apenas os genes anotados contendo todas as informações necessárias e exigidas pelos parâmetros do programa. A partir dessa análise, os genes co-expressos foram agrupados de acordo com a similaridade estrutural e biológica que apresentam permitindo, assim, o estudo das redes biológicas com as quais os genes se relacionam⁹³.

2.2.3.2.2. Sobreposição entre ATAC-Seq *peaks* e genes expressos no músculo esquelético

Todas as sobreposições entre os recursos descobertos neste artigo foram realizadas pela ferramenta de interseção BEDtools (v.2.27.1)⁹⁴.

2.2.3.2.3. Identificação de sequências motifs em genes co-expressos

As sequências *motifs* foram identificadas, analisadas e caracterizadas pelo *software* MEME suíte, um kit de ferramentas e banco de dados, que permite a análise e a caracterização de sinais biológicos presentes nas regiões regulatórias por meio do modelo *Position Weight Matrix* (PWM)^{95,96}. O banco de dados utilizado para comparar as sequências *motifs* foi o referente à

espécie Homo sapiens (HOCOMOCOv11_full_HUMAN) disponibilizado pelo software MEME suíte.

2.2.3.2.4. Descoberta de cis-acting elements (motifs)

A ferramenta escolhida para executar a descoberta de *motifs* foi o algoritmo MEME que, ao utilizar a position weight matrix (PWM), descreve a probabilidade de cada base (nucleotídeo) se situar em cada posição do motif. O PWM, além de ser obtido a partir de um conjunto de sequências genéticas funcionalmente relacionadas, possui uma linha para cada símbolo do alfabeto genético - 4 linhas para nucleotídeos (DNA) ou 20 linhas para aminoácidos (sequência proteica) - e uma coluna para cada localização do motif. Primeiramente é criada a Position Frequency Matrix (PFM) por meio da contagem das ocorrências de cada nucleotídeo em cada posição e, então, uma matriz de probabilidade de posição (PPM) é formada distribuindo e normalizando a contagem nucleotídica anterior em cada sítio genético. À medida em que os motifs vão sendo descobertos o algoritmo é programado para removê-los, fazendo com que apenas um único motif possa ser trabalhado por vez e que não haja a detecção de motifs alternativos^{97,98}. Os motifs são representados por logos escalonados em bits nos quais um símbolo duplicado é usado quando há probabilidade de duas bases ocorrerem em mais de 75% dos sítios e um símbolo triplicado, ou seja, 3 nucleotídeos aparecem caso uma base não ocorra. Posições perfeitamente conservadas contêm 2 bits de informação. Enquanto 1 bit é destacado quando duas das quatro bases ocorrem 50% do tempo em um sítio genético. As posições onde todas as quatro bases ocorrem igualmente não contêm nenhuma informação⁹⁹.

2.2.3.2.5. Identificação de motifs em ATAC-Seq peaks

A fim de analisar quais são os elementos regulatórios presentes nas regiões regulatórias representadas pelos ATAC-Seq *peaks*, foi utilizado o recurso SpaMo responsável por identificar padrões de espaçamento genético em um arquivo primário (*ATAC-Seq peaks*) entre dois grupos de *motifs*: os encontrados pelo software MEME e os anotados em bancos de dados¹⁰⁰ A principal análise fornecida por essa ferramenta se baseia em prever a localização de um sítio TFBS dentro das regiões de eucromatina (regulatórias) e testar a hipótese de que há uma relação interativa e, portanto, bioquímica entre *motifs* primários e secundários. Assim sendo, esse programa calcula qual a probabilidade de ocorrência (ao acaso) de espaçamentos próximos de modo a encontrar a melhor correspondência (*hit*) de *motif (motif secundário inferido)*¹⁰⁰.

2.2.3.2.6. Ocorrência de sequência motifs em bancos de dados

O algoritmo utilizado para pesquisar a ocorrência de sequências *motifs* em bancos de dados foi o FIMO. Essa análise, denominada *scanning*, se baseia em procurar outras ocorrências dos *motifs* descobertos pelo MEME. O algoritmo FIMO, visto que localiza todas as ocorrências de *motifs* individuais e pontua (P-*value*) a correspondência de cada *motif* em cada posição do banco de dados, é sugerido para escanear genomas complexos⁹⁵.

2.2.3.2.7. Descrição dos fatores de transcrição identificados

As informações sobre a função dos fatores de transcrição (proteínas) foram obtidas por meio do banco de dados *UniProt Knowledgebase* (UniProtKB) que consiste em duas sessões: Swiss-Prot cujos registros são informações proteicas revisadas e TrEMBL que abrange dados de proteínas computacionalmente analisadas porém ainda não anotadas¹⁰¹. Todas as informações descritas nesse trabalho foram retiradas dos registros revisados.

3. RESULTADOS

3.1. Identificação dos ATAC-Seq peaks

Em nosso estudo foram utilizadas duas replicatas biológicas e quatro replicatas técnicas: I, II, III e IV. A amostra I apresentou 35.204 regiões de eucromatina enquanto a amostra II, proveniente da mesma replicata biológica, 39.882. Já as amostras III e IV obtidas a partir da segunda replicata biológica, abrangeram 37.678 e 44.850 sítios de cromatina aberta, respectivamente. Para trabalharmos com as regiões de eucromatina que representam os locais de *ATAC-Seq peaks* mais prováveis de serem associados ao fenótipo de interesse e a fim de evitar dados falsos negativos, optamos por selecionar apenas os *ATAC-Seq peaks* identificados em todas as replicatas técnica. Em vista disso, no total realizamos todas as seguintes análises com 18.942 picos de eucromatina.

3.2. Determinação da localização dos elementos regulatórios em ATAC-Seq peaks

Para validar ainda mais os ATAC-Seq *peaks*, usamos a lista completa de genes anotados para o genoma de *Bos taurus* (versão ARS-UCD 1.2) e selecionamos apenas aqueles que continham informações sobre o site 5'UTR (correspondente aos TSSs ativos). O mapa *heatmap* (figura 1) indica como as *reads* dos ATAC-Seq *peaks* (em azul) e as regiões 5'UTR dos genes anotados do *Bos taurus* se sobrepõem quando suas informações são cruzadas. O sinal de cobertura das *reads* é mostrado em uma janela em escala de 2 kb *upstream* e *downstram* aos TSSs e a densidade do *heatmap* é proporcional ao número de regiões analisadas. É possível observar que os locais localizados ao redor dos TSS genéticos têm um enriquecimento significativamente maior nas regiões reguladoras representadas pelos ATAC-Seq *peaks*.



Figura 1. Heatmap representando as regiões regulatórias dos genes (2kb upstream e downstream aos TSSs) expressos em Bos taurus enriquecidas por reads de ATAC-seq peaks (regiões de eucromatina).

3.3. *Clusterização* de genes com o mesmo perfil de expressão e associação dos módulos gênicos ao fenótipo de interesse

Para melhor direcionar as buscas por *motifs* genéticos e tornar a análise exploratória mais robusta e precisa, optamos por trabalhar, inicialmente, com os genes expressos no músculo esquelético⁹². A partir desses dados, agrupamos os genes co-expressos uma vez que genes com o mesmo perfil de expressão podem compartilhar locais de ligação ao fator de transcrição (TFBS). Construímos uma árvore hierárquica de *clustering* usando o programa WGCNA. Para decidir quais agrupamentos devem ou não ser agrupados e qual a diferença entre conjunto de genes, a distância entre objetos (módulos) foi calculada. No total, os genes co-expressos foram agrupados em 20 módulos (Figura 2) contendo um número diferente de genes. A associação entre genes coexpressos (módulos) e 16 caracteres - Extrato Etéreo (EE), Área do Lombo (AOLf), Espessura de Gordura Subcutânea (EGSf), Ácido mirístico (C14_0), Ácido miristoleico (C14_1_C9), Ácido palmítico (C16_0), Ácido palmitoléico (C16_1_C9), ácido esteárico (C18_0), ácido oleico (C18_1_C9), ácido linoléico (C18_2_C9_C12), ácido vaccênico (C18_2_C9_T11), soma de ácido graxo saturado (sum_sat), soma de ácidos graxos monossaturados (sum_mono), soma de ácidos graxos poliinsaturados (soma_pol), soma de ômega-3 (soma_n3) e soma de ômega-6 (soma_n6) - associados à gordura intramuscular foi analisada de modo a definirmos quais genes mais se correlacionam com o fenótipo de interesse GIM (Figura 3). Os módulos Grey, Pink, Red e Royalblue foram significativamente correlacionados (pvalor <0,05) com as características C16_0, C18_1_C9, sum_sat e sum_mono, indicando associação com genes co-expressos nesses módulos e nesses fenótipos. O módulo Green apresentou correlação significativa com o EE, a principal característica associada ao GIM. O valor acima do valor-p que representa o coeficiente do modelo linear: se for positivo, a correlação é positiva e se for negativa, a correlação é negativa. Os motivos identificados nesses 5 módulos foram utilizados para descrever a associação entre elementos reguladores e fenótipo de deposição intramuscular de gordura.



Figura 2. Representação dos módulos de genes co-expressos definidos por meio da *Hierarchical Clustering Tree* de acordo com as medidas de dissimilaridade biológica entre os genes (co-expressão.

							MOD	ULE-TRAI	TASSOCIA	TION						
MElightyellow	-0.0082 (0.16)	=0.0034 (0.59)	=0.0023 (0.69)	0.0052 (0.34)	0.0042 (0.46)	0.01 (0.07)	0.0063 (0.3)	-0.00015 (0.98)	-0.0057 (0.35)	-0.0035 (0.52)	-0.0048 (0.46)	0.0075 (0.17)	-0.0038 (0.51)	-0.0042 (0.45)	-0.0096 (0.42)	-0.0058 (0.28)
MEgreen	-0.012 (0.044)	-0.0042 (0.52)	-0.003 (0.61)	0.0088 (0.11)	0.006 (0.29)	0.0092 (0.1)	0.0027 (0.66)	0.00042 (0.94)	-0.0065 (0.3)	0.0014 (0.8)	-0.0046 (0.49)	0.008 (0.16)	-0.0051 (0.36)	2.8e-05 (1)	-0.0058 (0.63)	-0.0054 (0.31)
MEred	-0.0092 (0.1)	-0.0038 (0.54)	-0.0052 (0.35)	0.009 (0.085)	0.0074 (0.17)	0.014 (0.0063)	-0.001 (0.86)	0.0049 (0.34)	-0.015 (0.0085)	0.0018 (0.73)	-0.0073 (0.25)	0.015 (0.0061)	-0.013 (0.012)	0.0014 (0.8)	-0.015 (0.18)	0.0028 (0.59)
MEgrey	0.0046 (0.41)	-0.0048 (0.43)	-0.0044 (0.42)	0.0042 (0.42)	0.0033 (0.54)	0.015 (0.0042)	-0.0025 (0.66)	0.0081 (0.11)	-0.018 (0.002)	-0.0023 (0.66)	-0.0072 (0.24)	0.016 (0.0021)	-0.015 (0.0038)	-0.0019 (0.71)	-0.02 (0.08)	0.0073 (0.15)
MEmagenta	-0.0082 (0.17)	-0.0051 (0.44)	-0.0017 (0.77)	0.01 (0.057)	0.0065 (0.25)	0.0069 (0.21)	-0.0029 (0.63)	0.0026 (0.63)	-0.0089 (0.15)	0.006 (0.28)	-0.0025 (0.71)	0.0083 (0.14)	-0.008 (0.15)	0.0048 (0.4)	-0.0067 (0.58)	0.00041 (0.94)
MEpink	-0.0017 (0.76)	-0.0067 (0.26)	-0.0041 (0.45)	0.0067 (0.18)	0.0051 (0.33)	0.012 (0.023)	-0.0068 (0.22)	0.0066 (0.18)	-0.015 (0.0077)	0.004 (0.43)	-0.0036 (0.56)	0.014 (0.0082)	-0.014 (0.0064)	0.0042 (0.42)	-0.015 (0.18)	0.0091 (0.064)
MEturquoise	0.0015 (0.79)	-0.0017 (0.79)	0.00078 (0.89)	0.0062 (0.25)	0.0033 (0.55)	0.00028 (0.96)	-0.0057 (0.34)	0.0027 (0.61)	-0.0052 (0.4)	0.0067 (0.21)	0.0022 (0.74)	0.0031 (0.57)	-0.0052 (0.35)	0.0089 (0.22)	-0.00024 (0.98)	0.0059 (0.28)
MEsalmon	0.0086 (0.12)	-0.0016 (0.79)	-0.0027 (0.63)	-0.011 (0.029)	-0.0059 (0.27)	0.0061 (0.24)	-0.014 (0.016)	0.0045 (0.37)	-0.0042 (0.47)	-0.00042 (0.94)	-0.0012 (0.84)	0.006 (0.25)	-0.0056 (0.28)	0.0014 (0.8)	-8.2e-05 (0.99)	0.0084 (0.097)
MEtan	0.0071 (0.19)	-0.0031 (0.61)	-0.0019 (0.73)	-0.0049 (0.34)	-0.0037 (0.47)	0.0074 (0.15)	-0.014 (0.014)	0.0055 (0.26)	-0.0076 (0.18)	0.0037 (0.46)	-0.0021 (0.73)	0.0086 (0.093)	-0.0087 (0.087)	0.0055 (0.29)	0.0038 (0.73)	0.011 (0.024)
MEyellow	-0.0033 (0.55)	-0.0044 (0.46)	0.0044 (0.42)	0.005 (0.33)	0.00033 (0.95)	-0.0094 (0.058)	0.0066 (0.24)	-0.0032 (0.52)	0.0092 (0.11)	-0.002 (0.7)	0.0087 (0.16)	-0.0082 (0.12)	0.0076 (0.14)	-0.0032 (0.54)	0.0029 (0.8)	-0.0057 (0.26)
MEmidnightblue	0.007 (0.24)	-0.0014 (0.83)	-0.0047 (0.42)	-0.012 (0.031)	-0.008 (0.17)	-0.0016 (0.77)	-0.0045 (0.47)	0.0018 (0.75)	0.00077 (0.9)	-0.0014 (0.81)	-0.0019 (0.78)	-0.0023 (0.69)	0.0011 (0.85)	-0.001 (0.86)	-0.0013 (0.91)	0.0045 (0.41)
MEbrown	0.0022 (0.7)	0.00011 (0.99)	0.0013 (0.82)	-0.0077 (0.15)	-0.0054 (0.32)	-0.0051 (0.34)	0.0028 (0.64)	-0.0032 (0.54)	0.0089 (0.14)	-0.0043 (0.42)	0.00082 (0.9)	-0.0072 (0.19)	0.0083 (0.12)	-0.0047 (0.38)	0.0057 (0.62)	-0.0056 (0.28)
MEgreenyellow	-0.0017 (0.75)	-0.00073 (0.9)	0.0038 (0.5)	-0.0011 (0.83)	-0.0023 (0.65)	-0.008 (0.11)	0.0072 (0.18)	-0.0047 (0.33)	0.011 (0.044)	-0.0046 (0.35)	0.0038 (0.52)	-0.0091 (0.069)	0.01 (0.043)	-0.008 (0.24)	0.005 (0.64)	-0.0099 (0.038)
MElightgreen	-0.008 (0.16)	0.0049 (0.42)	-0.0024 (0.66)	0.0055 (0.29)	0.0061 (0.25)	0.01 (0.051)	0.0035 (0.54)	-0.00016 (0.98)	-0.0083 (0.15)	-0.0029 (0.58)	-0.0012 (0.85)	0.0078 (0.14)	-0.0061 (0.24)	-0.00041 (0.94)	-0.0074 (0.51)	0.0039 (0.45)
MEroyalblue	-0.0037 (0.5)	0.0053 (0.38)	-0.00097 (0.86)	0.0058 (0.25)	0.006 (0.26)	0.011 (0.026)	-0.0028 (0.64)	0.0029 (0.56)	-0.012 (0.029)	-0.00068 (0.89)	-0.00036 (0.95)	0.011 (0.034)	-0.011 (0.038)	0.0018 (0.73)	-0.0084 (0.45)	0.0077 (0.12)
MEblack	0.0071 (0.22)	0.013 (0.039)	-0.0017 (0.76)	-0.0064 (0.23)	-0.0046 (0.4)	-0.0032 (0.56)	-0.004 (0.5)	0.0011 (0.84)	0.0015 (0.81)	-0.00071 (0.89)	-0.0041 (0.53)	-0.0024 (0.66)	0.0021 (0.7)	0.00083 (0.88)	0.016 (0.17)	0.0032 (0.54)
MEblue	-0.0026 (0.66)	0.0058 (0.36)	0.00052 (0.93)	-0.0046 (0.4)	-0.0022 (0.7)	6.6e-05 (0.99)	0.005 (0.4)	-0.0025 (0.63)	0.0038 (0.54)	-0.0059 (0.28)	-0.0063 (0.34)	-0.0026 (0.64)	0.0046 (0.41)	-0.0063 (0.26)	0.00049 (0.97)	-0.0075 (0.16)
MElightcyan	0.0017 (0.77)	0.013 (0.045)	-0.0017 (0.77)	-0.0082 (0.13)	-0.0047 (0.4)	0.0011 (0.85)	-0.00046 (0.94)	-0.00048 (0.93)	0.00086 (0.89)	-0.0036 (0.5)	-0.0089 (0.17)	-0.00091 (0.87)	0.0017 (0.76)	-0.0029 (0.6)	0.0046 (0.69)	-0.00032 (0.95)
MEcyan	0.0051 (0.38)	0.014 (0.028)	-0.0011 (0.85)	-0.0065 (0.24)	-0.0059 (0.29)	-0.0048 (0.39)	0.0013 (0.83)	-0.00038 (0.94)	0.0041 (0.5)	-0.0044 (0.42)	-0.0074 (0.26)	-0.0048 (0.39)	0.005 (0.37)	-0.0039 (0.48)	0.015 (0.2)	-0.0025 (0.63)
MEpurple	0.0016 (0.79)	0.013 (0.047)	0.00014 (0.98)	-0.0071 (0.2)	-0.0052 (0.36)	-0.0026 (0.65)	0.0015 (0.81)	-0.00066 (0.9)	0.0032 (0.61)	-0.0056 (0.31)	-0.01 (0.13)	-0.0035 (0.54)	0.004 (0.48)	-0.0057 (0.32)	0.0081 (0.61)	-0.004 (0.48)
	45	POL	4.05h	C**>	20 ⁰	. er.	29 10		200	Contraction of the second	est.	um sat	more	In Sold	aum_na	sum 18
				C	<u>, </u>	c	S	c	,	رهي تر	2.1	~ «	J. L.	e,	-	~

- **Figura 3.** Associação módulo-característica. Módulos e traços (EE^x, AOLf^a, EGSf^b, C14_0^c, C14_1_C9^d, C16_0^e, C16_1_C9^f, C18_0^g, C18_1_C9^h, C18_2_C9_C12ⁱ, C18_2_C9_T11^j, sum_sat^k, sum_mono^l, sum_poly^m, sum_n3ⁿ, sum_n6^o) são representados por cada fila e de cada coluna, respectivamente. O coeficiente de correlação e o valor p são fornecidos em cada célula. Os módulos *Grey, Pink, Red e Royalblue* foram associados às mesmas características.
 - x = Extrato Etéreo
 - a = Área do olho lombo
 - b = Espessura de gordura subcutânea
 - c = ácido mirístico
 - d = ácido miristoleico
 - e = ácido palmítico
 - f = ácido palmitoleico
 - g = ácido esteárico
 - h = ácido oleico
 - i = ácido linoleico
 - j = ácido vacênico
 - k = Soma de ácido graxo saturado
 - l = Soma de ácido graxo monoinsaturado
 - m = Soma de ácido graxo poliinsaturado
 - n = Soma de ômega-3
 - o = Soma de ômega-6

3.4. Overlap entre ATAC-Seq peaks e genes expressos no músculo esquelético

Para analisar como ATAC-Seq *peaks* são distribuídos nos genes expressos no tecido muscular, sobrepusemos todos os 18.942 ATAC-Seq *peaks* a todas as regiões regulatórias (2 Kb

upstream e downtream aos TSSs) dos 10.117 genes expressos no músculo esquelético. No total, 6.481 ATAC-Seq *peaks* podem ser encontrados em 8.524 genes.

3.5. Identificação de sequências motifs em genes co-expressos

A fim de estudar apenas as regiões regulatórias que pertencem aos genes co-expressos de cada módulo, foi considerado um arquivo contendo a sequência de DNA 2.000 pares de bases *apstream* e 2.000 pares de bases a *downstream* ao TSS de cada gene. Em seguida, os *motifs* genéticos foram identificados pelo programa MEME. A descoberta do motivo foi projetada para encontrar 8 logotipos (*motifs*) com um comprimento mínimo e máximo de 4 e 12 nucleotídeos, respectivamente, em cada módulo genético co-expresso. Os *outputs* do *Motif Discovery* são liberadas na forma de Logotipos e arquivos que especificam as estimativas estatísticas. Cada logotipo apresenta um valor calculado para indicar a probabilidade de o *motif* não ser apenas um artefato estatístico. Além disso, cada sequência pode conter qualquer número de ocorrências não sobrepostas.

3.6. Ocorrência das sequências motifs em banco de dados

O *output* obtido por meio da ferramenta FIMO que descreveu as melhores correspondências possíveis entre os *motifs* descobertos pelo MEME e os *motifs* anotados nos bancos de dados é descrita na Tabela 1. Apenas os *motifs* correspondentes aos descritos em bancos de dados com *p-values* inferiores a 0,0001 foram descritos. Os *p-values* inferiores a 0,0001 indicaram que todos os *motifs* identificados tinham maior probabilidade de corresponder a um determinado local genético anotado quando comparados aos *motifs* aleatórios testados pelo programa. As estimativas dos q-values indicaram que houve taxas insignificativas de descobertas falsas. O módulo *Blaek* apresentou 17.193 ocorrências de sequências de motivos, enquanto os módulos *Blue e Brown* apresentaram 48.295 e 70.208 ocorrências, respectivamente. O módulo *Cyan* apresentava 5.375 motivos correspondentes e o módulo *Green*, 28.628. Os módulos *Greenyellony*, *Grey, Lighteyan e Lightgreen* apresentaram, respectivamente, 5.910, 3.804, 5.902, 3.172 representações de motivos. O módulo Lightyellow contou com 2.665 partidas e o módulo Magenta, 18.942. Os módulos *Midnightblue e Pink* apresentaram 2.927 e 17.308 ocorrências de motivos, respectivamente. Os módulos *Purple*, *Red e Royalblue* tiveram respectivamente 14.058, 32.545 e 1.823 correspondências entre os *motifs*. Por fim, os módulos *Salmon, Tan, Turquoise e*

Yelllow apresentaram 5.460, 9.082, 253.093 e 34.580 motivos correspondentes aos já mencionados na literatura, respectivamente.

Motif (MEME)	Best Possible Match (database)	Motif (MEME)	Best Possible Match (database)			
Module	Black	Module Blue				
CSCSGCCSCSGC	CCCCGCCGCGGC	TTTTYWYTTTYW	ттттттттст			
RGRRRAGRARRG	GGAAGAGGAAAG	TTCTTGCCTGGA	TTCTTGCCTGGA			
TGTGTGTGTGTG	TGTGTGTGTGTG	GCCTGSYRGGCT	GCCTGGCAGGCT			
WTTTWTWTWTWW	TTTTATTTTTT	ACCYWCTCCAGT	ACCCACTCCAGT			
CHGTCCATGGGR	CAGTCCATGGGG	GTCCATGGGRT	GTCCATGGGGT			
GGACACGACTGA	GGACACGACTGA	GGACACGACTGA	GGACACGACTGA			
RAYCCCATGGAC	GACCCCATGGAC	RATCCCATGGA	AATCCCATGGA			
GGRGHGGG	GGGGTGGG	AARRAAAWRRAA	AAGAAAAAGAAA			
Module	Brown	Modul	e Cyan			
GGGGCGGGGSVG	GGGGCGGGGCGG	WAAAAAAWAAA	ΑΑΑΑΑΑΑΑΑΑΑ			
TGGAGWGGGDDG	TGGAGTGGGTTG	TTTTWWTTTTWW	ΤΤΤΤΤΤΤΤΤΑΑ			
TCCATGGGRTY	TCCATGGGATC	GGVGGVGGGGGG	GGAGGCGGGGGG			
GTGTCCGACTCT	GTGTCCGACTCT	CCTCCYCYBCCC	CCTCCTCCTCCC			
AGAGTCGGACAC	AGAGTCGGACAC	TTCTCCAGGCAA	TTCTCCAGGCAA			
AAAAAAWAAAA	ΑΑΑΑΑΑΑΑΑΑ	AGAGTCRGACAC	AGAGTCGGACAC			
RAYCCCATGGAC	GACCCCATGGAC	RAYCCCATGGAC	GACCCCATGGAC			
CCCSCCCCYCC	CCCCCCCCTCC	AAATGGCAACCC	AAATGGCAACCC			
Module	Green	Module Greenyellow				
RGARARRGARR	AGAGAAAGAAG	SCCCCRCCCYS	CCCCCGCCCCC			
GTGTGTGTGTGT	GTGTGTGTGTGT	GGGGGGGGGGG	GGGGGGGCGGGG			
CCYYCCCTYCCC	CCTCCCCTCCCC	WTTTTWTTTTT	ТТТТТТТТТТТТ			
YCCCAKGGACWG	CCCCATGGACTG	CACACACACACA	CACACACACACA			
GGGGSKGGRGGG	GGGGCGGGGGGG	CCCAGG	CCCAGG			
TTTTWTTTTTWW	ΤΤΤΤΤΤΤΤΤΤΤΤ	WKAAAAADRAAA	ΤΤΑΑΑΑΑΑΑΑΑ			
TTGCCTGGAGAA	TTGCCTGGAGAA	CATGGGGTCGCA	CATGGGGTCGCA			
TYCTCCAGGSAA	TTCTCCAGGGAA	AGAATACTGGAG	AGAATACTGGAG			
Module	Grey	Module I	ightcyan			
TTTCYTTYTYYT	TTTCCTTTTCTT	SSCGGSSSCGCS	CCCGGCCCCGCC			
AAAAAAAARAAA	ΑΑΑΑΑΑΑΑΑΑΑ	WYTTTTTTTT	ATTTTTTTTTT			
GGRGRAGGRRR	GGGGGAGGGGG	WRAAADWRAAA	AAAAAATGAAA			
CCCCHCCCCYCC	CCCCTCCCTCC	GGACACGACTGA	GGACACGACTGA			
YYHCCAGGSAAG	TCACCAGGCAAG	TTSCCTGGARAA	TTCCCTGGAGAA			
AYCCCAKGGAY	ATCCCATGGAC	GGGGGWGGGGG	GGGGGAGGGGG			
GTATTCTTGCCT	GTATTCTTGCCT	CCHBBCCCYTCY	ссстссссстст			

Tabela 1. As melhores correspondências entre os motifs MEME e os motifs anotados nos bancos de dados

AGTCCATGGGRT	AGTCCATGGGGT	ACCCACTCCAGT	ACCCACTCCAGT			
Module Li	ghtgreen	Module Lightyellow				
GGSGGGGSHGGG	GGCGGGGGAGGG	AAAAAAWAAAA	ΑΑΑΑΑΑΑΑΑ			
CCCCGCCCCSC	CCCCGCCCCGC	GGRGGMGGGGR	GGGGGAGGGGA			
TTTTHTTYTTT	TTTTCTTTTTT	WWTTTWHTTTTW	TTTTTTTTTTTT			
TCCATGGGRTY	TCCATGGGGTC	YCCCWBCCCCMS	CCCCTGCCCAG			
GACAYGACTGAG	GACATGACTGAG	RAYYCCATGGAC	AACCCCATGGAC			
RATCCCATGGAC	AATCCCATGGAC	TCTCCTGCATTG	TCTCCTGCATTG			
WTTTTWAAAWKW	TTTTTAAAAATT	TMCTGGAGTGGG	TACTGGAGTGGG			
AWRAAAAADAAA	ΑΑΑΑΑΑΑΑΑΑΑ	GCCTGGYRGGCT	GCCTGGCAGGCT			
Module N	/agenta	Module Mi	dnightblue			
SSCSGSSCSSGS	CCCGGCGCCGGG	GGRSHRGG	GGGGAGGG			
TTTCYTTYTYT	TTTCTTTCTTT	TTTTTWWTTTTW	TTTTTTTTTTTT			
WKAAAAAAAAA	ΤΤΑΑΑΑΑΑΑΑΑ	CCYCCCYCYCCC	ССТСССССТССС			
CWYCCCHCYCCM	CTCCCCACTCCA	GACACGACTGAG	GACACGACTGAG			
RGTGGGYTGCCR	AGTGGGTTGCCA	RAAAAARAAARA	AAAAAAGAAAAA			
TCCAGGCAAGAA	TCCAGGCAAGAA	GGGATTYTCCAG	GGGATTTTCCAG			
RAYCCCATGGAC	AATCCCATGGAC	CCCWTGGACTGC	CCCATGGACTGC			
GGAGRRGG	GGAGGAGG	GTCRCAAAGAGT	GTCGCAAAGAGT			
Module	e Pink	Module Purple				
GSSCGGSGSSGG	GGGCGGGGGCCGG	ΑΑΑΑΑΑΑΑΑΑ	ΑΑΑΑΑΑΑΑΑΑΑ			
CCHBCCCCYCC	CCCTCCCCTCC	ΤΥΤΥΥΤΥΥΤΥΥΤ	TCTCCTTTTCTT			
CTCCAGGSAAGM	CTCCAGGCAAGA	CCCCGCCCCCC	CCCCGCCCCCC			
TGTCCGACTCTT	TGTCCGACTCTT	ACACACACACAC	ACACACACACAC			
TTTTWTTTTTYT	TTTTTTTTTTCT	GGVGGGGGMGGG	GGAGGGGGCGGG			
GRARRAGAARR	GGAGGAGAAAG	RAYCCCATGGAC	AATCCCATGGAC			
GTCCATGGGRTY	GTCCATGGGATT	GTCCATGGGRTY	GTCCATGGGGTC			
TACTGGAGTGGG	TACTGGAGTGGG	CTGCCTSTGATT	CTGCCTGTGATT			
Modul	e Red	Module Royalblue				
ΤΤΤΤΥΤΤΥΤΤΥ	TTTTCTTTTTT	GSCGGCGSCGSS	GCCGGCGGCGGC			
AAAAAARAAAR	AAAAAAAAAAG	RAYCCCATGGAC	AATCCCATGGAC			
ССССКСССССС	222222222222	TTTTWTTTYTTT	ТТТТТТТТТТТТТ			
GGRGGWGGGGR	GGGGGAGGGGG	ΑΑΑΑΑΑΑΑΑ	ΑΑΑΑΑΑΑΑΑ			
RAYCCCATGGAC	AATCCCATGGAC	ACCCACTCCAGT	ACCCACTCCAGT			
ACTCCAGTATT	ACTCCAGTATT	GACACGACTGAG	GACACGACTGAG			
WWWTAWWTWTW	AAATATATTTT	GTCCATGGGRTY	GTCCATGGGGTC			
SCSGSGSCGGCC	GCGGCGGCGGCC	AGARTACTGGAG	AGAATACTGGAG			
Module	Salmon	Modu	le Tan			
SSSSSCGSSSCS	GCGGGCGCCGCC	CSSCGSCSSSSS	CCCCGGCGGCGC			
ACACACACACAC	ACACACACACAC	YTTYYYTCYYY	сттстстстст			
GRGRARGRGRAG	GGGAAAGAGAAG	AAAADAAAAAA	ΑΑΑΑΑΑΑΑΑΑ			
ССҮҮҮТССТҮСҮ	ссттстсстссс	GGAGGRGGSGGG	GGAGGGGGGGGG			
GTCCATGGGRTY	GTCCATGGGGTT	TTSCCTGGARAA	TTGCCTGGAGAA			
	тттттттт	GGGTBGSRAAGA	GGGTCGCAAAGA			

TTSCYTGGAAAA	TTGCCTGGAAAA	CCCATGGACWG	CCCATGGACTG			
Module T	urquoise	Module Yellow				
GGCGGCGGCGGC	GGCGGCGGCGGC	AWRAAARWAAAA	ΑΑΑΑΑΑΑΑΑΑΑ			
GCCCCGCCCCC	GCCCCGCCCCC	CCCCCDCCCCC	ССССТССССС			
GGGGSCGGGCGS	GGGGGCGGGCGC	ТТТТТТТТТТТ	ТТТТТТТТТТТ			
GGACACGACTGA	GGACACGACTGA	GSTGGAGRRGG	GGTGGAGAGGG			
CCSCCGCCSCC	CCCCCGCCCCC	CWHCCYWCTCCA	CAACCCACTCCA			
CCATGGACWGAR	CCATGGACTGAA	RAYCCCATGGAC	AATCCCATGGAC			
GGCAACCCACTC	GGCAACCCACTC	AAAGAGTCKGAC	AAAGAGTCGGAC			
SCTYCYCTKYCC	GCTCCTCTGTCC	TCCGACTCTTTG	TCCGACTCTTTG			

3.7. Identificação de sequências motifs em regiões de ATAC-Seq peaks

Dentro das regiões de ATAC-Seq *peaks*, a ferramenta SpaMo foi empregada para identificar locais de ligação específicos para *motifs* descobertos pela ferramenta MEME nos 5 módulos associados à deposição de gordura intramuscular (*motifs* primários) e aqueles descritos no banco de dados HOCOMOCOv11_full_HUMAN (*motifs* secundários)¹⁰⁰. Em outras palavras, o programa SpaMo pesquisa *motifs* secundários inferidos nas regiões ATAC-Seq, sobrepondo os TFBSs mais significativos revelados pela ferramenta MEME com os locais de ligação anotados para fatores de transcrição específicos para os *motifs* secundários. No total, os 4 módulos (*Grey, Pink, Red e Royalblue*) associados às mesmas características e o módulo *Green* correlacionados com EE apresentaram, respectivamente, 498 e 331 *motifs* secundários não redundantes (inferidos) que também podem ser encontrados nos ATAC-Seq *peaks* (Tabela 2a e Tabela 2b).

Para saber em quais ATAC-Seq *peaks* estão localizados os 32 motivos (8 motivos descobertos para cada um dos 4 módulos *Grey, Pink, Red* e *Royalblue* associados aos mesmos fenótipos), sobrepusemos a sequência fasta de todos os ATAC-Seq *peaks* com as sequências consenso dos *motifs*. Encontramos 14.685 ATAC-seq *peaks* dentro dos quais os TFBSs estão posicionados. Da mesma forma, comparamos a sequência fasta dos ATAC-Seq *peaks* com a sequência consenso dos *motifs* descobertos no módulo *Green* e os resultados mostraram que todas as 8 sequências dos *motifs* pertencentes ao módulo *Green* estão distribuídas entre 673 ATAC-Seq *peaks*.

INFERRED SECONDARY MOTIFS							
AHR	ESR1	HXA2	NFAC1	RARG	THAP1		
ALX1	ESR2	HXA5	NFAC2	RAX2	THB		
ANDR	ESX1	HXA7	NFAC3	REL	TLX1		
AP2A	ETS1	HXB2	NFAC4	RELB	TWST1		
AP2B	ETS2	HXB3	NFAT5	REST	TYY1		
AP2C	ETV1	HXB4	NFE2	RFX1	TYY2		
ARI3A	ETV2	HXC10	NFIA	RFX2	UBIP1		
ARI5B	ETV3	HXC11	NFIB	RFX4	UNC4		
ARNT	ETV4	HXC12	NFIC	RFX5	USF2		
ARNT2	ETV5	HXC13	NFIL3	RHXF1	VAX1		
ASCL1	ETV6	HXC8	NFKB1	RORA	VAX2		
ATF1	ETV7	HXD10	NFKB2	RORG	VDR		
ATF2	FEV	HXD11	NFYA	RREB1	VENTX		
ATF3	FEZF1	HXD12	NFYB	RUNX1	VEZF1		
ATF6A	FIGLA	HXD13	NFYC	RUNX2	VSX1		
BACH1	FLI1	HXD3	NKX21	RUNX3	WT1		
BACH2	FOS	HXD4	NKX22	RX	XBP1		
BARH1	FOSB	HXD8	NKX23	RXRA	Z324A		
BARH2	FOSL1	HXD9	NKX25	RXRB	Z354A		
BATF	FOXA1	IKZF1	NKX31	SALL4	ZBT17		
BATF3	FOXA2	IRF1	NKX61	SCRT1	ZBT18		
BC11A	FOXB1	IRF2	NKX62	SCRT2	ZBT48		
BCL6	FOXC1	IRF3	NOBOX	SIX1	ZBT7A		
BHA15	FOXC2	IRF4	NR0B1	SIX2	ZBT7B		
BHE23	FOXD1	IRF5	NR1D1	SMAD1	ZBTB4		
BPTF	FOXD3	IRF7	NR1H3	SMAD2	ZBTB6		
BRAC	FOXG1	IRF8	NR1H4	SMAD3	ZEB1		
BRCA1	FOXH1	IRF9	NR1I2	SMAD4	ZEP1		
BSH	FOXI1	IRX2	NR1I3	SMCA1	ZEP2		
CDX1	FOXJ2	IRX3	NR2C1	SMCA5	ZF64A		

Tabela 2. a) Nome dos *motifs* secundários inferidos identificados nos 4 módulos (*Grey, Pink, Red and Royalblue*) e nos ATAC-Seq *peaks*.

CDX2	FOXJ3	ISL2	NR2C2	SOX1	ZFP42
CEBPA	FOXM1	ITF2	NR2E1	SOX10	ZFP82
CEBPB	FOXO1	JUN	NR4A1	SOX13	ZFX
CEBPD	FOXO3	JUNB	NR4A2	SOX15	ZIC1
CEBPE	FOXO4	JUND	NR5A2	SOX17	ZIC2
CEBPG	FOXP1	KAISO	NR6A1	SOX2	ZIC3
CEBPZ	FOXP2	KLF1	OLIG2	SOX21	ZIC4
CLOCK	FOXQ1	KLF12	ONEC3	SOX3	ZIM3
COE1	FUBP1	KLF14	OSR2	SOX4	ZKSC1
COT1	GABPA	KLF15	OTX2	SOX5	ZKSC3
COT2	GATA2	KLF16	OVOL1	SOX8	ZN134
CPEB1	GATA3	KLF3	P53	SOX9	ZN136
CREB1	GATA4	KLF4	P63	SP1	ZN148
CREB3	GATA5	KLF5	P73	SP2	ZN219
CREM	GBX1	KLF6	PATZ1	SP3	ZN232
CRX	GBX2	KLF8	PAX2	SP4	ZN257
CTCFL	GCM1	KLF9	PAX4	SPDEF	ZN263
CUX1	GCR	LEF1	PAX5	SPI1	ZN274
CUX2	GFI1	LHX2	PAX6	SPIB	ZN281
CXXC1	GLI1	LHX4	PAX8	SPIC	ZN282
DDIT3	GLI3	LHX9	PBX1	SPZ1	ZN322
DLX1	GLIS3	LMX1A	PBX3	SRBP1	ZN331
DLX2	GMEB2	LMX1B	PDX1	SRBP2	ZN335
DLX3	GRHL2	LYL1	PEBB	SRF	ZN341
DLX4	GSC2	MAF	PIT1	SRY	ZN350
DLX5	GSX1	MAFA	PITX1	STA5B	ZN384
DLX6	GSX2	MAFB	PITX2	STAT1	ZN394
DMBX1	HAND1	MAFF	PITX3	STAT2	ZN418
DMRT1	HEN1	MAFG	PKNX1	STAT3	ZN436
DRGX	HES1	MAFK	PLAG1	STAT4	ZN449
E2F1	HESX1	MAX	PLAL1	STAT6	ZN467
E2F3	HEY2	MAZ	PO2F1	STF1	ZN524

E2F4	HIC1	MECP2	PO2F2	SUH	ZN528
E2F6	HIC2	MEF2A	PO2F3	TAF1	ZN547
E2F7	HIF1A	MEF2B	PO3F1	TAL1	ZN549
E2F8	HLTF	MESP1	PO3F2	TBP	ZN554
EGR1	HMBX1	MGAP	PO3F3	TBR1	ZN563
EGR2	HME1	MITF	PO4F3	TBX1	ZN589
EGR3	HME2	MIXL1	PO5F1	TBX15	ZN652
EGR4	HMGA1	MNX1	PPARA	TBX21	ZN667
EHF	HMX2	MSX1	PPARD	TBX3	ZN708
ELF1	HMX3	MSX2	PPARG	TBX4	ZN713
ELF2	HNF1A	MXI1	PRD14	TCF7	ZN740
ELF3	HNF1B	MYB	PRDM1	TEAD1	ZN768
ELF5	HNF4A	MYC	PRDM6	TEAD2	ZN770
ELK1	HNF4G	MYCN	PRGR	TEAD4	ZN784
ELK3	HSF1	MYNN	PROP1	TF65	ZN816
ELK4	HSF2	MYOD1	PROX1	TF7L1	ZNF18
EMX1	HSF4	MYOG	PRRX1	TF7L2	ZNF41
EOMES	HSFY1	MZF1	PTF1A	TFCP2	ZNF76
EPAS1	HTF4	NANOG	PURA	TFDP1	ZNF85
ERG	HXA10	NF2L1	RARA	TGIF1	ZSC22
ERR1	HXA11	NF2L2	RARB	THA	ZSC31

Tabela 2. b) Nome dos motifs secundários inferidos identificados no módulo Green e nos ATAC-Seq peaks

Motifs secundários inferidos							
ZEP1	KLF16	AHR	STA5B	NFE2	ZN528		
P53	ZN467	TWST1	STAT4	BACH2	LYL1		
NFKB2	ZN219	ZN263	COE1	NFIC	ZN554		
NFKB1	SALL4	KLF4	ETV2	PRGR	NFIA		
NKX23	COT2	PITX1	EHF	PLAL1	HNF4G		
HIC1	HAND1	FOXK1	ZN436	PAX5	ZKSC1		
EGR3	SPIC	ZN770	TGIF1	NF2L1	JUNB		
PAX8	FEV	EGR1	SPI1	RELB	FOS		

VEZF1	SOX10	STF1	ETS1	HES1	NF2L2
ARI5B	SP2	E2F1	IRF4	STAT1	ZBT7A
NKX21	KLF6	PPARG	SPIB	IRF2	TEAD2
TF65	SMAD3	HME2	GABPA	HIF1A	ZN667
ZN274	BCL6	ZNF41	ELF2	ATF1	ZBT48
RUNX2	SP1	FOXC1	IRF8	GLI1	OTX2
NKX25	ZEB1	BSH	ELK3	NR1H3	TFCP2
P73	PROX1	ETS2	ELK4	NFAC1	ZN549
HMGA1	HXC8	SRY	ELK1	MXI1	ZN524
SPDEF	KLF5	E2F6	ESR2	ZNF85	ZN784
REL	SMCA5	TBX1	PBX1	ZN563	REST
PITX2	KLF9	PAX1	SUH	GRHL2	E2F8
NKX22	PRDM6	ZN121	MESP1	ZEP2	IKZF1
SRBP2	ZN281	NANOG	HTF4	ATF3	SRF
BRCA1	ZN713	BHA15	ARNT	PTF1A	HLTF
ZN257	MCR	E2F4	ZN816	TEAD4	UBIP1
NFAC3	KLF1	FOXP2	RXRB	GLIS3	NR1H4
MAZ	HXD13	EGR4	THA	KLF8	RARA
NFAC2	SOX3	SIX2	RARB	ALX1	ZF64A
ETV7	IRF3	GATA5	THB	IRF7	CTCFL
NFAT5	HSF1	PAX2	NR1I2	LMX1A	LEF1
ELF5	PPARA	SMAD2	VDR	HXB3	AP2C
TBX15	GATA3	BACH1	NR1I3	KLF14	OSR2
SOX5	BC11A	PBX3	RARG	NR0B1	NDF1
PATZ1	ZN350	PKNX1	ZKSC3	FOXA1	MAFB
ZN341	PPARD	HESX1	STAT6	NR5A2	HXC13
ETV6	ZN740	FOSB	ARNT2	PLAG1	SRBP1
ZN148	BARX1	FOSL1	ZFX	HNF4A	LHX2
PURA	CUX2	SMAD4	HMBX1	CDX1	TLX1
SP3	ITF2	ZN136	KAISO	FOXD3	NFYC
ZN394	ZIC1	RFX2	RFX1	MEF2A	BATF
SMCA1	TFDP1	SOX9	E2F3	PO3F3	MSX2

AP2A	E2F7	ZFP42	PO5F1	HNF1B	OVOL1
KLF15	FOXO1	NKX61	ESR1	HXD11	ZN652
EGR2	FOXO3	TYY2	GCR	HXA11	RFX4
TAF1	GFI1	NR4A1	ANDR	HXD12	HXC12
FLI1	ETV4	HXB4	SOX4	HSFY1	SPZ1
SP4	ELF1	COT1	ONEC3	MYB	PEBB
ZSC22	KLF3	FEZF1	NFIB	HXC10	RUNX3
ETV5	ERG	AP2B	JUN	CDX2	RUNX1
WT1	KLF12	MZF1	JUND	PO2F1	ZIC2
P63	ELF3	RHXF1	MAFG	PAX6	NOBOX
ZBT17	PITX3	STAT3	MAFF	RFX5	NKX32

3.8. Módulos de interesse

Com base nos resultados obtidos no item anterior, organizamos os dados dos 5 módulos para analisar em quais genes as sequências consenso dos *motifs* situados nos ATAC-Seq *peaks* se sobrepõem. No total, 8.236 genes pertencentes ao módulo *Grey, Pink, Red e Royalblue* contêm TFBSs localizados em 6.213 ATAC-Seq *peaks* e 8.001 genes pertencentes ao módulo *Green* associado a EE contêm TFBSs localizado em 5.985 ATAC -Seq *peaks*.

3.9. Descrição dos fatores de transcrição identificados

Selecionamos 5 fatores de transcrição descobertos pelo SpaMo e específicos para os 5 módulos que, além de anotados para *Homo sapiens*, também apresentam descrição para *Bos taurus*.

A proteína *Early Growth Response Protein 1* (UniprotKB - Q29W20), conhecida como EGR1, é um regulador transcricional que reconhece e se liga à sequência de DNA 5'-GCG (T / G) GGGCG-3 '. É responsável por regular a resposta a fatores de crescimento e é necessária para o processo normal de mitose e para a biossíntese do hormônio luteinizante (LHB) na hipófise.

Os membros *Runt-related transcription factor (*RUNX1) (UniprotKB - E1BAD4) são proteínas que controlam a transcrição de seus genes alvo por meio do reconhecimento das sequências 5'-TGTGGT-3 ' e 5'-TGCGGT-3', dentro de suas regiões regulatórias

A proteína NFKB1 (*Nuclear factor kappa B subunit 1*) (UniProtKB - F1MKW9) possui fator transcricional e ligação à cromatina como função molecular. Esse fator de transcrição atua no controle do sistema imunológico, modulando as respostas inflamatórias.

O fator de transcrição PPARG (UniProtKB - O18971) é uma proteína que controla a via da betaoxidação. Também regula a diferenciação de adipócitos e a homeostase da glicose no organismo.

Forkhead box protein 01 (FOXO1) (UniProtKB - E1BPQ1) é uma proteína que medeia a ação da insulina no tecido adiposo e regula a expressão de genes como o PPARG durante a diferenciação dos pré-adipócitos, o tamanho dos adipócitos e a expressão gênica específica do tecido adiposo.

4. DISCUSSÃO

Os objetivos propostos por esse trabalho permitiram que análises exploratórias fossem realizadas a fim de se predizer *motifs* e fatores de transcrição associados à deposição de gordura intramuscular. Alguns dados precisaram ser extrapolados para o genoma referência *Homo sapiens* devido à escassez de anotação de elementos regulatórios pertencentes ao genoma *Bos taurus*.

Para garantir que os resultados fossem coerentes com o tema abordado nessa pesquisa precisávamos, primeiramente, entender se os ATAC-Seq peaks encontrados realmente representavam regiões de eucromatina. A melhor maneira de obtermos essa resposta era analisando se os ATAC-Seq *peaks* coincidiam com as regiões regulatórias dos genes expressos já anotados para Bos taurus. Por meio do gráfico heatmap (Figura 1) podemos visualizar que as regiões gênicas - cujos TSSs foram considerados ponto central de referência para que 2000 pares de base upstream e 2000 pares de base downstream fossem estendidos - foram enriquecidas pelas regiões de cromatina aberta isoladas pela metodologia ATAC-seq. Visto que a região promotora principal (core promoter) distancia-se do TSS em aproximadamente 50 pares de base¹⁰², filtrar as regiões regulatórias de acordo com a localização do sitio de início de transcrição é uma estratégia adequada para que encontremos informações relacionadas à maquinaria transcricional. Assim sendo, nossos resultados reforçaram a ideia de que as regiões de ATAC-Seq peaks estão associadas aos TSSs abrangendo, possivelmente, elementos regulatórios cis-acting nas distâncias de pelo menos 1500 pares de base upstream e 1500 pares de base downstream ao sítio ativo de início da transcrição^{103,104}. A literatura descreve que a função dos elementos regulatórios pode ser modulada tanto pelos promotores quanto pelos enhancers uma vez que o primeiro grupo determina o início da transcrição e o segundo, aumenta o nível transcricional^{105,106}.

A definição de módulos de genes co-expressos auxiliou na busca por *motifs* genéticos uma vez que os resultados indicaram quais genes são co-expressos e, portanto, compartilham elementos regulatórios associados à expressão gênica. A escolha por essa ferramenta foi devido à necessidade de nortear nossa procura por *motifs* de acordo com algum critério molecular que fizesse sentido. Ao predizer quais *motifs* estão super-representados em genes que carregam similaridade molecular de expressão, podemos apontar quais são os elementos regulatórios associados ao mecanismo transcricional de determinada característica. Quando obtidos, esses *motifs* podem ser pesquisados dentro das regiões de eucromatina a fim de elucidar a localização dos elementos *cis-acting*. O programa WGCNA realiza uma *clusterização* hierárquica a partir da qual é possível definir e encontrar grupos de objetos (no nosso caso, genes) que interagem e que compartilham similaridades funcionais. Para identificar as ramificações (*branches*), utiliza o *'dynamic* *tree cut*' que estabelece condições para que *cluster*s sejam agrupados ou definitivamente separados. Toda a análise se baseia em construir *networks* por meio dos padrões de interação gênica entendendo, portanto, o sistema biológico e não somente as partes individuais que o compõe¹⁰⁷. Assim sendo, os 20 módulos identificados (Figura 2) ajudaram a direcionar nossas buscas por *motifs* presentes no genoma Nelore.

Os logos correspondentes ao *motifs* encontrados para cada módulo reportaram a significância estatística de cada *motif* encontrado (*e-value*) e a estimativa da precisão da ocorrência *motif* (*p-value*). Quanto menor o *e-value* maior é a confiança de predição do *motif*. Os resultados do *Motif Discovery* são liberados em formato de Logotipos (logos) e de arquivos que especificam as estimativas estatísticas. O termo *bits* (entropia) corresponde à soma do conteúdo informacional das colunas do logotipo. Nos logos uma única base só é mostrada caso ocorra em mais da metade dos sítios genéticos e pelo menos duas vezes mais que a segunda base mais frequente. Grande parte dos nossos resultados apresentaram baixos *e-values* e *bits* com valores entre 1 e 2 demonstrando, assim, a otimização determinística dos *motifs* encontrados, ou seja, apenas os *motifs* genéticos com maior probabilidade de ocorrer, foram considerados.

Após identificarmos os 160 *motifs* nos conjuntos de genes co-expressos (8 *motifs* para cada um dos 20 módulos), precisávamos saber se esses sítios de ligação à fatores de transcrição correspondiam a algum *motif* já descrito pela literatura. Para isso a ferramenta utilizada precisa ser capaz de considerar todos os possíveis viés existentes entre os dois dados e que possam interferir no *match* final¹⁰⁸. Embora a quantidade de TFBSs coincidentes entre o nosso arquivo *input* e os bancos de dados tenha sido relativa para cada *motif* descoberto, pudemos comparar e associar diferentes funções e atividades moleculares à cada *motif* analisado. Não é preciso que todas as bases alocadas nas posições genéticas sejam idênticas entre o grupo de comparação. O software utiliza estimativas estatísticas que determinam o limite mínimo aceitável para que um *motif* possa ser considerado semelhante a algum outro.

Os *motifs* direcionam a ligação entre os fatores de transcrição e sítios genéticos uma vez que determinam a especificidade dessas proteínas aos domínios de ligação ao DNA (DBD), responsáveis por efetivar essa interação molecular. No entanto, a maquinaria transcricional também é composta por interações proteína-proteína indispensáveis para a integração de sinais biológicos¹⁰⁹. As interações entre dois fatores de transcrição culminam na formação de dímeros que, ao serem alvos de *loopings* genéticos, podem resultar em um grande complexo proteico. O espaçamento existente entre os *motifs* e seus respectivos fatores de transcrição-DBD é rígido visto que qualquer perturbação bioquímica ocasiona disruptura na interação proteica. Assim sendo, esses espaçamentos representam padrões regulatórios importantes na busca por elementos

distais^{110,111}. A análise de espaçamento foi realizada por meio do arquivo ATAC-Seq *peaks (input)*, *motifs* analisados pelo MEME (*motif* primário) e *motifs* descritos em bancos de dados (*motif* secundário) (Tabela 2). Nossos resultados indicaram que houve espaçamento de *motifs* estatisticamente significativos quando o primário e o secundário foram idênticos ou semelhantes entre si (homodímeros), mas também quando ambos se diferiram totalmente. Essa observação é válida visto que direciona nossas ideias sobre a ocorrência de homodímeros nas regiões de eucromatina ou a presença de apenas 1 fator de transcrição de interesse enriquecido nos ATAC-Seq *peaks*.

Para detalharmos a localização das sequências *motifs* (TFBSs) em regiões promotoras já descritas e anotadas em bancos de dados utilizamos a ferramenta FIMO para a qual o parâmetro *'--parse-genomic – coord'*, foi determinado. Assim sendo, as posições *start e end* estão em coordenadas baseadas em 0 em relação ao início da localização genética dos *motifs*. Já os q-values nos mostraram que houve taxas irrisórias de falsas descobertas.

Todos os *motifs* identificados e descritos nesse trabalho apresentaram algum tipo de função biológica associada ao evento transcricional. A maioria deles atua regulando positivamente a expressão de determinados genes reforçando, portanto, a importância genética que desempenham quando ativados por fatores de transcrição.

Os fatores de transcrição EGR1 e EGR4, além de desempenharem um papel nas funções adipocitárias, também interagem com as vias PI3K / Akt e Erk / MAPK, modulando o sistema glicêmico e o evento inflamatório^{112,113}. A proteína RUNX1 está associada ao fenótipo da obesidade, uma vez que desempenha um papel importante na regulação da adipogênese marrom. Além disso, esse TF está envolvido em eventos epigenéticos associados à característica de ácidos graxos ômega-3¹¹⁴⁻¹¹⁶. Tanto o EGR quanto o RUNX1 também são encontrados em regiões *hotspot* que afetam a expressão de genes correlacionados ao FMI, associados ao metabolismo lipídico^{92,117}.

O fator de transcrição NFKB, identificado como uma das proteínas de ligação de motivos encontradas e associadas aos fenótipos intramusculares de deposição de gordura, corresponde a uma molécula de ligação de potenciador associada à regulação de respostas próinflamatórias e anti-inflamatórias⁶⁴. No entanto, estudos indicaram que esta proteína está diretamente associada à adipogênese, uma vez que os sinais regulatórios correlacionados a esse evento são mediados pelo adenosina monofosfato 3'-5'-cíclico (cAMP), uma molécula que atua como mensageiro secundário celular e tem papel significativo na adipose e tecido muscular¹¹⁸. A importância do cAMP na contração do músculo esquelético aumenta os níveis dessa molécula que estimula a proteína quinase A (PKA) que ativa a cascata de fosforilação e as proteínas transcricionais associadas à adipogênese, incluindo o receptor γ ativado por proliferador de peroxissomo (PPAR- γ) e o CCAAT / proteína de ligação ao intensificador (C / EBP) α e $\beta^{118,119}$.

A expressão gênica da insulina é modulada por vários fatores de transcrição em interação, formando complexas redes de proteínas. Alguns estudos indicam que o ISL atua junto com PDX1 (fator de transcrição identificado neste estudo) e BETA2, que, ao se ligarem a elementos de ação cis até 410 pares de bases a montante do TSS do gene responsável - mais precisamente no site E2 de o promotor - regula a expressão gênica associada ao nível de glicose no sangue^{120,121}. Sabe-se que o acúmulo excessivo de gordura intramuscular interfere na resistência muscular à insulina. Quanto maior a concentração de lipídios nos músculos, menor a captação de glicose e, consequentemente, há inibição da enzima glicogênio sintetase mediada por insulina e hormônio. Além disso, o acúmulo de lipídios entre os feixes musculares tende a dificultar a difusão da insulina no tecido, comprometendo sua ação^{122,123}.

O *Peroxisome proliferator-activated receptor* (PPAR) desempenha um papel fundamental no controle do conteúdo de gordura intramuscular e na diferenciação de adipócitos. Esse fator de transcrição, quando ligado a um motivo específico, regula o processo de adipogênese e lipogênese, auxiliando na expressão gênica associada ao metabolismo intramuscular da deposição de gordura^{124,125}.

FOXO1 é uma proteína que atua como fator de transcrição uma vez que promove a expressão de genes associados à gliconeogênese⁶⁴. Estudos mostram que a proteína FOXO1 modula os níveis de expressão do receptor α do fator de crescimento derivado de plaquetas α (PDGFR α) - expresso nos tecidos adiposos - responsável pela deposição do FMI ao ativar a via de sinalização Erk e, portanto, a depleção desse fator de transcrição inibe a lipogênese ao desregular o PDGFR $\alpha^{126,127}$.

Selecionamos, também, alguns genes identificados nos arquivos *output* de nossas análises para verificarmos, por meio do banco de dados GeneCards, como eles estão envolvidos no processo de deposição de gordura intramuscular. O gene GBE1, localizado no fígado e músculo, é responsável por catalisar a transferência de unidades de glucosil alfa-1,4 ligadas da extremidade externa de uma cadeia de glicogênio para uma posição alfa-1,6, estando associado ao armazenamento de glicogênio¹²⁸. O gene CAPN7 codifica a proteína calpaína que desempenha um papel importante nas proteases de cisteína dependentes de cálcio (sistema proteolítico). Esse processo bioquímico é um dos fatores responsáveis pelo amaciamento *post mortem* e pela maciez da carne em *Bos taurus*¹²⁹. Os genes MYNN e ACTRT3 que regulam o processo de transcrição e estão associados a eventos de obesidade e diabetes, respectivamente, foram descritos como genes

candidatos putativos afetados pelos maiores efeitos de loci de características quantitativas (QTL) associados a características de crescimento em bovinos de corte¹³⁰.

Não podemos afirmar quais *motifs* correspondem apenas à *enhancers* visto que não validamos as regiões identificadas. No entanto, uma vez que os *enhancers* podem apresentar vários sitios TFBSs e comprimento entre 10 e 1.000 pares de base, supomos que os *enhancers* associados a atividades complexas – tais como interações proteicas mais abrangentes - apresentam maior comprimento e, portanto, maior probabilidade de serem elementos regulatórios distais.

A localização das sequências essenciais à regulação da transcrição pode ser mapeada pela técnica *footprinting*. Quando as sequências *motifs* (elementos *cis-acting*) interagem com fatores de transcrição, elas ficam protegidas da ação enzimática da ATAC-Seq revelando um perfil *'heatmap'*²⁶. Esse princípio molecular explica o porquê de alguns *motifs* encontrados nas análises de *clusters* de genes co-expressos não terem sido identificados nos ATAC-Seq *peaks*. Assim sendo, vários genes e diversos *motifs* e fatores de transcrição desencadeiam a sinalização biológica necessária à demanda do organismo tendo, cada um deles, seu papel principal na manifestação do fenótipo.

5. CONCLUSÃO

Nossos estudos demonstraram que as regiões de ATAC-Seq *peaks* identificadas no tecido muscular de bovinos Nelore correspondem às regiões regulatórias associadas à deposição de gordura intramuscular. Os *motifs* e os fatores de transcrição encontrados por meio de nossas análises reforçam a relação bioquímica existente entre regiões de eucromatina e expressão gênica do fenótipo de interesse.

REFERÊNCIAS

- 1. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 2019. doi:10.1038/s41576-018-0089-8
- 2. García A, González S, Antequera F. Nucleosomal organization and DNA base composition patterns. *Nucleus*. 2017. doi:10.1080/19491034.2017.1337611
- Chen CC, Mellone BG. Chromatin assembly: Journey to the CENter of the chromosome.
 J Cell Biol. 2016;214(1):13-24. doi:10.1083/jcb.201605005
- 4. Srivastava R, Ahn SH. The Epigenetic Pathways to Ribosomal DNA Silencing. *Microbiol Mol Biol Rev.* 2016;80(3):545-563. doi:10.1128/MMBR.00005-16
- 5. Wang J, Jia ST, Jia S. New Insights into the Regulation of Heterochromatin. *Trends Genet*. 2016. doi:10.1016/j.tig.2016.02.005
- Grigoryev SA, Woodcock CL. Chromatin organization the 30 nm fiber. *Exp Cell Res*. 2012;318(12):1448-1455. doi:10.1016/j.yexcr.2012.02.014
- Morales V, Giamarchi C, Chailleux C, et al. Chromatin structure and dynamics: functional implications. *Biochimie*. 2001;83(11-12):1029-1039. https://www.ncbi.nlm.nih.gov/pubmed/11879731.
- Cutter AR, Hayes JJ. A brief review of nucleosome structure. FEBS Lett. 2015. doi:10.1016/j.febslet.2015.05.016
- Bintu L, Ishibashi T, Dangkulwanich M, et al. Nucleosomal elements that control the topography of the barrier to transcription. *Cell*. 2012;151(4):738-749. doi:10.1016/j.cell.2012.10.009
- Yu S, Yang F, Shen WH. Genome maintenance in the context of 4D chromatin condensation. *Cell Mol Life Sci.* 2016;73(16):3137-3150. doi:10.1007/s00018-016-2221-2
- Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*. 2017. doi:10.1038/nrm.2017.26
- 12. Kobayashi W and KH. Structural transition of the nucleosome during chromatin remodeling and transcription. *Elsevier*. 2019;59:107-114.

- Ojolo SP, Cao S, Priyadarshani SVGN, et al. Regulation of plant growth and development: a review from a chromatin remodeling perspective. *Front Plant Sci*. 2018. doi:10.3389/fpls.2018.01232
- 14. Tyagi M, Imam N, Verma K, Patel AK. Chromatin remodelers: We are the drivers!! *Nucleus*. 2016. doi:10.1080/19491034.2016.1211217
- 15. Meas R, Smerdon MJ, Wyrick JJ. The amino-terminal tails of histones H2A and H3 coordinate efficient base excision repair, DNA damage signaling and postreplication repair in Saccharomyces cerevisiae. *Nucleic Acids Res.* 2015. doi:10.1093/nar/gkv372
- 16. Javaid N, Choi S. Acetylation- and methylation-related epigenetic proteins in the context of their targets. *Genes (Basel)*. 2017. doi:10.3390/genes8080196
- Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. *Bioessays*. 2015;37(2):148-154. doi:10.1002/bies.201400138
- Jenuwein T, Allis CD. Translating the histone code. *Science (80-)*. 2001;293(5532):1074-1080. doi:10.1126/science.1063127
- Singh NP, Madabhushi SR, Srivastava S, et al. Epigenetic profile of the euchromatic region of human Y chromosome. *Nucleic Acids Res.* 2011;39(9):3594-3606. doi:10.1093/nar/gkq1342
- Banine F, Bartlett C, Gunawardena R, et al. SWI/SNF chromatin-remodeling factors induce changes in DNA methylation to promote transcriptional activation. *Cancer Res*. 2005. doi:10.1158/0008-5472.CAN-04-3554
- 21. Greer EL, Shi Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012. doi:10.1038/nrg3173
- 22. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013. doi:10.1038/npp.2012.112
- 23. Jang HS, Shin WJ, Lee JE, Do JT. CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes (Basel)*. 2017. doi:10.3390/genes8060148
- 24. Tamaru H. Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev.* 2010;24(14):1465-1478. doi:10.1101/gad.1941010
- Geiman TM, Robertson KD. Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? *J Cell Biochem*. 2002;87(2):117-125. doi:10.1002/jcb.10286

- Pacis A, Mailhot-Léonard F, Tailleux L, et al. Gene activation precedes DNA demethylation in response to infection in human dendritic cells. *Proc Natl Acad Sci U S A*. 2019. doi:10.1073/pnas.1814700116
- Sims RJ, Millhouse S, Chen CF, et al. Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Mol Cell*. 2007. doi:10.1016/j.molcel.2007.11.010
- Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev.* 2014. doi:10.1021/cr400730x
- 29. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol.* 2004. doi:10.1038/nsmb763
- Han Y, He Y. Eukaryotic transcription initiation machinery visualized at molecular level.
 Transcription. 2016. doi:10.1080/21541264.2016.1237150
- 31. Roeder RG. Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett*. 2005;579(4):909-915. doi:10.1016/j.febslet.2004.12.007
- 32. Goodfellow SJ, Zomerdijk JCBM. Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Subcell Biochem*. 2013. doi:10.1007/978-94-007-4525-4_10
- Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 2002. doi:10.1101/gad.1018902
- 34. Schneider DA. RNA polymerase I activity is regulated at multiple steps in the transcription cycle: Recent insights into factors that influence transcription elongation. *Gene*. 2012. doi:10.1016/j.gene.2011.08.006
- 35. Vannini A, Cramer P. Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Mol Cell*. 2012. doi:10.1016/j.molcel.2012.01.023
- Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006;7:29-59. doi:10.1146/annurev.genom.7.080505.115623
- Latchman DS. Transcription factors: An overview. Int J Biochem Cell Biol. 1997. doi:10.1016/S1357-2725(97)00085-X
- Peter I, Davidson E. Genomic Control Process: Development and Evolution. Elsevier;
 2015.

- 39. Noonan JP, McCallion AS. Genomics of Long-Range Regulatory Elements. *Annu Rev Genomics Hum Genet*. 2010. doi:10.1146/annurev-genom-082509-141651
- 40. Kornberg RD. Eukaryotic transcriptional control. *Trends Biochem Sci.* 1999. doi:10.1016/S0968-0004(99)01489-9
- 41. Bhuiyan T, Timmers HTM. Promoter Recognition: Putting TFIID on the Spot. *Trends Cell Biol*. 2019. doi:10.1016/j.tcb.2019.06.004
- 42. Maston GA, Evans SK, Green MR. Transcriptional Regulatory Elements in the Human Genome. *Annu Rev Genomics Hum Genet*. 2006. doi:10.1146/annurev.genom.7.080505.115623
- 43. Zabidi MA, Stark A. Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet*. 2016. doi:10.1016/j.tig.2016.10.003
- Chow CN, Chiang-Hsieh YF, Chien CH, et al. Delineation of condition specific Cis- and Trans-acting elements in plant promoters under various Endo- and exogenous stimuli. *BMC Genomics*. 2018. doi:10.1186/s12864-018-4469-4
- 45. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013;49(5):825-837. doi:10.1016/j.molcel.2013.01.038
- 46. Ptashne M, Gann A. Transcriptional activation by recruitment. *Nature*. 1997. doi:10.1038/386569a0
- 47. Claessens F, Gewirth DT. DNA recognition by nuclear receptors. *Essays Biochem*. 2015. doi:10.1042/bse0400059
- 48. Atchison M. Enhancers: Mechanisms Of Action And Cell Specificity. *Annu Rev Cell Dev Biol*. 1988. doi:10.1146/annurev.cellbio.4.1.127
- 49. Reményi A, Schöler HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol*. 2004. doi:10.1038/nsmb820
- Merika M, Williams AJ, Chen G, Collins T, Thanos D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell*. 1998;1(2):277-287. https://www.ncbi.nlm.nih.gov/pubmed/9659924.
- 51. Panne D. The enhanceosome. *Curr Opin Struct Biol.* 2008;18(2):236-242. doi:10.1016/j.sbi.2007.12.002
- 52. Arnosti DN, Kulkarni MM. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem*. 2005;94(5):890-898. doi:10.1002/jcb.20352
- 53. Harris MB, Mostecki J, Rothman PB. Repression of an interleukin-4-responsive

promoter requires cooperative BCL-6 function. *J Biol Chem*. 2005. doi:10.1074/jbc.M412649200

- 54. Ogbourne S, Antalis TM. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J*. 1998.
- 55. Fourel G, Magdinier F, Gilson É. Insulator dynamics and the setting of chromatin domains. *BioEssays*. 2004. doi:10.1002/bies.20028
- 56. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. Locus control regions. *Blood*. 2002. doi:10.1182/blood-2002-04-1104
- 57. D'haeseleer P. What are DNA sequence *motifs*? *Nat Biotechnol*. 2006. doi:10.1038/nbt0406-423
- 58. Zambelli F, Pesole G, Pavesi G. Motif discovery and transcription factor binding sites before and after the next-generation sequencing era. *Brief Bioinform*. 2013. doi:10.1093/bib/bbs016
- 59. Davie K, Jacobs J, Atkins M, et al. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. *PLoS Genet*. 2015;11(2):e1004994. doi:10.1371/journal.pgen.1004994
- 60. Reznikoff WS. Tn5 as a model for understanding DNA transposition. *Mol Microbiol*. 2003;47(5):1199-1206. http://www.ncbi.nlm.nih.gov/pubmed/12603728.
- 61. O'Donnell KA, Burns KH. Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mob DNA*. 2010;1(1):21. doi:10.1186/1759-8753-1-21
- 62. Muñoz-López M, García-Pérez JL. DNA transposons: nature and applications in genomics. *Curr Genomics*. 2010;11(2):115-128. doi:10.2174/138920210790886871
- 63. Naumann TA, Reznikoff WS. Tn5 transposase with an altered specificity for transposon ends. J Bacteriol. 2002;184(1):233-240. http://www.ncbi.nlm.nih.gov/pubmed/11741865.
- 64. Dos Santos Silva DB, Fonseca LFS, Pinheiro DG, et al. Prediction of hub genes associated with intramuscular fat content in Nelore cattle. *BMC Genomics*. 2019. doi:10.1186/s12864-019-5904-x
- Cesar AS, Regitano LC, Koltes JE, et al. Putative regulatory factors associated with intramuscular fat content. *PLoS One.* 2015;10(6):e0128350. doi:10.1371/journal.pone.0128350

- 66. Oliveira GB, Regitano LCA, Cesar ASM, et al. Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics*. 2018;19(1):126. doi:10.1186/s12864-018-4514-3
- Hocquette JF, Gondret F, Baéza E, Médale F, Jurie C, Pethick DW. Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers. *Animal.* 2010;4(2):303-319. doi:10.1017/S1751731109991091
- 68. Wood JD, Enser M, Fisher A V, et al. Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* 2008;78(4):343-358. doi:10.1016/j.meatsci.2007.07.019
- 69. Luchiari Filho A. Pecuária da carne bovina. *São Paulo LinBife*. 2000. doi:10.1037/e574772011-001
- 70. Laborde FL, Mandell IB, Tosh JJ, Wilton JW, Buchanan-Smith JG. Breed effects on growth performance, carcass characteristics, fatty acid composition, and palatability attributes in finishing steers. *J Anim Sci*. 2001. doi:10.2527/2001.792355x
- 71. Hocquette JF, Cassar-Malek I, Jurie C, Bauchart D, Picard B, Renand G. Relationships between muscle growth potential, intramuscular fat content and different indicators of muscle fibre types in young Charolais bulls. *Anim Sci J*. 2012. doi:10.1111/j.1740-0929.2012.01021.x
- 72. Hocquette JF, Cassar-Malek I, Jurie C, Bauchart D, Picard B, Renand G. Relationships between muscle growth potential, intramuscular fat content and different indicators of muscle fibre types in young Charolais bulls. *Anim Sci J*. 2012;83(11):750-758. doi:10.1111/j.1740-0929.2012.01021.x
- 73. Martínez-Álvaro M, Agha S, Blasco A, Hernández P. Muscle lipid metabolism in two rabbit lines divergently selected for intramuscular fat. *J Anim Sci*. 2017;95(6):2576-2584. doi:10.2527/jas.2017.1371
- 74. Oliveira GB, Regitano LCA, Cesar ASM, et al. Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics*. 2018. doi:10.1186/s12864-018-4514-3
- 75. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823-837. doi:10.1016/j.cell.2007.05.009
- 76. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693-705.doi:10.1016/j.cell.2007.02.005

- 77. Petesch SJ, Lis JT. Overcoming the nucleosome barrier during transcript elongation. *Trends Genet.* 2012;28(6):285-294. doi:10.1016/j.tig.2012.02.005
- Studitsky VM, Nizovtseva E V. Nucleosomal Barrier to Transcription: Structural Determinants and Changes in Chromatin Structure. *Biochem Mol Biol J.* 2016. doi:10.21767/2471-8084.100017
- 79. Nizovtseva E V., Clauvelin N, Todolli S, et al. Nucleosome-free DNA regions differentially affect distant communication in chromatin. *Nucleic Acids Res.* 2017. doi:10.1093/nar/gkw1240
- 80. Riechmann JL. Transcriptional Regulation: a Genomic Overview. *Arab B.* 2002. doi:10.1199/tab.0085
- Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the vertebrate insulator protein CTCFbinding sites in the human genome. *Cell*. 2007;128(6):1231-1245. doi:10.1016/j.cell.2006.12.048
- 82. Fatma A. Hashim MSM and WA-A. Review of Different Sequence Motif Finding Algorithms. *Avicenna J Med Biotechnol*. 2019;11.
- Morosini NS. Identificação e caracterização de regiões de eucromatina associadas à regulação da expressão gênica e à gordura intramuscular em bovinos da raça Nelore.
 2018.
- 84. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213-1218. doi:10.1038/nmeth.2688
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
 2012;9(4):357-359. doi:10.1038/nmeth.1923
- Ebbert MT, Wadsworth ME, Staley LA, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*. 2016;17 Suppl 7:239. doi:10.1186/s12859-016-1097-3
- 87. Feng J, Liu T, Zhang Y. Using MACS to identify peaks from ChIP-Seq data. *Curr Protoc Bioinforma*. 2011;Chapter 2:Unit 2.14. doi:10.1002/0471250953.bi0214s34
- 88. García-Alcalde F, Okonechnikov K, Carbonell J, et al. Qualimap: evaluating nextgeneration sequencing alignment data. *Bioinformatics*. 2012;28(20):2678-2679.

doi:10.1093/bioinformatics/bts503

- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-91. doi:10.1093/nar/gku365
- 90. Maertens A, Tran V, Kleensang A, Hartung T. Weighted Gene Correlation Network Analysis (WGCNA) Reveals Novel Transcription Factors Associated With Bisphenol A Dose-Response. Front Genet. 2018. doi:10.3389/fgene.2018.00508
- 91. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005. doi:10.2202/1544-6115.1128
- 92. Cesar ASM, Regitano LCA, Reecy JM, et al. Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. *BMC Genomics*. 2018. doi:10.1186/s12864-018-4871-y
- 93. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008. doi:10.1186/1471-2105-9-559
- 94. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010. doi:10.1093/bioinformatics/btq033
- 95. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res.* 2015. doi:10.1093/nar/gkv416
- 96. Bailey TL, Boden M, Buske FA, et al. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res*. 2009. doi:10.1093/nar/gkp335
- 97. Tran NTL, Huang CH. A survey of motif finding Web tools for detecting binding site *motifs* in ChIP-Seq data. *Biol Direct*. 2014. doi:10.1186/1745-6150-9-4
- 98. Liu B, Yang J, Li Y, McDermaid A, Ma Q. An algorithmic perspective of de novo cisregulatory motif finding based on ChIP-seq data. *Brief Bioinform*. 2018. doi:10.1093/bib/bbx026
- 99. D'Haeseleer P. What are DNA sequence *motifs*? *Nat Biotechnol*. 2006. doi:10.1038/nbt0406-423
- 100. Whitington T, Frith MC, Johnson J, Bailey TL. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res*. 2011. doi:10.1093/nar/gkr341
- 101. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 2017. doi:10.1093/nar/gkw1099
- 102. Andersson R, Sandelin A, Danko CG. A unified architecture of transcriptional regulatory

elements. Trends Genet. 2015. doi:10.1016/j.tig.2015.05.007

- 103. Ando M, Saito Y, Xu G, et al. Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. *Nat Commun*. 2019. doi:10.1038/s41467-019-09937-w
- 104. Yu CP, Lin JJ, Li WH. Positional distribution of transcription factor binding sites in Arabidopsis thaliana. *Sci Rep.* 2016. doi:10.1038/srep25164
- 105. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018. doi:10.1038/s41580-018-0028-8
- 106. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: From properties to genome-wide predictions. *Nat Rev Genet*. 2014. doi:10.1038/nrg3682
- 107. Li J, Zhou D, Qiu W, et al. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Sci Rep.* 2018. doi:10.1038/s41598-017-18705-z
- 108. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between *motifs*. *Genome Biol*. 2007. doi:10.1186/gb-2007-8-2-r24
- 109. Wolberger C. Multiprotein-DNA complexes in transcriptional regulation. *Annu Rev Biophys Biomol Struct*. 1999. doi:10.1146/annurev.biophys.28.1.29
- 110. Ravasi T, Suzuki H, Cannistraci CV, et al. An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *Cell*. 2010. doi:10.1016/j.cell.2010.01.044
- 111. Chen L, Glovert JNM, Hogan PG, Rao A, Harrison SC. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature*. 1998. doi:10.1038/32100
- 112. Decker EL, Nehmann N, Kampen E, Eibel H, Zipfel PF, Skerka C. Early growth responce proteins (EGR) and nuclear factors of activated T cells (NFAT) form heterodimers and regulate proinflammatory cytokine gene expression. *Nucleic Acids Res.* 2003. doi:10.1093/nar/gkg186
- 113. Huang W, Guo Y, Du W, Zhang X, Li A, Miao X. Global transcriptome analysis identifies differentially expressed genes related to lipid metabolism in Wagyu and Holstein cattle. *Sci Rep.* 2017. doi:10.1038/s41598-017-05702-5
- 114. Trajkovski M, Lodish H. MicroRNA networks regulate development of brown adipocytes. *Trends Endocrinol Metab.* 2013. doi:10.1016/j.tem.2013.05.002
- 115. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: From gene data mining to

disease genome sequence analyses. Curr Protoc Bioinforma. 2016. doi:10.1002/cpbi.5

- 116. Boddicker RL, Koltes JE, Fritz-Waters ER, et al. Genome-wide methylation profile following prenatal and postnatal dietary omega-3 fatty acid supplementation in pigs. *Anim Genet*. 2016. doi:10.1111/age.12468
- 117. Muñoz M, García-Casco JM, Caraballo C, et al. Identification of Candidate Genes and Regulatory Factors Underlying Intramuscular Fat Content Through Longissimus Dorsi Transcriptome Analyses in Heavy Iberian Pigs. *Front Genet*. 2018. doi:10.3389/fgene.2018.00608
- 118. Ravnskjaer K, Madiraju A, Montminy M. Role of the cAMP pathway in glucose and lipid metabolism. *Handb Exp Pharmacol*. 2015. doi:10.1007/164_2015_32
- 119. Birsoy K, Chen Z, Friedman J. Transcriptional Regulation of Adipogenesis by KLF4. *Cell Metab.* 2008. doi:10.1016/j.cmet.2008.02.001
- 120. Zhang H, Wang WP, Guo T, et al. The LIM-Homeodomain Protein ISL1 Activates Insulin Gene Promoter Directly through Synergy with BETA2. *J Mol Biol.* 2009. doi:10.1016/j.jmb.2009.07.036
- 121. Wang W, Shi Q, Guo T, et al. PDX1 and ISL1 differentially coordinate with epigenetic modifications to regulate insulin gene expression in varied glucose concentrations. *Mol Cell Endocrinol*. 2016. doi:10.1016/j.mce.2016.03.019
- Phillips DIW, Caddy S, Ilic V, et al. Intramuscular triglyceride and muscle insulin sensitivity: Evidence for a relationship in nondiabetic subjects. *Metabolism*. 1996. doi:10.1016/S0026-0495(96)90260-7
- 123. Hermsdorff HHM, Monteiro JBR. Gordura visceral, subcutânea ou intramuscular: onde está o problema? *Arq Bras Endocrinol Metabol*. 2004. doi:10.1590/s0004-27302004000600005
- 124. Cui J, Chen W, Liu J, Xu T, Zeng Y. Study on quantitative expression of PPARγ and ADRP in muscle and its association with intramuscular fat deposition of pig. *Springerplus*. 2016. doi:10.1186/s40064-016-3187-0
- 125. Cui JX, Zeng QF, Chen W, Zhang H, Zeng YQ. Analysis and preliminary validation of the molecular mechanism of fat deposition in fatty and lean pigs by high-throughput sequencing. *Mamm Genome*. 2019. doi:10.1007/s00335-019-09795-3
- 126. Sun YM, Qin J, Liu SG, et al. PDGFRα regulated by miR-34a and FoxO1 Promotes adipogenesis in porcine intramuscular preadipocytes through erk signaling pathway. *Int*

J Mol Sci. 2017. doi:10.3390/ijms18112424

- 127. Mei Y, Wang Z, Zhang L, et al. Regulation of neuroblastoma differentiation by forkhead transcription factors FOXO1/3/4 through the receptor tyrosine kinase PDGFRA. *Proc Natl Acad Sci U S A*. 2012. doi:10.1073/pnas.1119535109
- 128. Iijima H, Iwano R, Tanaka Y, et al. Analysis of GBE1 mutations via protein expression studies in glycogen storage disease type IV: A report on a non-progressive form with a literature review. *Mol Genet Metab Reports*. 2018. doi:10.1016/j.ymgmr.2018.09.001
- 129. Curi RA, Chardulo LAL, Mason MC, Arrigoni MDB, Silveira AC, De Oliveira HN. Effect of single nucleotide polymorphisms of CAPN1 and CAST genes on meat traits in Nellore beef cattle (Bos indicus) and in their crosses with *Bos taurus. Anim Genet.* 2009. doi:10.1111/j.1365-2052.2009.01859.x
- Seabury CM, Oldeschulte DL, Saatchi M, et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*. 2017. doi:10.1186/s12864-017-3754-y