



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS AGRÁRIAS
PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

SHIRLIANE DE ARAÚJO SOUSA

IDENTIFICAÇÃO DE GENES CANDIDATOS E PROCESSOS BIOLÓGICOS
RELACIONADOS A CARACTERÍSTICAS DE CARÇA EM *OVIS ARIES*
POR MEIO DA GENÔMICA COMPARATIVA E ENRIQUECIMENTO
GÊNICO

TERESINA-PI

2020

SHIRLIANE DE ARAÚJO SOUSA

**IDENTIFICAÇÃO DE GENES CANDIDATOS E PROCESSOS BIOLÓGICOS
RELACIONADOS A CARACTERÍSTICAS DE CARCAÇA EM *OVIS ARIES*
POR MEIO DA GENÔMICA COMPARATIVA E ENRIQUECIMENTO
GÊNICO**

Tese apresentada ao Programa de Pós-Graduação em Ciência Animal, da Universidade Federal do Piauí, como requisito para obtenção do título de Doutora em Ciência Animal.
Área de concentração: Produção Animal.

Orientador: Prof. Dr. José Lindenberg Rocha Sarmiento

Coorientador: Prof. Dr. Romuere Rodrigues Veloso e Silva

TERESINA-PI

2020

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Setorial do Centro de Ciências Agrárias
Serviço de Processos Técnicos

S725i

Sousa, Shirliane de Araújo.

Identificação de genes candidatos e processos biológicos relacionados a características de carcaça em *Ovis aries* por meio da genômica comparativa e enriquecimento gênico / Shirliane de Araújo Sousa. -- 2020.

128 f. : il.

Tese (Doutorado) – Universidade Federal do Piauí, Centro de Ciências Agrárias, Programa de Pós-Graduação em Ciência Animal, Teresina, 2020.

“Orientador: Prof. Dr. José Lindenberg Rocha Sarmento.”

“Co-orientador: Prof. Dr. Romuere Rodrigues Veloso e Silva.”

1. Ovinocultura de corte. 2. Longissimus dorsi. 3. GWAS. 4. BLAST.
I. Sarmento, José Lindenberg Rocha. II. Silva, Romuere Rodrigues Veloso.
III. Título.

CDD 636.3

Elaborada por Thais Vieira de Sousa Trindade - CRB-3/1282

IDENTIFICAÇÃO DE GENES CANDIDATOS E PROCESSOS BIOLÓGICOS
RELACIONADOS A CARACTERÍSTICAS DE CARÇA EM *OVIS ARIES*
POR MEIO DA GENÔMICA COMPARATIVA E ENRIQUECIMENTO
GÊNICO

SHIRLIANE DE ARAUJO SOUSA

Tese aprovada em: 20/12/2019

Banca Examinadora:



Prof. Dr. José Lindenberg Rocha Sarmiento (Presidente) / DZO/CCA/UFPI



Prof. Dr. Fábio Barros Britto (Interno) / CCN/UFPI



Prof. Dr. Romuere Rodrigues Veloso e Silva (Interno) / CSHNB/UFPI



Prof. Dr. Natanael Pereira da Silva Santos (Interno) / CPCE/UFPI



Prof. Dr. Luiz Antonio Silva Figueiredo Filho (Externo) / IFMA



Profa. Dra. Ester Miranda Pereira (Externa) / CET

Dedico esta árdua caminhada aos meus pais, Ariosvaldo Pereira de Sousa e Eliana de Jesus Araújo; aos meus irmãos, Ariosvaldo Filho e Amanda Beatriz; ao meu amado esposo, Igor Soares; e aos meus filhos (que ainda virão), que compartilharam comigo toda essa trajetória.

AGRADECIMENTOS

Gratidão é o sentimento mais nobre que um ser humano pode manifestar, com ela dissipa-se o amor, o carinho, a compaixão e a humildade. E é com esse sentimento de gratidão que eu agradeço primeiro à Deus, por me manter firme e não me deixar fraquejar nos momentos mais difíceis, que não foram poucos. Obrigada Senhor, por preservar minha sanidade mental durante essa caminhada e por ter feito com que eu conhecesse pessoas incríveis que me acompanharam durante toda essa luta.

Agradeço ao Programa de Pós-Graduação em Ciência Animal (PPGCA/UFPI), pela oportunidade de realizar este curso de doutorado. Nesse contexto, agradeço especialmente ao meu orientador, professor Dr. José Lindenberg Rocha Sarmiento, pela oportunidade de realizar esse trabalho, pela paciência, e principalmente, por acreditar que eu pudesse concluí-lo com sucesso.

Agradeço pelo apoio do professor Dr. Romuere Rodrigues Veloso e Silva, por suas valiosas contribuições na co-orientação deste trabalho, pela paciência, tranquilidade, disponibilidade, amizade e pelo incentivo durante essa caminhada.

Agradeço ao Instituto Nacional de Ciência e Tecnologia em Ciência Animal (INCT-CA) do CNPq, pelo suporte nos dados utilizados nessa pesquisa.

Agradeço a todos os amigos do Grupo de Estudos em Genética e Melhoramento Animal (GEMA – UFPI), representados nas pessoas dos amigos Bruna, LÍlian, Luciano, Débora, Tatiana, Arthur, Expedito e Max. Meus amigos, sem a contribuição de cada um de vocês nada disso seria possível.

Agradeço a todos os servidores administrativos, do PPGCA, em especial a pessoa do Sr. Luís, que sempre foi muito querido e eficiente em tudo que lhe foi solicitado. Aos servidores da Biblioteca Setorial do CCN, que me acompanharam com muita força durante a escrita do meu documento de tese, sempre torcendo positivamente pelo meu sucesso.

Agradeço aos professores Dr. Fabio Barros Britto, Dr. Natanael Pereira da Silva Santos, Dr. Ester Miranda Pereira e Dr. Luiz Antônio Silva Figueiredo Filho, por disponibilizarem sua presença na banca examinadora da defesa de tese, assim como por suas valiosas colaborações.

Agradeço ao apoio de toda minha família e amigos, que sempre de alguma forma se fizeram presente nessa caminhada.

Agradeço aos amigos de trabalho da UECE, representados nas pessoas de Jones Baroni e Débora Ximenes, pelo incentivo diário e motivação para prosseguir.

Agradeço a família que ganhei e que me recebeu de braços abertos neste ano de 2018/19, “os vizinhos do 701”, representados nas pessoas de Danyelle, Ricardo, Luma, Gabriel, Wedson, e o nosso querido Pedro!

Agradeço a Família Soares (Maria das Neves, Ellen, Aldenora, Sr. Antônio Soares) pelo suporte, apoio, carinho e muito amor durante essa caminhada.

Agradeço a minha família, meus pais (Ariosvaldo e Eliana), meus irmãos (Amanda e Ariosvaldo Filho), meu cunhado (Herderson) e meu esposo (Igor Soares), pela motivação e força para concluir esse trabalho. Tudo foi por vocês!!!

Agradeço ao meu amigo Luciano Silva Sena, que fez mais do que um irmão poderia ter feito. Obrigada meu amigo! Eu não tenho palavras e nem sentimento o suficiente para demonstrar minha gratidão por você. Somos sonhadores e vamos conseguir realizar nossos sonhos, porque somos fortes!

Enfim, agradeço a oportunidade de viver para realizar sonhos como esse!

Obrigada a todos!

SUMÁRIO

LISTA DE ILUSTRAÇÕES.....	09
LISTA DE TABELAS.....	10
RESUMO.....	13
ABSTRACT.....	14
1 INTRODUÇÃO	16
2 REVISÃO DE LITERATURA	19
2.1 PANORAMA GERAL DA PRODUÇÃO E CONSUMO DE CARNE OVINA.....	19
2.2 PRINCIPAIS CARACTERÍSTICAS INDICADORAS DE QUALIDADE E RENDIMENTO DE CARCAÇA EM OVINOS.....	20
2.3 O USO DE INFORMAÇÕES GENÉTICAS E GENÔMICAS PARA O AUMENTO DA PRODUÇÃO E QUALIDADE DE CARNE.....	22
2.4 PRINCIPAIS APLICAÇÕES DA GENÔMICA NA PRODUÇÃO ANIMAL....	24
2.4.1 <i>Genômica animal: um breve histórico</i>	25
2.5 USO DA BIOINFORMÁTICA EM ESTUDOS GENÔMICOS	27
2.6 ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA NA PECUÁRIA	31
2.6.1 <i>Aplicação de GWAS quando há poucos animais genotipados</i>	32
2.6.2 <i>Utilização de janelas de SNPs em GWAS</i>	33
2.7 O USO DA GENÔMICA COMPARATIVA NA PECUÁRIA	34
2.7.1 <i>Alinhamento e comparação de sequências biológicas.....</i>	36
2.8 ANÁLISE DE ANOTAÇÃO FUNCIONAL.....	39
2.8.1 <i>Termos de ontologia gênica (GO).....</i>	39
2.8.2 <i>Enciclopedia de Kyoto de Genes e Genomas (KEGG).....</i>	40
2.8.3 <i>Principais metodologias estatísticas utilizadas.....</i>	40
2.8.4 <i>Utilização da ferramenta DAVID para anotação funcional.....</i>	41
REFERÊNCIAS.....	42
CAPÍTULO I - Aplicação de genômica comparativa para a identificação de genes relacionados a características de carcaça na espécie <i>Ovis aries</i>	48

1 Introdução	52
2 Material e Métodos	53
3 Resultados e Discussão	60
4 Conclusões	72
Referências	74
CAPÍTULO II - Análise de enriquecimento gênico para características de carcaça em ovinos.....	80
1 Introdução	84
2 Material e Métodos	85
3 Resultados e Discussão	89
4 Conclusões.....	106
Referências	108
CONSIDERAÇÕES FINAIS.....	111
APÊNDICES	113

LISTA DE ILUSTRAÇÕES

REVISÃO DE LITERATURA

	Pág.
Figura 1. Coleta das medidas de área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS) com uso de ultrassonografia em ovinos da raça Santa Inês.	21
Figura 2. Esquema de organização estruturação e gerenciamento dos principais bancos, bases e sub-bases de dados biológicos.	29
Figura 3. Representação esquemática das duas formas de alinhamento: global e local.	37

CAPITULO I: Aplicação de genômica comparativa para a identificação de genes relacionados a características de carcaça na espécie *Ovis aries*

Figura 1. Esquema que representa o processo utilizado para a delimitação das regiões (janelas) utilizadas para a busca de genes candidatos no genoma da espécie <i>Ovis aries</i> .	54
Figura 2. Esquema que descreve o método de busca de genes utilizado neste trabalho, desde a análise dos alinhamentos significativos até a genômica comparativa para a identificação de genes candidatos às características de carcaça: 1- Etapa de seleção dos segmentos de DNA a serem alinhados. 2- Alinhamentos e filtragens gerados pelo BLASTn. 3- Análise dos genes oriundos dos melhores alinhamentos, em busca de candidatos no genoma de <i>Ovis aries</i> via genômica comparativa.	59

CAPITULO II: Análise de enriquecimento gênico para características de carcaça em ovinos

Figura 1. Delimitação de cada janela da espécie <i>Ovis aries</i> utilizada.	86
Figura 2. Via de sinalização de PI3K-Akt (oas04151) enriquecida com genes presentes nas regiões delimitadas para área de olho de lombo em ovinos da raça Santa Inês.	94

Figura 3. Via de sinalização de ErbB (oas04012) enriquecida com genes presentes nas regiões delimitadas para área de olho de lombo em ovinos da raça Santa Inês.	95
Figura 4. Via de sinalização de insulina (oas04910).	101

LISTA DE TABELAS

REVISÃO DE LITERATURA	Pág.
Tabela 1. Principais bancos de dados biológicos utilizados no presente estudo.	30
Tabela 2. Descrição dos programas BLASTs disponíveis para cada tipo de comparação de sequências.	38
CAPITULO I: Aplicação de genômica comparativa para a identificação de genes relacionados a características de carcaça na espécie <i>Ovis aries</i>	
Tabela 1. Processo utilizado para a delimitação das regiões (janelas) utilizadas para a busca de genes candidatos no genoma da espécie <i>Ovis aries</i> neste estudo.	55
Tabela 2. Quantidade de alinhamentos (AOL) por janelas antes e após os filtros de <i>E-value</i> , Identidade e RAG.	66
Tabela 3. Quantidade de alinhamentos (EGS) por janelas antes e após os filtros de <i>E-value</i> , Identidade e RAG.	67
Tabela 4. Regiões de alinhamentos entre <i>Ovis aries</i> e as espécies referência para a característica área de olho de lombo.	69
Tabela 5. Regiões de alinhamentos entre <i>Ovis aries</i> e as espécies referência à característica espessura de gordura subcutânea.	72
CAPITULO II: Análise de enriquecimento gênico para características de carcaça em ovinos	

Tabela 1. Regiões do genoma da espécie <i>Ovis aries</i> utilizadas para a busca por genes.	87
Tabela 2. Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de genes presentes nas regiões selecionadas no genoma ovino.	90
Tabela 3. Vias KEGG significativamente enriquecidas ($P < 0,05$) para as características em estudo a partir de genes presentes nas regiões selecionadas no genoma ovino.	92
Tabela 4. Análise de anotação de termos de ontologia gênica (GO) enriquecida ($P < 0,05$) a partir de genes presentes nas regiões selecionadas no genoma ovino.	96
Tabela 5. Vias KEGG significativamente enriquecidas ($P < 0,05$) a partir do conjunto de genes presentes nas regiões selecionadas no genoma ovino.	99
Tabela 6. Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de potenciais genes candidatos presentes no genoma ovino, por característica	102
Tabela 7. Vias KEGG significativamente enriquecidas ($P < 0,05$) para as características em estudo a partir da plataforma DAVID, a partir de potenciais genes candidatos presentes no genoma ovino, por característica	104
Tabela 8. Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de potenciais genes candidatos presentes no genoma ovino	105
Tabela 9. Vias KEGG significativamente enriquecidas ($P < 0,05$) a partir de potenciais genes candidatos para as características em estudo.	107

APÊNDICES

- Tabela Suplementar 1.** Regiões genômicas relacionadas a área de olho de lombo (AOL) em ovinos Santa Inês* e regiões extrapoladas para análise no presente estudo. 114
- Tabela Suplementar 2.** Regiões genômicas relacionadas a espessura de gordura subcutânea (EGS) em ovinos Santa Inês* e regiões extrapoladas para análise no presente estudo. 116
- Tabela Suplementar 3.** Regiões genômicas relacionadas à área de olho de lombo mensurada por ultrassonografia em tempo real em ovinos Santa Inês. 118
- Tabela Suplementar 4.** Regiões genômicas relacionadas à espessura de gordura subcutânea mensurada por ultrassonografia em tempo real em ovinos Santa Inês. 120
- Tabela Suplementar 5.** Genes candidatos imediatos presentes nas regiões relacionadas a AOL na espécie *Ovis aries* e funções biológicas mais relevantes. 121
- Tabela Suplementar 6.** Genes candidatos imediatos presentes nas regiões relacionadas a EGS na espécie *Ovis aries* e funções biológicas mais relevantes. 124

RESUMO

SOUSA, Shirliane de Araújo. **Identificação de genes candidatos e processos biológicos relacionados a características de carcaça em *Ovis aries* por meio da genômica comparativa e enriquecimento gênico.** 2020. Tese (Doutorado em Ciência Animal) – Universidade Federal do Piauí, Teresina, 2020.

Informações sobre a arquitetura genética de características de carcaça em ovinos ainda são escassas na literatura. Para a identificação de marcadores moleculares e genes associados com características de carcaça, as análises de associação genômica ampla (GWAS) representam uma das principais metodologias. A partir de resultados de GWAS, é possível a realização de análises de genômica comparativa entre diferentes espécies e a extrapolação de informações entre estas. Isto possibilita o melhor entendimento da arquitetura genética de uma espécie de interesse e a identificação de funções biológicas controladas por conjuntos de genes associados à expressão de diferentes características de interesse econômico. Portanto, nesta pesquisa objetivou-se identificar potenciais genes (ou regiões genômicas) candidatos relacionados a desenvolvimento muscular e deposição de gordura, em regiões específicas do genoma da espécie *Ovis aries*, associadas a características de carcaça nesta espécie, a partir da comparação com genes já descritos no genoma de espécies referência, além de identificar termos de ontologia e vias metabólicas que incluem os genes da espécie ovina. Para identificação dos genes candidatos, foram delimitadas janelas no genoma da espécie *Ovis aries* relacionadas às características área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS). Posteriormente, foram realizadas análises comparativas entre as janelas e genes das espécies espécie *Ovis aries* com regiões dos genomas das espécies *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Gallus gallus* e *Homo sapiens*. Para visualizar os genomas, verificar função de genes, comparar DNA e visualizar resultados, foram utilizados o navegador Ensembl, a base de dados UniProt e ferramenta BLASTn, respectivamente. As análises de anotação funcional foram realizadas considerando o p-valor < 0,05 para termos de ontologia gênica (GO) e vias KEGG, com uso da ferramenta DAVID. Foram identificados no genoma ovino, 26 genes candidatos imediatos relacionados a AOL e 12 para EGS. A metodologia de investigação através da genômica comparativa possibilitou a identificação dos seguintes genes candidatos e regiões candidatas, respectivamente: ENSOARG00000005535,

LOC101104530, ENSOARG00000005104, para AOL; CR2:127924874-127925647 e CR2:127924874-127925647 para EGS. Importantes termos de ontologia e vias KEGG relacionados a metabolismo proteico e metabolismo lipídico foram identificados com utilização dos conjuntos gênicos analisados. Essas informações poderão dar suporte a estudos genômicos em ovinos, assim como, à identificação e seleção de animais com maior rendimento e melhor qualidade de carcaça.

Palavras-chave: Ovinocultura de corte; *Longissimus dorsi*; GWAS; BLAST.

ABSTRACT

SOUSA, Shirliane de Araújo. **Identification of candidate genes and enrichment analysis for carcass traits in the *Ovis aries* species.** 2020. Thesis (Doctor of Philosophy in Animal Science) – Universidade Federal do Piauí, Teresina, 2020.

There is still a paucity of information on the genetic architecture of carcass traits in sheep. For the identification of molecular markers and genes associated with carcass traits, genome wide association (GWA) analyses are among the main methodologies. From GWA results, it is possible to perform comparative genomic analyses between different species in order to extrapolate information between them. This enables a better understanding of the genetic architecture of a species of interest and the identification of biological functions controlled by gene sets associated with the expression of different traits of economic interest. Therefore, this research aimed to identify candidate genes for carcass traits in sheep by comparing specific regions of *Ovis aries* DNA with regions of genomes of *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Gallus gallus* species and *Homo sapiens*, as well as identifying ontology terms and metabolic pathways that include the genes of the sheep species. Information from 45,465 SNPs and 388 samples was used for genome-wide association analyses, which enabled the identification of windows from 10 adjacent SNPs that explained at least 1.00% of the additive genetic variance for loin eye area (LEA) and subcutaneous fat thickness (SFT). From the extrapolation of these windows, a comparative genomic study was performed to identify genes or regions candidate to the characteristics under study. Nucleotide sequences were subjected to the BLASTn tool. After filtering based on the e-value, identity and RAG parameters, there were 2,164 significant alignments, which had their

biological functions analyzed and selected based on the characteristics associated with muscularity and fat deposition. In total, after comparison with the sheep genome, seven genes and two genomic regions were found with significant information, which may be possible candidates for the characteristics analyzed. The use of comparative genomics favored predictions about which genes and / or regions are candidates for carcass traits, which may help in understanding the genetic architecture and contribute to the selection of Santa Inês sheep, favoring the supply of higher quality meat.

Keywords: meat sheep; *Longissimus dorsi*; GWAS; BLAST.

1 INTRODUÇÃO

A demanda por proteína de origem animal para a alimentação humana tenderá a crescer com o aumento da população previsto para os próximos anos (OECD/FAO, 2018). Nesse contexto, estratégias que possibilitem o aumento da eficiência produtiva dos animais são decisivas e necessárias para reduzir os impactos ambientais causados pela pecuária, garantir a segurança alimentar à população, e para a manutenção da viabilidade financeira das atividades de produção animal. Um investimento promissor consiste em melhorar o potencial genético das principais espécies de interesse zootécnico, que representam a maior fonte de proteína alimentar (LUDU; PLASTOW, 2013).

As três espécies quantitativamente mais importantes para o fornecimento de proteína (carne) para a alimentação humana são aves, suínos e bovinos, nesta ordem. Estas espécies já receberam investimentos maciços em suas cadeias produtivas, inclusive em melhoramento genético. A quarta categoria animal de maior importância comercial é representada pelos ovinos (FAO, 2019), que diferente daquelas citadas anteriormente, recebeu relativamente poucos investimentos no setor produtivo e ainda menos em relação à melhoria genética (MRODE et al., 2018).

Na maior parte dos países, as avaliações genéticas em ovinos ainda se baseiam, prioritariamente, apenas em informações de fenótipo e pedigree, que são escassas na maioria dos rebanhos (MRODE et al., 2018), principalmente com relação a características que indiquem com maior precisão o rendimento de carcaça, com base na deposição de músculo e gordura, características estas de difícil mensuração. Esses fatores são limitantes do progresso genético para estas características (NAVAJAS, 2014).

Uma alternativa viável para acelerar o progresso genético de ovinos para características de carcaça consiste na utilização de informações oriundas de marcadores moleculares, juntamente com fenótipos de interesse (por exemplo, área de olho de lombo e espessura de gordura subcutânea) (RUPP et al., 2016).

A utilização de informação de genótipos de animais com base em marcadores densos de polimorfismos de nucleotídeo único (SNPs), aliados a informações de fenótipos de interesse podem possibilitar a descoberta da estrutura genética de características complexas (SCHERER; CHRISTENSEN, 2016).

A seleção genômica ampla (GWS) é uma estratégia para acelerar o progresso genético dessas características. Em complemento, a metodologia de estudos de associação genômica ampla (GWAS) é a mais recomendada para a identificação dos efeitos de SNPs associados a características de interesse zootécnico.

Para a realização de análises de GWAS, o método *single-step* GWAS ou ssGWAS (WANG et al., 2012) é um dos mais eficientes. Este método permite a estimação simultânea de todos os efeitos de SNPs e possibilita a utilização de informações de todos os animais de uma determinada população, independentemente de apresentarem informação genotípica e/ou fenotípica (WANG et al., 2012).

Os desafios para a geração de conhecimentos úteis e aplicáveis a partir do emprego de abordagens analíticas neste âmbito, as quais invariavelmente geram enormes quantidades de informação, não se assentam apenas nas etapas de processamento e análise de dados, mas principalmente naquelas que envolvem sua interpretação. Nesse contexto, técnicas de mineração de dados, capazes de extrair informações biológicas consistentes, são necessárias para interpretar o significado biológico de um dado experimento.

Sendo assim, pode-se dizer que um dos maiores desafios da biologia computacional atual é o de prover métodos robustos que permitam inferir a funcionalidade de uma longa lista de genes derivada de estudos de associação genômica ampla (GWAS). Por isso, é necessário o investimento e a inclusão de análises de enriquecimento gênico em dados gerados a partir de estudos de associação genômica, com o objetivo de possibilitar a melhor identificação dos processos biológicos e vias metabólicas associados a características de interesse zootécnico, como características indicadoras de rendimento de carcaça.

Em ovinos, ainda há carência de informações sobre a arquitetura genética de fenótipos de carcaça e os mecanismos fisiológicos em que os genes ligados a estas características participam, e quando se considera características de difícil mensuração, como a deposição de músculo e gordura, os estudos ainda são mais escassos.

Várias espécies de vertebrados, como aves e mamíferos, apresentam mecanismos genéticos e fisiológicos semelhantes relacionados ao desenvolvimento muscular e de tecido adiposo (BOFFELLI; NOBREGA; RUBIN, 2004). Assim, o estudo comparativo da arquitetura genética de características indicadoras de desenvolvimento muscular e deposição de gordura, simultaneamente, em diferentes espécies de vertebrados, pode trazer maiores esclarecimentos sobre a biologia dessas características com relação ao rendimento de carcaça em espécies produtoras de carne (SHAHJAHAN, 2015).

Portanto, a realização de análises de genômica comparativa com o objetivo de avaliar e comparar regiões genômicas envolvidas no metabolismo de músculos esqueléticos e tecido adiposo, em diferentes espécies de mamíferos e aves, poderá auxiliar na identificação de potenciais genes candidatos associados com rendimento e qualidade de carcaça em ovinos, por meio da extrapolação de informações entre as espécies.

Neste trabalho, objetivou-se identificar potenciais genes (ou regiões genômicas) candidatos relacionados ao desenvolvimento muscular e deposição de gordura, em regiões específicas do genoma da espécie *Ovis aries*, com base em regiões associadas a características de carcaça nesta espécie, e a partir da comparação com genes já descritos no genoma de espécies referência. Além de identificar termos de ontologia e vias metabólicas que incluem os genes da espécie ovina.

Esta tese está estruturada com base nas normas do Programa de Pós-Graduação em Ciência Animal da UFPI estabelecidas na Resolução 001/03-CCMCA. A estrutura é apresentada da seguinte forma:

- **Introdução e Revisão de Literatura**, em que são discutidos tópicos gerais e relevantes da pesquisa dentro da temática da produção animal; a importância de investimentos em estudos no melhoramento genético para animais de produção; e o uso da genômica em estudos comparativos de sequências biológicas.

- **Capítulo I:** Aplicação de genômica comparativa para a identificação de genes relacionados a características de carcaça na espécie *Ovis aries*. (Este capítulo foi elaborado com base nas normas da revista *Livestock Science* ISSN 1871-1413).

- **Capítulo II:** Análise de enriquecimento gênico para características de carcaça em ovinos. (Este capítulo foi elaborado com base nas normas da revista *Livestock Science* ISSN 1871-1413).

Finalmente, são apresentadas as considerações finais, onde são discutidos os principais resultados encontrados nesta pesquisa, a contribuição do estudo para avanços no melhoramento genético animal, e as sugestões e perspectivas da continuação desta pesquisa em estudos futuros.

2 REVISÃO DE LITERATURA

2.1 PANORAMA GERAL DA PRODUÇÃO E CONSUMO DE CARNE OVINA

A espécie ovina (*Ovis aries*) é uma das principais utilizadas na pecuária para a produção de carne, e isto se deve à alta adaptabilidade desses animais a diversas condições ambientais. Os ovinos estão presentes em praticamente todas as regiões do planeta, inclusive, criados em condições de manejo inviáveis, ou seja, em regiões com condições ambientais (clima e relevo) inadequadas para a exploração pecuária de outras espécies (FAO, 2009; KNAPIK; ROPKA-MOLIK; PIESZKA, 2017).

A ovinocultura apresenta uma maior importância em países subdesenvolvidos e/ou emergentes, em que os ovinos são criados em sua maioria para a subsistência de famílias de produtores rurais com baixo poder aquisitivo (FAO, 2009). Por outro lado, alguns países desenvolvidos como Nova Zelândia, Austrália e União Europeia (UE) também têm se destacado na produção e exportação de carne ovina (FAO, 2019).

Em 2018, a produção mundial de carnes foi estimada em 336,4 milhões de toneladas (t). As carnes de aves, suínos, bovinos e ovinos, nesta ordem, representaram as categorias de carne mais consumidas mundialmente. Neste cenário, a produção de carne ovina foi estimada em 15,2 milhões de toneladas, o que representou aumento de 0,6% em relação ao ano de 2017, principalmente devido a aumentos nos mercados produtivos da China (0,8%) e Austrália (4,2%) (FAO, 2019).

Os sete maiores produtores de carne ovina no mundo são China, União Europeia, Austrália, Índia, Paquistão, Nova Zelândia e Turquia. A China é o maior produtor e consumidor mundial de carne ovina, e tem se destacado também como maior importador deste produto (348 mil t). Os números de importação da China representam

mais que o dobro do total importado pela UE (140 mil t), que apareceu como o segundo maior importador em 2018 (FAO, 2019).

Em relação às exportações, dados da FAO (2019) apontam que, em 2018, a Austrália liderou o ranking mundial (493 mil t), seguida pela Nova Zelândia (409 mil t) e UE (25 mil t). No continente americano, o Uruguai teve destaque no cenário mundial da ovinocultura, representando o sexto maior exportador de carne ovina em 2018 (12 mil t). Embora o Brasil tenha apresentado a maior produção de carne ovina da América Latina, no período de 2015 a 2017, a exportação anual média de carne ovina brasileira foi menor que mil toneladas. Neste mesmo período, o Brasil produziu em média 118 mil t de carne ovina e importou em média 6 mil toneladas (OECD/FAO, 2018).

No Brasil, assim como nos demais países da América Latina e alguns outros países em desenvolvimento, a produção de ovinos ainda apresenta baixo nível de organização da cadeia produtiva, com reflexos nos índices de produtividade, na qualidade dos produtos, na falta de regularidade da oferta e, principalmente, na informalidade da comercialização da produção (OECD/FAO, 2018; LOBO, 2019).

Em países desenvolvidos como Austrália e Nova Zelândia, grande parte do destaque na produção e comercialização de carne ovina está relacionado ao fato de que estes países apresentam cadeias produtivas bem estruturadas e avançaram bastante no melhoramento genético de características de produção em ovinos (ROWE, 2010; AUVRAY et al., 2014; SWAN et al., 2014; MORTIMER et al., 2017).

Esses fatores têm permitido que a produção de carne ovina ocorra de forma mais eficiente, ou seja, em maior quantidade, com utilização de menos recursos naturais, em menor espaço de tempo e com qualidade condizente com as exigências do mercado consumidor.

Apesar dos diversos desafios enfrentados, a carne ovina, destaca-se pela sua qualidade e sabor, entretanto, é necessário e importante o investimento em programas de manejo e melhoramento genético nos rebanhos, tanto por parte dos ovinocultores, como pelas instituições competentes, com o objetivo de alcançar uma melhor seleção destes, principalmente em países emergentes, como o Brasil (LOBO, 2019).

2.2 PRINCIPAIS CARACTERÍSTICAS INDICADORAS DE RENDIMENTO DE CARÇA EM OVINOS

Para avaliação do potencial genético de animais produtores de carne, com base em características de carcaça (porção comestível do animal, após a retirada de vísceras e membros), normalmente é realizado o abate para a condução de testes de progênie. Contudo, técnicas como ultrassonografia em tempo real (UTR) permitem a inclusão da composição de carcaça nos esquemas de seleção com uso de medidas coletadas *in vivo* em animais candidatos à seleção (SUGUISAWA et al., 2009).

A área de olho-de-lombo (AOL) e a espessura de gordura subcutânea (EGS), medidas entre a 12^a e 13^a vértebras costais no músculo *Longissimus dorsi* (em bovinos, caprinos e ovinos, por exemplo), representam duas das principais características de carcaça utilizadas para indicação do potencial genético do animal para musculosidade, composição da carcaça, rendimento de cortes cárneos de alto valor comercial e precocidade de acabamento da carcaça.

A AOL normalmente é mensurada em centímetros quadrados (cm²) e representa literalmente a área da região mostrada pelo aparelho de ultrassom ao ser realizado o posicionamento da probe (transdutor) perpendicularmente ao comprimento do músculo *Longissimus dorsi*. A medida de EGS, que normalmente é coletada em milímetros (mm), representa a espessura de gordura entre a pele e o músculo, na mesma região em que é mensurada a AOL (Figura 1).



Figura 1 – Coleta das medidas de área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS) com uso de ultrassonografia em ovinos da raça Santa Inês. Fonte: Sena (2019).

Na literatura, há vários estudos que mostram a existência de variabilidade genética suficiente para se obter progresso genético por meio da seleção para AOL e EGS, em diferentes espécies, como bovinos e ovinos de corte (SOUZA et al., 2016; SILVA et al.,

2019). Portanto, o melhoramento genético destas características é de grande importância para o aumento quantitativo e qualitativo produção de carne ovina.

2.3 O USO DE INFORMAÇÕES GENÉTICAS E GENÔMICAS PARA O AUMENTO DA PRODUÇÃO E QUALIDADE DE CARNE

O aumento constante no efetivo da população humana tem sido acompanhado pelo aumento na demanda por alimentos, ao mesmo tempo em que a disponibilidade de áreas agricultáveis se mantém constante e tenderá a diminuir com o passar dos anos (SIMIANER, 2016).

Diante deste cenário, um grande desafio para o futuro será a produção de produtos de origem animal e vegetal com a utilização dos recursos naturais da maneira mais eficiente possível e, ao mesmo tempo, atender às exigências do mercado consumidor para manter a viabilidade econômica das atividades agropecuárias (ROTHSCHILD; PLASTOW, 2014).

Neste sentido, a realização de ações direcionadas ao melhoramento genético das espécies pecuárias será fundamental, pois estas ações têm o potencial para promoverem mudanças nos genótipos existentes nos rebanhos, proporcionando ganhos cumulativos e permanentes em diferentes características (NAVAJAS, 2014).

Como a maioria das características de interesse econômico exploradas comercialmente na pecuária está sob controle genético, há a possibilidade de explorar sua variabilidade genética dentro de uma mesma raça ou entre diferentes raças de uma mesma espécie. Por conta disso, a utilização de informações genéticas é fundamental para a elaboração de esquemas de melhoramento genético animal e para o desenvolvimento de uma exploração pecuária visando a maior eficiência produtiva (LÔBO; LÔBO, 2007; NAVAJAS, 2014; BERRY et al., 2017).

Um exemplo do potencial de melhoria da produção de proteína animal com base no uso de informações genéticas pode ser representado pelo aumento de aproximadamente 90% nos ganhos em desempenho atingidos em várias décadas no setor de avicultura de corte, devido à aplicação de estratégias de melhoramento genético (HAVENSTEIN et al., 2003).

Outro exemplo clássico, resultante de ações de melhoramento genético diz respeito à suinocultura, em que a conversão alimentar diminuiu quase 50% entre os anos de 1960 e 2005 (EGGEN, 2012). Em ovinos, a seleção genética resultou em

aumento de 83% em kg de cordeiro produzido por ovelha e 28% (+4,1 kg) de aumento no peso de carcaça, entre 1990 e 2012, na Nova Zelândia (BRITO et al., 2017).

De modo geral, ainda há relativamente poucos resultados sobre melhoramento genético para características de carcaça e qualidade de carne, pois estas têm recebido menor ênfase em programas de melhoramento genético de espécies utilizadas para a produção de carne, em comparação a outros fenótipos (por exemplo, peso e tamanho corporal). Isto se deve, em parte, às dificuldades e custos relativamente altos para a mensuração precisa de fenótipos indicadores de composição de carcaça e qualidade de carne (RUPP et al., 2016).

As limitações para mensuração adequada e os poucos estudos relacionados à obtenção de informação fenotípica sobre carcaça e qualidade de carne, ainda representam as maiores restrições para o melhoramento genético destas características (NAVAJAS, 2014).

A avaliação do mérito genético dos animais foi realizada durante vários anos apenas com base em informações de fenótipo e pedigree. Contudo, a expressão fenotípica das principais características de interesse na pecuária é resultante da ação conjunta de vários genes e efeitos de ambiente (MEUWISSEN; HAYES; GODDARD, 2016; SARMENTO; SENA, 2017).

Apesar da contribuição dos métodos tradicionais de melhoramento genético para o aumento da produção de diferentes espécies, a eficiência destes métodos diminui quando se considera fenótipos de difícil mensuração, com baixa herdabilidade, ou que não podem ser mensurados rapidamente, a baixo custo e de maneira correta em grande número de animais, como é o caso das medidas de carcaça (NAVAJAS, 2014).

As limitações para identificação de genótipos superiores que permitam inferir sobre a qualidade de carcaça, devido às dificuldades de coleta de fenótipo, podem ser amenizadas ou superadas com a utilização de mensurações realizadas por ultrassonografia em animais vivos (por exemplo, área do músculo *Longissimus dorsi* ou espessura de gordura subcutânea), e com o uso de informações oriundas de marcadores moleculares. Estas ferramentas permitem maior acurácia na avaliação do mérito genético de animais para a seleção (SIMIANER, 2016; MEIRA et al., 2019).

Além dos benefícios proporcionados pelo aumento da acurácia, o desenvolvimento de tecnologias que permitam o estudo amplo e análise funcional do genoma de diferentes espécies também possibilita o maior entendimento de mecanismos genéticos e fisiológicos relacionados à expressão de características de interesse

econômico (NAVAJAS, 2014). Estas informações podem ser aplicadas diretamente no melhoramento genético com uso da seleção assistida por marcadores (ROTHSCHILD; PLASTOW, 2014).

2.4 PRINCIPAIS APLICAÇÕES DA GENÔMICA NA PRODUÇÃO ANIMAL

A genômica é o campo da genética que estuda o genoma dos organismos para compreender o conteúdo, organização, função e evolução da informação genética contida em genomas inteiros (PIERCE, 2016; ARMSTRONG et al., 2018). A genômica pode ser estudada e compreendida em três abordagens: genômica estrutural, que se concentra no sequenciamento dos genomas inteiros e na catalogação, ou anotação das sequências dentro de um determinado genoma; genômica evolutiva, que é a comparação dos genomas dentro e entre espécies; e a genômica funcional, que utiliza as sequências genômicas para compreender a função gênica em um organismo. Juntas, essas três abordagens contribuem para o entendimento do papel de cada gene contido em um determinado genoma, visando um conhecimento amplo da estrutura e função dos genes (BOWMAN; SANDERS, 2014).

As informações contidas nas sequências genômicas favorecem a compreensão de como a variação genética influencia uma característica de interesse, pois permitem mapear esta característica em local preciso no genoma (SCHMUTZ; GRIMWOOD, 2004). A análise do genoma de animais de produção (por exemplo, aves, suínos, bovinos e ovinos) proporciona a obtenção de informações sobre a arquitetura genética de características poligênicas (como por exemplo: ganho de peso, balanço energético, qualidade de carne, rendimento de carcaça, fertilidade e resistência a doenças). Estas informações possibilitam o melhor entendimento da base genética da variação fenotípica entre e dentro de espécies e abrem novos rumos para a melhoria e evolução das características de interesse econômico na pecuária (LANDER; WEINBERG, 2000; HALL; FIEBIG; PREUSS, 2002; ANDERSSON; GEORGES, 2004).

Avanços ocorridos nas áreas da biologia molecular, estatística e ciências da computação (com a bioinformática), alavancaram e otimizaram os estudos na área da genômica (HAGEN, 2001; DEROME et al., 2018). Nesse contexto, a utilização da genômica aplicada à produção animal (genômica animal) passou por uma mudança de paradigma desde o advento do sequenciamento genômico (MEUWISSEN; HAYES; GODDARD, 2016).

Nas últimas décadas, diferentes ferramentas genômicas têm sido utilizadas amplamente para o estudo e caracterização de diversidade genética e estrutura populacional de diferentes espécies pecuárias (ROTHSCHILD; PLASTOW, 2014). Além disso, a utilização de informações genômicas tem contribuído para o manejo e conservação de recursos genéticos animais, identificação de assinaturas de seleção artificial, estimação de parentesco genômico e coeficientes de endogamia, identificação e rastreabilidade de produtos para autenticação e garantia de qualidade, e identificação de mutações letais ou outras mutações de interesse específico (BISCARINI et al., 2015).

Em termos econômicos, as principais aplicações da genômica animal compreendem os Estudos de Associação Genômica Ampla (GWAS) e Predição Genômica Ampla (WGP) ou Seleção Genômica (GS) (KADARMIDEEN, 2014). De modo geral, a genômica aplicada à produção animal consiste basicamente em: encontrar e caracterizar variações de sequência; em seguida, estabelecer associações entre variantes causais e fenótipos de interesse econômico; e então utilizar estas associações para auxiliar no refinamento do melhoramento genético de características de interesse em programas de melhoramento genético animal (LUDU; PLASTOW, 2013).

2.4.1 Genômica animal: um breve histórico

As primeiras pesquisas com utilização de genética molecular tiveram início na década de 1970. Desde então, as aplicações da genética molecular possibilitaram melhorias no âmbito dos programas de melhoramento genético animal e vegetal, devido à possibilidade do uso de marcadores de DNA para a identificação de genes ou regiões genômicas que controlavam características de interesse econômico (ANDERSSON, 2001; DEKKERS, 2012).

A primeira aplicação dos métodos de genética molecular consistiu na descoberta da base genética e desenvolvimento de testes genéticos para defeitos em genes individuais. Para características quantitativas de interesse econômico, os avanços proporcionados pela genética molecular possibilitaram a identificação de QTL (loci de características quantitativas) e o desenvolvimento de testes de DNA para auxiliar na escolha de animais candidatos à seleção precocemente, com base na combinação de informações fenotípicas e dados de marcadores genéticos associados a QTLs. Isso

passou a ser denominado de seleção assistida por marcadores (SAM) (DEKKERS, 2012).

Para a execução da SAM, grande número de estudos de genes candidatos e mapeamento de QTL foram conduzidos, o que resultou na descoberta de vários QTLs, associações de marcadores de DNA com fenótipos e mutações causais (DEKKERS, 2004; ROTHSCILD; PLASTOW, 2014). De acordo com Dekkers (2012), a implementação dessas informações em programas de melhoramento genético foi limitada, por conta destes motivos: a maioria dos estudos de QTL foram conduzidos em cruzamentos experimentais, ao invés de serem testados nas populações que seriam utilizadas para o melhoramento genético; a maioria dos efeitos descobertos tendiam a explicar apenas uma quantidade limitada de variação genética para cada característica; muitos QTL e associações não podiam ser replicados; a realização de genotipagem de candidatos à seleção tinha custo muito alto.

Até meados dos anos 1990, a maioria dos QTLs foi detectada com uso de marcadores microssatélites, que não apresentavam boa distribuição nos genomas das principais espécies pecuárias e tinham baixa associação com os genes alvo (SIMIANER, 2016). Portanto, devido à difícil detecção de genes que causavam variação significativa nas características quantitativas, foram iniciados o re-mapeamento e mapeamento fino dos QTLs mapeados inicialmente (GEORGES, 2007).

No início dos anos 2000, vários avanços ocorreram na área da genômica humana, devido ao grande feito realizado em 2001, quando foi completado o primeiro esboço da sequência do genoma humano (VENTER et al., 2001), que deu início à “revolução genômica” e também resultou em mudança de paradigma na área da produção animal. Em 2004 foi realizado o primeiro sequenciamento do genoma de uma espécie de interesse pecuário, no caso o frango, que foi seguido pelos sequenciamentos do genoma bovino, ovino, suíno, caprino, equino e outras espécies (ROTHSCILD; PLASTOW, 2014).

A divulgação das sequências dos genomas das principais espécies pecuárias levou à disponibilização de marcadores genéticos de alto rendimento variando de aproximadamente 10.000 até um milhão de polimorfismos de nucleotídeo único (SNP). Inicialmente, os dados genotípicos de SNPs de alta densidade eram principalmente utilizados para a condução de estudos de associação genômica ampla (GWAS), com a combinação de variantes genéticas (com inclusão ou não de informação de pedigree)

com o fenótipo observado para gerar estimativas para milhares de marcadores relacionados a cada fenótipo considerado (KADARMIDEEN, 2014).

As rápidas melhorias em tecnologias de alto desempenho de genotipagem de SNPs, arranjos de SNPs mais densos, juntamente com a redução de custos para genotipagem e sequenciamento expandiram as possibilidades de utilização de informações genômicas, inclusive para auxiliar na seleção de animais de produção (MEUWISSEN; HAYES; GODDARD, 2016). Desde a introdução da seleção genômica em bovinos leiteiros a partir de um conceito teórico de Meuwissen et al. (2001), a revolução genômica promoveu mudanças na estrutura do melhoramento genético e, assim, a seleção genômica ampla apresenta-se como a principal ferramenta para condução do melhoramento genético animal atualmente (BAGNATO; ROSATI, 2012).

Todo o êxito que tem sido observado na área da genômica se deve a avanços em seus principais componentes, ou seja, biologia molecular, estatística e ciência da computação (bioinformática e matemática) (SUÁREZ-DÍAZ, 2010; NICOLAZZI et al., 2015; SIMIANER, 2016). Entre esses componentes, a bioinformática se destaca por abranger ferramentas conceituais e práticas para a propagação, geração, processamento e entendimento de ideias científicas e informação biológica (FADIEL; ANIDI; EICHENBAUM, 2005; GARCIA et al., 2012).

2.5 USO DA BIOINFORMÁTICA EM ESTUDOS GENÔMICOS

O sequenciamento de genomas completos, dos mais diferentes organismos, produz uma quantidade enorme de dados de sequências. Um dos maiores desafios da genética moderna, com o advento das novas tecnologias de sequenciamento, é buscar ferramentas para catalogar, armazenar, recuperar e gerenciar soluções computacionais para analisar o conjunto de dados gerados e que viabilizem a extração de informações valiosas a respeito das sequências genômicas dos organismos (COSTA, 2015; FERRARI, 2018).

Neste contexto, as ferramentas da bioinformática são necessárias, pois estas fazem parte de um campo emergente composto por biologia molecular e ciência computacional que se concentra no desenvolvimento de bancos de dados, algoritmos de pesquisa em computador, programa de previsão gênica e outras ferramentas analíticas usadas para dar sentido aos dados de sequência de DNA, RNA e proteínas (PIERCE, 2016).

A bioinformática traz uma variedade de tarefas e processos que possuem o objetivo de resolver e solucionar os principais questionamentos relacionados com a genômica animal, como por exemplo, alinhamento de sequências de DNA, procura por genes (por associação e prévio conhecimento de outros genes), comparação de genomas, predição de estruturas de proteínas (secundárias, terciárias e quaternárias), bem como outras questões associadas, como a evolução (FOURMENT; GILLINGS, 2008; FERRARI, 2018).

Bancos de dados biológicos (BDB) foram criados para a coleta e análise de sequências de DNA, proteína e outras informações adicionais. O crescente aumento e acúmulo de informações sobre espécies de interesse econômico depositadas nesses bancos de dados, juntamente com o desenvolvimento de ferramentas de análise adequadas, têm permitido a realização de análises comparativas com dados de organismos modelo, facilitando a descoberta de genes envolvidos em características fenotípicas economicamente importantes (SOARES et al. 2019).

Os bancos de dados podem ser divididos em três categorias a partir do tipo e do conteúdo dos dados neles armazenados: bancos de dados primários, que possuem as informações das sequências originais (de nucleotídeos ou proteínas), junto com informações que descrevem a fonte da sequência e sua determinação; bancos de dados secundários, que armazenam os resultados de análises feitas a partir dos dados primários, como as informações sobre padrões de sequência específicos, variações, mutações e relações evolutivas; e bancos de dados adicionais ou especializados, que atendem a um interesse particular da pesquisa, e que têm informações sobre a diversidade na sequência, como e onde um genoma varia entre os organismos individuais (ASHBURNER; GOODMAN, 1997; COHEN, 2004).

Portanto, é importante conhecer as principais bases e sub-bases de dados biológicos, como são alimentadas e como estão organizadas. O gerenciamento de dados genômicos baseia-se na existência do consórcio *International Nucleotide Sequence Database Collaboration* (INSDC). Este consórcio é constituído pelos bancos de dados **DNA Data Bank of Japan** (DDBJ), **European Nucleotide Archive** (ENA) e **GenBank**, que operam independentemente, mas sincronizam seus dados. A partir dessa fonte primária de dados de sequência, muitos outros bancos de dados secundários e terciários são construídos. Eles fornecem propósitos muito mais específicos e alimentam os dados para os principais bancos de dados. Os genes são vinculados usando identificadores exclusivos (FARRELL, 2019) (Figura 1).

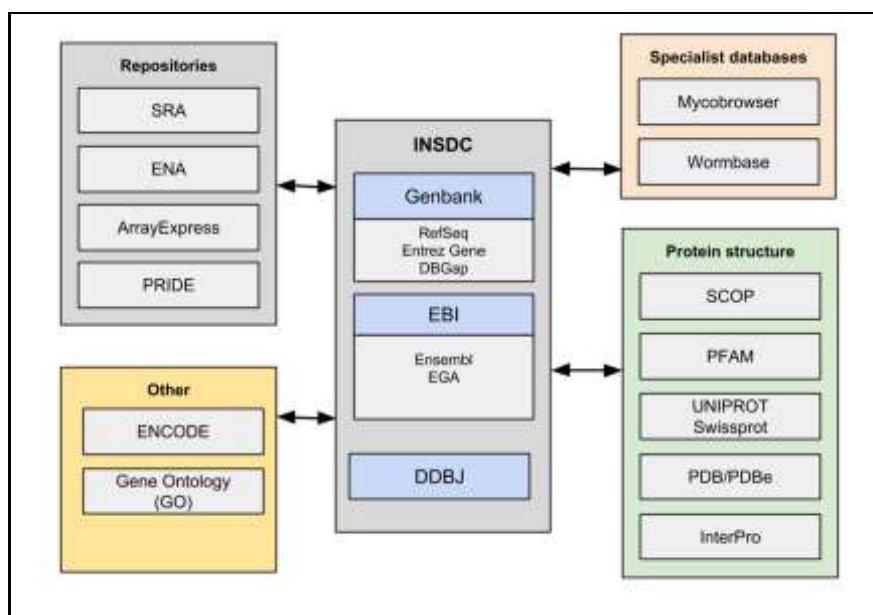


Figura 2 - Esquema de organização, estruturação e gerenciamento dos principais bancos, bases e sub-bases de dados biológicos. Fonte: Farrell (2019).

A maioria dos bancos de dados relacionados a sequências biológicas e suas demais funções possuem acesso gratuito, ou seja, são dados públicos, que através da internet tornam-se facilmente pesquisáveis, proporcionando para cientistas e estudantes o livre acesso a esse importante recurso. Alguns bancos de dados utilizados no presente estudo estão exemplificados na Tabela 1.

Tabela1 - Principais bancos de dados biológicos utilizados no presente estudo.

Nome	Descrição	URL
GenBank	Informação primária da sequência de DNA mantida pelo U.S National Institutes of Health.	http://www.ncbi.nlm.nih.gov/Genbank/
Ensembl	Contém dados de genomas de vertebrados associados com pesquisas de genômica comparativa, evolução, variação de sequências e regulação transcricional.	http://www.ensembl.org
UniProt	Dados de sequências de proteínas e outras informações sobre proteínas de uma variedade de organismos.	http://www.ebi.ac.uk/uniprot/

No campo da genômica, após o sequenciamento de um genoma, é preciso identificar e agregar informação à sequência, por meio da identificação dos genes potenciais da sequência. Após a identificação, cada gene tem que passar por uma notação, ou seja, conectar as informações da sua sequência a outras informações sobre sua função e expressão, a proteína que codifica e informações sobre genes semelhantes em outras espécies (SANTOS; ORTEGA, 2003).

Devido à complexidade de achados sobre os genes em uma determinada sequência, foram desenvolvidos programas com o objetivo de analisar regiões ou sequências específicas no DNA que estejam associadas a alguns genes. O desenvolvimento de todas essas aplicações envolve uma série de etapas experimentais e de análise de dados, exigindo infraestrutura e especialistas em diferentes áreas como biologia molecular, zootecnia, veterinária, estatística e ciência da computação (GIACHETTO; HIGA, 2014).

As análises oriundas da bioinformática usam diversas linguagens de programação distintas para solucionar e lidar com os problemas biológicos, de modo que a escolha de uma linguagem pode ser determinada por três fatores básicos: recursos disponíveis, facilidade de uso e necessidade. De uma forma geral, as linguagens mais utilizadas são Perl, Python e R. Estas três linguagens disponibilizam grandes acervos de ferramentas

para trabalho com dados biológicos já bem estabelecidas, como BioPerl (www.bioperl.org), BioPython (www.biopython.org) e Bioconductor (www.bioconductor.org), todos apoiados pela *Open Bioinformatics Foundation* (www.openbio.org), que trabalha focada em apoiar programação em código aberto na área de bioinformática (FOURMENT, 2008; FERRARI, 2018).

2.6 ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA NA PECUÁRIA

Estudos de associação genômica ampla (GWAS) consistem, basicamente, na utilização de painéis densos de SNPs para a investigação de variações genéticas em uma determinada população que possa desempenhar alguma função em um fenótipo de interesse (GARCIA et al., 2012). A variante genética que causa uma determinada característica pode não ser testada em análises de GWAS, mas sua assinatura poderá ser evidente por meio da associação de SNPs que ocorrem dentro de um mesmo haplótipo (conjunto específico de alelos) ou cromossomo (SCHERER; CHRISTENSEN, 2016). A pressuposição geral em GWAS é de que associações significativas acontecem devido à presença de desequilíbrio de ligação (proximidade) entre um SNP e uma mutação causal que afeta uma determinada característica (HAYES; GODDARD, 2010).

A aplicação de GWAS foi inicialmente realizada na análise de doenças humanas e em seguida foi estendida a outras áreas de estudo, como a genética e melhoramento de animais domésticos, a partir da descoberta e disponibilização de sequências genômicas e grande número de SNPs para várias espécies domésticas (ZHANG et al., 2012). Atualmente, há vários tipos de chips de SNPs comerciais, com diferentes densidades, disponíveis para as principais espécies de interesse zootécnico (por exemplo, bovinos, ovinos, suínos, equinos e aves). Os painéis de SNPs utilizados para animais são construídos com informações oriundas de genotipagens usando plataformas de alto rendimento das empresas Affymetrix e Illumina.

Há vários exemplos de estudos de GWAS bem-sucedidos em diferentes espécies pecuárias, que já permitiram a identificação de milhares de genes associados a características como qualidade e rendimento de carne, facilidade de parto, produção de leite, porcentagem de gordura e proteína no leite, fertilidade e produção de ovos (SHARMA et al., 2015). Devido às diferentes arquiteturas genéticas das diversas raças existentes e à natureza poligênica das características complexas, diferentes genes e

regiões já foram encontrados em associação com uma característica em comum em diferentes raças de uma mesma espécie (NAHA et al., 2016).

Com o advento da GWAS e outras metodologias genômicas, a quantidade de genes descobertos que afetam características complexas (por exemplo, produção de leite em bovinos) aumentou mais que 100 vezes nas últimas duas décadas. Isto tornou essas características ainda mais complexas do que se imaginava antes, pois se descobriu que esses fenótipos são afetados por muitos genes com pequenos efeitos individuais (MEUWISSEN; HAYES; GODDARD, 2016). Por conta deste e outros fatores, o mapeamento de genes associados a características complexas ainda é uma questão muito importante na genética animal (SCHMID; BENNEWITZ, 2017).

Os resultados oriundos de GWAS podem aumentar o conhecimento biológico sobre a expressão de diferentes características, assim como o entendimento sobre a arquitetura genética destas. Além disso, as funções de genes identificados com base em dados de SNPs, por meio de análises de GWAS, podem ser utilizadas para derivar e validar hipóteses sobre expressão fenotípica (MEUWISSEN; HAYES; GODDARD et al., 2016). Conseqüentemente, essas informações contribuirão para melhorar a seleção de animais em programas de melhoramento genético que utilizam informação genômica.

2.6.1 Aplicação de GWAS quando há poucos animais genotipados

Normalmente, as metodologias utilizadas para análises de GWAS na pecuária utilizam informações fenotípicas de animais genotipados e pseudofenótipos (projeções dos fenótipos de indivíduos aparentados com aqueles que foram genotipados) de animais não genotipados (LEGARRA et al., 2014). Esses procedimentos são realizados em múltiplas etapas (*multistep*) e estão sujeitas a estimativas viesadas e perda de acurácia, em casos em que a maior parte da informação fenotípica seja oriunda de animais não genotipados (WANG et al., 2014), como é a situação na maior parte dos rebanhos, principalmente em países em desenvolvimento (MRODE et al., 2018).

Uma metodologia alternativa para análise de GWAS foi proposta por Wang et al. (2012), com a finalidade de integrar todas as informações genotípicas, fenotípicas e de pedigree disponíveis (de animais genotipados ou não) em um procedimento de etapa única, denominado de *single-step* GWAS para a estimação dos efeitos de SNPs. O método ssGWAS possibilita a utilização de qualquer modelo que utilize BLUP (melhor

predição linear não viesada) para a estimação de valores genéticos (WANG et al., 2014). Com a utilização de ssGWAS, a associação genômica é realizada por meio da conversão de valores genômicos preditos (GEBVs) com uso do método *single-step* GBLUP para efeitos de marcadores e ponderações (pesos) desses marcadores, os quais são então utilizados em um processo iterativo para atualizar as soluções (valores genéticos e efeitos de SNPs) (WANG et al., 2012).

2.6.2 Utilização de janelas de SNPs em GWAS

Normalmente, nas análises de GWAS (métodos *multistep*) há a realização do teste de cada SNP de uma vez para verificar sua associação com uma característica de interesse (HAYES; GODDARD, 2009). Contudo, vários SNPs podem estar em desequilíbrio de ligação com um QTL, por estarem localizados ao redor deste no cromossomo, de modo que cada SNP explica uma parte do efeito do QTL. De acordo com Habier et al. (2011), a quantidade de QTL pode ser estimada de maneira mais precisa por meio do cálculo da variância de valores genômicos (GEBVs) explicados pelos efeitos de SNPs consecutivos (adjacentes) presentes em uma região cromossômica específica (janela). De acordo com estes autores, a partir da movimentação da janela ao longo do cromossomo e da observação de picos maiores em relação a SNPs individuais, a quantidade de QTL pode ser inferida de maneira mais eficiente, com uso de qualquer metodologia aplicada à estimação de efeitos de SNPs.

Além disso, se forem utilizadas janelas móveis de SNPs (*moving windows*), há redução na possibilidade de sub ou superestimação das variâncias das janelas de SNPs, pois, ao considerar um número fixo de SNPs, pode não haver concordância com os padrões dos blocos de haplótipos (WANG et al., 2014). Na metodologia ssGWAS, é preconizada a utilização de janelas de SNPs adjacentes para a realização da busca por genes associados às características em estudo (WANG et al., 2012, 2014).

Vários estudos de associação genômica ampla com utilização da metodologia ssGWAS já foram conduzidos, por exemplo, em aves (FRAGOMENI et al., 2014; WANG et al., 2014), bovinos (VALENTE et al., 2016; GUARINI et al., 2019; SILVA et al., 2019), ovinos (BERTON et al., 2017; AMORIM et al., 2018; ROVADOSCKI et al., 2018), suínos (SUWANNASING et al., 2018), peixes (GONZALEZ-PENA et al., 2016; BARRIA et al., 2019) e dados simulados (ZHANG et al., 2016). Nestes estudos, já foram identificados vários genes candidatos a associados a peso corporal, rendimento

de carcaça, qualidade de carne, prolificidade, resistência a doenças, temperamento e presença de chifre, por exemplo.

A escolha apropriada de genes candidatos (polimorfismos que podem influenciar a variabilidade de características produtivas), de acordo com suas supostas funções fisiológicas, ainda é uma tarefa difícil (NAVAJAS et al., 2014), de modo que a técnica do gene candidato pode não identificar novos genes e vias associadas a algumas características. Além disso, as regiões com QTL identificados são geralmente amplas, de modo que se faz necessária a realização de mais mapeamento fino e a consistência de resultados de mapeamento de QTL é limitada em populações diferentes daquela em que o estudo foi realizado (ROTHSCHILD et al., 2007).

A metodologia GWAS é uma das mais promissoras para superar essas limitações (FAN et al., 2010). Para a detecção de genes candidatos para características que não estão sob intensa seleção artificial por muitas gerações, uma maneira interessante consiste na utilização de diferentes abordagens em um mesmo estudo, por exemplo, GWAS e genômica comparativa (IGOSHIN et al., 2019).

2.7 O USO DA GENÔMICA COMPARATIVA NA PECUÁRIA

Antes da disponibilização das sequências completas dos genomas das diferentes espécies, genes candidatos eram selecionados por análise comparativa entre espécies, com limitações diretamente proporcionais à distância filogenética entre as espécies avaliadas. Com a disponibilização de informação genômica, especialmente mapas de SNPs, muitas destas limitações passaram a ser amenizadas.

A evolução das ferramentas de predição genética juntamente com avanços na biologia genômica possibilitaram a elaboração de diversos projetos de sequenciamento de genomas. Esses projetos fornecem informações detalhadas sobre o conteúdo genético e a organização dos genes tanto em diferentes indivíduos da mesma espécie, como em espécies diferentes. Com a utilização desses dados, é possível extrair informações importantes sobre as relações evolutivas entre os organismos e quais os fatores que influenciam a velocidade e a direção dessa evolução (MILLER, 2004; GIUFFRÀ; TUGGLE; FAANG, 2019).

Na pecuária, a associação de marcadores moleculares com características de interesse econômico é o caminho para analisar as regiões próximas a esses marcadores e identificar possíveis genes ou regiões candidatos à expressão e controle da característica

desejada. Entretanto, às vezes, essa investigação pode revelar pouca ou nenhuma informação sobre os genes próximos aos marcadores. Nesses casos, a análise de sequências individuais do genoma fornece muita informação da estrutura do genoma, mas pouca informação sobre a função deste (UTSUNOMIYA et al., 2013).

A genômica comparativa é o campo da genômica que vem contribuir para solucionar esse problema através da comparação das semelhanças e diferenças no conteúdo, função do gene e a organização entre os genomas de diferentes organismos, permitindo a comparação entre materiais genéticos de diferentes espécies e a extrapolação de informações entre estas (ARMSTRONG et al. 2018; PIERCE, 2016). A genômica comparativa se aplica a genomas inteiros ou regiões de diferentes espécies, subespécies diferentes ou estirpes diferentes da mesma espécie, e inclui tanto desenvolvimento de ferramentas computacionais, quanto o uso das ferramentas para analisar genomas para descobertas biológicas (WEI et al., 2002).

As análises comparativas podem investigar, entre outros objetivos, a estrutura do genoma, regiões codificadoras, e regiões sem codificação. A estrutura de diferentes genomas pode ser comparada em três níveis: estatísticas gerais de nucleotídeos, estrutura do genoma no nível do DNA e estrutura do genoma ao nível do gene. A análise comparativa das regiões de codificação entre genomas diferentes normalmente envolve a identificação de regiões codificadoras de genes, comparação do conteúdo gênico e comparação do teor de proteínas.

Para fazer essas comparações no estudo genômico e poder extrapolar informações de um organismo para o outro, é importante, entre outros conceitos evolutivos, entender as diferenças entre os termos homologia e similaridade. Homologia implica ancestralidade comum de dois genes ou produtos gênicos. Similaridade é o que se pode medir a partir do alinhamento de sequências ou estruturas, e a similaridade pode ser usada como evidência para homologia, mas não implica necessariamente homologia (WEI et al., 2002).

O sequenciamento de genomas de animais domésticos e a utilização dos princípios da genômica comparativa no contexto da pecuária podem trazer novas perspectivas para os programas de melhoramento genético animal. Esses fatores envolvem estudos de genes ou regiões de ação biológica conhecida e que estão envolvidos com o desenvolvimento ou a fisiologia de uma característica de interesse econômico. Esses estudos ajudarão a entender melhor sobre questões de saúde, bem-estar, compreensão sobre base genética de doenças, melhoramento de características de

difícil mensuração, seleção de animais superiores, entre outros benefícios (POOL; WADDELL, 2002; GODDARD; HAYES, 2009; GOES; JUNIOR; SANTOS, 2012).

2.7.1 Alinhamento e comparação de sequências biológicas

O alinhamento de sequências consiste no processo de comparação de duas sequências, de nucleotídeos ou proteínas, de forma a se observar seu nível de similaridade. A comparação de sequências biológicas é amplamente utilizada para determinar o grau de relação que pode ser estabelecido entre sequências, pois quando comparamos uma sequência de DNA de um organismo com a sequência de outros que possuem mais informações ou funções conhecidas, é possível estimar o nível de similaridade e inferir características comuns entre eles (PROSDOCIMI et al., 2007).

Essa busca por similaridade de sequências é umas das atividades de mais importância da bioinformática, e fornece a primeira evidência para a função de um novo gene sequenciado ou pedaço de sequência. Durante a comparação das sequências, para identificar e agregar informação, o que realmente importa é conseguir alinhamentos estatisticamente significativos conforme a identidade, número que indica a quantidade de nucleotídeos alinhados; similaridade, que considera a probabilidade do alinhamento ter ocorrido por acaso (*e-value*) e considera todos os outros possíveis alinhamentos; e a homologia, que se refere ao fato de dois genes partilharem ou não uma história evolutiva comum.

O alinhamento de sequências biológicas pode ser utilizado para diferentes objetivos, como por exemplo: identificação de genes e proteínas desconhecidas, na comparação da ordem de genes em genomas de organismos proximalmente relacionados (sintenia), no mapeamento de sequências expressas dentro de um genoma para identificação de genes, na montagem de genomas, e em diversas outras aplicações (ARMSTRONG, 2018; GIUFFRA; TUGGLE; FAANG, 2019).

O algoritmo de alinhamento de sequências biológicas tem que levar em consideração as possíveis mutações (inserções, deleções ou substituições) existentes ao longo do genoma. O alinhamento de duas sequências biológicas é definido como um pareamento entre os caracteres das sequências (correspondente de um nucleotídeo ou aminoácido), com a possível inserção de espaços (*gap*) entre os caracteres. Os pareamentos de caracteres iguais são chamados de *match*, e entre caracteres diferentes *mismatch* (GUSFIELD, 1997; ALURU, 2005).

Cada alinhamento possui um valor de pontuação (*score*) que indica o quanto o alinhamento tem posições similares/idênticas. Esse *score* é calculado por algoritmos que computam o valor do pareamento de cada letra no alinhamento e depois somam esses valores. O cálculo de cada alinhamento é feito com base no tipo de sequência, proteína ou DNA (MOUNT, 2004).

O alinhamento entre duas sequências pode ser feito de forma global ou local. O alinhamento global é feito quando comparamos uma sequência de aminoácidos ou nucleotídeos com outra, ao longo de toda sua extensão. O alinhamento local acontece quando a comparação entre duas sequências não é feita ao longo de toda sua extensão, mas sim através de pequenas regiões destas (Figura 2).

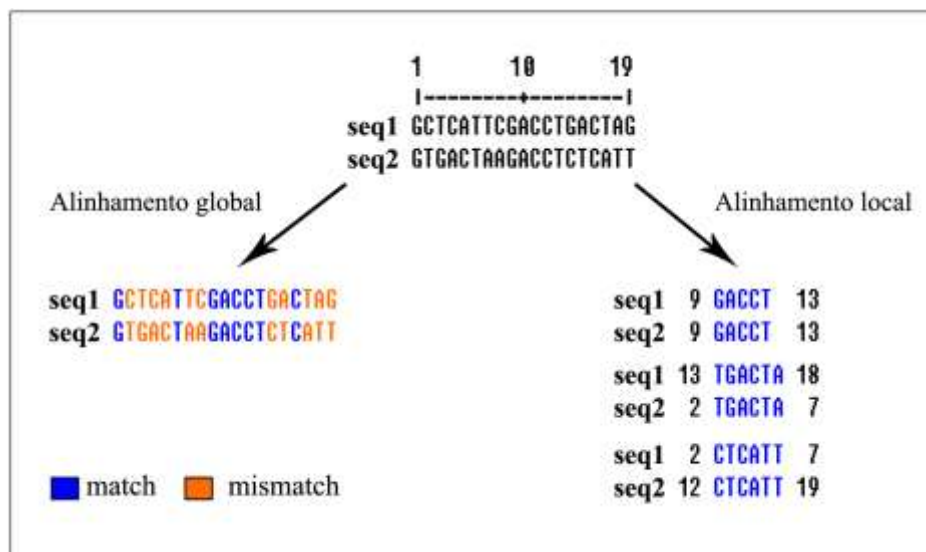


Figura 3 - Representação esquemática das duas formas de alinhamento: global e local.
 Fonte: Prosdocimi et al. (2007).

Existem alguns algoritmos de comparação que são utilizados especialmente em análises de biossequências armazenadas em bancos de dados. O programa mais usado destes é o *Basic Local Alignment Search Tool* (BLAST). Em síntese, para fazer uma pesquisa com o BLAST, um pesquisador submete uma sequência (*query*) e o programa pesquisa o banco de dados para qualquer sequência (*subject*) com regiões de alta semelhança com a sequência em questão. O programa retorna todas as sequências no banco de dados que são semelhantes, junto com informações sobre o grau de semelhança e a importância da combinação (qual probabilidade de que a semelhança ocorra por acaso sozinha) (ALTSCHUL et al., 1990, 1997; NCBI, 2008).

No BLAST, as informações sobre sequências são armazenadas no formato de apenas uma letra para cada nucleotídeo (A, T, C, G) ou aminoácidos. As informações sobre as sequências inseridas pelo pesquisador (*query*), e as sequências disponíveis nos bancos de dados (*subject*) são armazenadas no formato FASTA (PROSDOCIMI et al., 2003).

Os resultados do BLAST apresentam em seu arquivo de saída os seguintes parâmetros: o valor do escore (*Score bits*) e o valor E (*e-value*). O valor de escore depende do tamanho do alinhamento, do número de *match/mismatch/gap* e da matriz de comparação de sequências utilizada e é normalizado através de variáveis (ALTSCHUL et al., 1990, 1997).

O BLAST apresenta diferentes subprogramas que devem ser utilizados de acordo com o tipo de sequência de entrada e os bancos de dados que se deseja pesquisar. Todas as combinações entre sequências de consulta de DNA ou proteína com buscas contra bases de dados de DNA ou proteínas são possíveis (Tabela 2).

Tabela 2 - Descrição dos programas BLASTs disponíveis para cada tipo de comparação de sequências.

Programa BLAST adequado	Nome da Tarefa	Descrição
Blastp	Blastp	BLASTP tradicional para comparar uma consulta de proteína com um banco de dados de proteínas.
	Blastp-short	BLASTP otimizado para consultas com menos de 30 resíduos.
Blastn	Blastn	BLASTN tradicional que exige uma correspondência exata de 11.
	Blastn-short	Programa BLASTN otimizado para sequências menores que 50 bases.
	Megablast	Megablast tradicional usado para encontrar sequências muito semelhantes (por exemplo, intraespécies ou espécies intimamente relacionadas).
	Dcmegablast	Megablast descontínuo usado para encontrar sequências mais distantes (por exemplo, interespécies).

Fonte: Silva (2019).

Nesse contexto, para que os alinhamentos comparativos sejam confiáveis é necessário entender qual o objetivo do trabalho a ser executado e qual tipo e tarefa do BLAST é a ideal para essa finalidade.

2.8 ANÁLISE DE ANOTAÇÃO FUNCIONAL

A avaliação de propriedades funcionais de conjuntos de genes é um procedimento rotineiro para o entendimento de dados biológicos de alto desempenho e é comumente utilizada tanto para verificar que os genes implicados em um experimento biológico são funcionalmente relevantes e para descobrir funções inesperadas compartilhadas entre esses genes (GLASS; GIRVAN, 2014).

Análises de enriquecimento funcional possibilitam a identificação de enriquecimentos funcionais entre genes afetados por variantes estruturais. Por conta disso, esta metodologia tem fornecido significantes informações biológicas sobre a relação entre genótipo e fenótipo em diferentes espécies (WEBBER, 2011).

Algumas das principais fontes de informação de anotação funcional são baseadas em anotações da literatura, como *Gene Ontology* (Ashburner et al., 2000) e Enciclopedia de Kyoto de Genes e Genomas (KEGG; KANEHISA et al., 2008).

2.8.1 Termos de ontologia gênica (GO)

O projeto *Gene Ontology* (GO) representa a maior iniciativa em bioinformática com o objetivo de padronizar a representação do conhecimento biológico dos genes e produtos gênicos (por exemplo, proteínas e moléculas de RNA funcional). O projeto prevê um vocabulário controlado de termos para descrever as características dos produtos gênicos e seus dados de anotação através dos membros do Consórcio GO (<http://geneontology.org/>) e de ferramentas para acessar e processar esses dados. O projeto GO incentiva contribuições da comunidade tanto no conteúdo do GO quanto na anotação utilizando GO, e isso garante que o GO seja completo e acurado.

O projeto GO consiste de três domínios de conhecimento: função molecular (FM); processo biológico (PB) e componente celular (CC). Componente celular é definido como ‘um componente de uma célula, que é parte de um objeto maior’, como uma organela ou máquina molecular feita de vários produtos gênicos. Uma função molecular é definida como a capacidade potencial para conduzir uma atividade, como catálise no nível molecular que o produto gênico possui. Processo biológico é definido como ‘uma série de eventos atingidos por uma ou mais montagens de funções moleculares ordenadas’. Uma diferença básica entre função molecular e processo

biológico é que a FM apresenta abrangência biológica ao nível local, individual e molecular, enquanto processo biológico abrange a biologia em níveis mais altos, desde vias metabólicas à fisiologia e até mesmo comportamento de organismos (THOMAS; MI; LEWIS, 2007).

2.8.2. *Enciclopedia de Kyoto de Genes e Genomas (KEGG)*

KEGG representa uma base de dados para auxiliar no entendimento de funções de alto nível e utilidades de um determinado sistema biológico, como a célula, o organismo e o ecossistema, a partir do nível de informação molecular, especialmente bancos de dados moleculares de larga-escala gerados por sequenciamento genômico e outras tecnologias experimentais de alto rendimento (KANEHISA et al., 2008).

2.8.3. *Principais metodologias estatísticas utilizadas*

As estatísticas mais amplamente utilizadas para avaliar quais categorias funcionais são enriquecidas em um conjunto de genes são baseadas na contagem de genes e incluem o teste exato de Fisher, teste binomial e teste Chi-quadrado, por exemplo. Todas estas estatísticas dependem da pressuposição básica de que todos os genes têm igual probabilidade de serem selecionados sob a hipótese de nulidade (GLASS; GIRVAN, 2014).

Entre os testes mais utilizados, o teste exato de Fisher (TEF) é a estatística mais comum e utilizada por várias das ferramentas de enriquecimento funcional mais populares. O TEF estima o enriquecimento gênico por meio da avaliação da sobreposição entre genes em um determinado conjunto de genes experimental com genes anotados para um termo GO. Genes presentes no conjunto gênico experimental são assumidos como tendo igual probabilidade de serem identificados. Pela construção matemática, o TEF também assume que os genes anotados para um termo funcional são igualmente prováveis de serem identificados; no entanto, devido ao fato de que alguns genes são anotados para muitos termos funcionais, enquanto outros são anotados para poucos, segue-se que genes não têm probabilidade igual de serem identificados no contexto de análise funcional gênica, inconsistente com o modelo de nulidade do TEF.

Como a análise de enriquecimento funcional frequentemente envolve a comparação de um conjunto de genes a todos os termos GO, correções de múltiplas-hipóteses são geralmente aplicadas aos resultados dos testes estatísticos. Estas correções

diminuem o valor em que uma comparação entre um conjunto gênico e um termo GO deve ser considerada significativa (KHATRI; DRAGHICI, 2005).

Algumas das principais correções de múltiplas hipóteses incluem as correções de Bonferroni, Benjamini e a Taxa de Falsa Descoberta (FDR). Entre estas, a correção de Bonferroni é a mais conservativa e ajusta o valor em que um teste é considerado “significativo” pelo número de testes realizados (GLASS; GIRVAN, 2014). A FDR ajusta o valor em que um teste é considerado “significativo” baseado no rank do nível predito de significância (STOREY, 2002). Isto gera aproximadamente a mesma correção que Bonferroni para os p-valores mais significativamente ranqueados, mas não ajusta testes que sejam os menos significativos. É importante notar que, embora estas correções mudem o valor crítico dos testes individuais, elas não afetam a ordem de ranqueamento dos resultados (GLASS; GIRVAN, 2014).

2.8.4. Utilização da ferramenta DAVID para anotação funcional

A ferramenta de classificação funcional DAVID (*Database for Annotation, Visualization and Integrated Discovery*) utiliza um algoritmo de aglomeração para condensar uma lista de genes ou termos biológicos associados em classes organizadas de genes relacionados. Esta organização é alcançada por meio da mineração de complexas co-ocorrências biológicas encontradas em múltiplas fontes de anotação funcional. Assim, a ferramenta DAVID (<https://david.ncifcrf.gov/home.jsp>) apresenta um poderoso método para agrupar genes funcionalmente relacionados e termos em um número de módulos biológicos fácil de manusear para a interpretação eficiente de listas de genes (HUANG et al., 2007).

As informações aglomeradas pelo DAVID se referem a milhões de identificadores de gene/proteína oriundos a partir de uma variedade de recursos genômicos públicos agrupados dentro dos grupos do DAVID. O agrupamento desses identificadores melhora a capacidade de realização de referência cruzada, particularmente entre os sistemas NCBI e UniProt, possibilitando a integração ampla de mais de 40 fontes de anotação funcional disponíveis publicamente e a centralização pelos agrupamentos de genes (*clusters*) do DAVID. Os arquivos em formato simples de texto que compõem o DAVID são livres para baixar para vários tipos de análises de dados. Além disso, uma interface bem organizada de rede permite aos usuários gerarem

diferentes tipos de anotações heterogêneas com alta eficiência (SHERMAN et al., 2007).

REFERÊNCIAS

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403-410, 1990.

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., J. ZHANG, Z. ZHANG, W. MILLER, E, LIPMAN, D. J. Gapped blast and psi-blast: a new generation of protein database search programs. **Nucleic Acids Research**, v. 25, n. 17, p.3389-3402, 1997.

ALURU, S. Handbook of computational molecular biology. **CRC Press**, 2005.

AMORIM, S. T. et al. Genomic study for maternal related traits in Santa Inês sheep breed. **Livestock Science**, v. 217, p. 76–84, 2018.

ANDERSSON, L. Genetic dissection of phenotypic diversity in farm animals. **Nature Review Genetics**, v. 2, n. 2, p. 130–138, 2001.

ANDERSSON, L.; GEORGES, M. Domestic animal genomics: deciphering the genetics of complex traits. **Nature Review Genetics**, v. 5, n. 3, p. 202–212, 2004.

ARMSTRONG, E. et al. Novel genetic polymorphisms associated with carcass traits in grazing Texel sheep. **Meat Science**, 145, 202-208, 2018.

ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nature Genetics**, v. 25, n.1 p. 25-29, 2000.

ASHBURNER, M., GOODMAN, N. **Informatics** — genome and genetic databases. *Current Opinion in Genetics & Development*, v. 7, n. 6, p. 750-756, 1997.

AUVRAY, B. et al. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. **Journal of Animal Science**, v. 92, n. 10, p. 4375–4389, 2014.

BAGNATO, A.; ROSATI, A. From the editors – Animal selection: The genomics revolution. **Animal Frontiers**, v. 2, p. 1-2, 2012.

BARRIA, A. et al. Single-Step Genome-Wide Association Study for Resistance to *Piscirickettsia salmonis* in Rainbow Trout (*Oncorhynchus mykiss*). **G3 (Bethesda, Md.)**, v. 9, n. 11, p. 3833–3841, 2019.

BERRY, D. P. et al. Animal breeding strategies can improve meat quality attributes within entire populations. **Meat Science**, v. 132, p. 6-18, 2017.

BERTON, M. P. et al. Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. **Journal of Animal Science and Biotechnology**, v. 8, n. 73, p. 1-16, 2017.

- BOFFELLI, D.; NOBREGA, M. A.; RUBIN, E. M. Comparative genomics at the vertebrate extremes. **Nature Review Genetics**. v. 5, n. 6, p. 456–465, 2004.
- BOWMAN, J.; SANDERS, M. **Análise Genética: Uma abordagem integrada**. 1a edição. Ed. Pearson Brasil, 2014.
- BRITO, L. F. et al. Genetic parameters for various growth, carcass and meat quality traits in a New Zealand sheep population. **Small Ruminant Research**, v. 154, p. 81–91, 2017.
- COHEN, J. Bioinformatics: An Introduction for Computer Scientists. **ACM Computing Surveys**, v. 36, n. 2, p. 122–158, 2004.
- COSTA, D. A. **Genômica comparativa de linhagens de Saccharomyces e Kluyveromyces de interesse biotecnológico**. Tese (Doutorado) – Departamento de Microbiologia. Universidade Federal de Viçosa, MG, 2015.
- DEKKERS, J. C. M. Application of Genomics Tools to Animal Breeding. **Current Genomics**, v. 13, p. 207-212, 2012.
- DEROME, N. et al. A brief history of bioinformatics. **Briefings in Bioinformatics**, 2018.
- EGGEN, A. The development and application of genomic selection as a new breeding paradigm. **Animal Frontiers**, v. 2, n. 1, p. 10–15, 2012.
- FADIEL, A.; ANIDI, I.; EICHENBAUM, K. D. Farm animal genomics and informatics: an update. **Nucleic Acids Research**, v. 33, n. 19, 2005.
- FAN, B. et al. Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. **Asian-Australasian Journal of Animal Science**, v. 23, n. 7, p.833-847, 2010.
- FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Meat Market Review**. Rome, 2019. Disponível em: <<http://www.fao.org/3/ca3880en/ca3880en.pdf>>. Acesso em: 10 de agosto de 2019.
- FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Sheep and goats for diverse products and profits**. Rome, 2019. Disponível em: <<http://www.fao.org/3/i0524e/i0524e.pdf>>. Acesso em: 10 e agosto de 2019.
- FARRELL, D. Sequence, gene and protein databases: are you confused? **Bioinformatics and other bits**. 2019. Disponível em: <<https://dmnfarrell.github.io/bioinformatics/sequence-databases>>. Acesso em: 10 de outubro de 2019.
- FERRARI, R. C., GASTALDI, V. D. Precisamos falar sobre a Bioinformática. In: **VIII Botânica no Inverno 2018** / Org. Aline Possamai Della [et al.]. – São Paulo: Instituto de Biociências da Universidade de São Paulo, Departamento de Botânica, cap. 20. P. 246-261, 2018.

- FOURMENT, M., GILLINGS, M. R., A. Comparison of common programming languages used in bioinformatics. **BMC bioinformatics**. v. 9, n. 1, p. 82-91, 2008.
- FRAGOMENI, B. O. et al. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, v. 5, n. 332, 2014.
- GARCIA, J. F. et al. How Bioinformatics Enables Livestock Applied Sciences in the Genomic Era. In: de Souto M.C., Kann M.G. (eds.) *Advances in Bioinformatics and Computational Biology. BSB 2012. Lecture Notes in Computer Science*, v. 7409. Springer, Berlin, Heidelberg, 2012.
- GEORGES, M. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. **Annual Review of Genomics Human Genetics**, v. 8, p. 131–162, 2007.
- GIACHETTO, P. F.; HIGA, R. H. Bioinformática aplicada à agricultura. In: MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; LUCHIARI JUNIOR, A.; ROMANI, L. A. S. (Ed.). **Tecnologias da informação e comunicação e suas relações com a agricultura**. Brasília, DF: Embrapa, 2014. Cap. 4. p. 67-83.
- GIUFFRA E, TUGGLE CK, FAANG CONSORT. Functional Annotation of Animal Genomes (FAANG): current achievements and roadmap. **Annual Review of Animal Biosciences**, v. 7, p. 65–88, 2019.
- GLASS, K.; GIRVAN, M. Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets. **Scientific reports**, v. 4, 4191, 2014.
- GODDARD, M.E.; HAYES, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviv Genetics*, v. 10, n. 6, p. 381–391, 2009.
- GOES, P. R. N., JÚNIOR, R. A., SANTOS, J. M. G. Disponibilidade, usos e limitações dos marcadores moleculares em espécies de animais de produção. **Cesumar**, v. 14, n. 1, p. 5-16, 2012.
- GONZALEZ-PENA, D. et al. Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). **Frontiers in Genetics**. v. 7, n. 203, 2016.
- GUARINI, A. R. et al. Genetics and genomics of reproductive disorders in Canadian Holstein cattle. **Journal of Dairy Science**, v. 102, p. 1341-1353, 2019.
- GUSFIELD, D. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. **Cambridge University Press**, 1997.
- HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics**. v. 12, n. 186, 2011.
- HAGEN, J. B. The origins of bioinformatics. *Nature reviews*. **Genetics**, v. 1, p. 231–236, 2001.

HALL, A. E.; FIEBIG, A.; PREUSS, D. Beyond the arabidopsis genome: opportunities for comparative genomics. **Plant Physiology**, v. 129, n. 4, p. 1439–1447, 2002.

HAVENSTEING.; FERKET, P. R.; QURESHI, M.A. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. **Poultry Science**, v. 82, p.1500–1508, 2003.

HAYES, B.; GODDARD, M. Genome-wide association and genomic selection in animal breeding. **Genome**, v. 53, n. 11, p. 876–883, 2010.

HUANG, D.W. et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. **Genome biology**, v. 8, n. 9, R183, 2007.

IGOSHIN, A. V. et al. Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. **BMC Genetics**, v. 20, n. 26 (Suppl 1), 2019.

KADARMIDEEN, H. N. Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. **Livestock Science**, v. 166, p. 232–248, 2014.

KANEHISA, M. et al. KEGG for linking genomes to life and the environment. **Nucleic acids research**, 36(Database issue), D480–D484, 2008.

KHATRI, P.; DRAGHICI, S. Ontological analysis of gene expression data: current tools, limitations, and open problems. **Bioinformatics**, v.21, p. 3587–3595, 2005.

LANDER, E.S., WEINBERG, R.A. Genomics: Journey to the center of biology. **Science** 287: 1777–1782, 2000.

LEGARRA, A. et al. Single step, a general approach for genomic selection. **Livestock Science**, v. 166, p. 54-65, 2014.

LOBO, R.N.B. Opportunities for investment into small ruminant breeding programmes in Brazil. **Journal of Animal Breeding and Genetics**, v. 136, n.5, p. 313-318, 2019.

LÔBO, R.N.B.; LÔBO, A.M.B.O. Melhoramento genético como ferramenta para o crescimento e o desenvolvimento da ovinocultura de corte. **Revista Brasileira de Reprodução Animal**, v. 31, p. 247-253, 2007.

LUDU, J.; PLASTOW, G. Livestock and the promise of genomics. **Genome**, v. 56, p. 556–566, 2013.

MEIRA, A. N. et al. Single nucleotide polymorphisms in the growth hormone and IGF type-1 (IGF1) genes associated with carcass traits in Santa Ines sheep. **Animal**, v. 6, p. 1-9, 2019.

MEUWISSEN, T.; HAYES, B. J.; GODDARD M. E. Prediction of total genetic values using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MEUWISSEN, T.; HAYES, B. J.; GODDARD, M. E. Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers**, v.6, n.1, p. 6-14, 2016.

MORTIMER, S. I. et al. Genetic parameters for wool traits, live weight, and ultrasound carcass traits in Merino sheep. **Journal of Animal Science**, v. 95, n. 5, p. 1879-1891, 2017

MOUNT, D. M. Bioinformatics - sequence and genome analysis (2. Ed.). **Cold Spring Harbor Laboratory Press**, 2004.

MRODE, R. et al. Invited review: Genomic selection for small ruminants in developed countries: how applicable for the rest of the world? **Animal**, v. 12, n. 7, p. 1333-1340, 2018.

NAHA, B. C. et al. Concept of genome wide association studies and its progress in livestock. **International Journal of Science and Nature**, v.7, n. 1, 2016.

NAVAJAS, E.A. Animal breeding and genetics: DNA Markers and Marker-Assisted Selection in the Genomic Era. **Encyclopedia of Meat Sciences**, v.1, n.2, p.12-18, 2014.

NCBI. **Manual do usuário de aplicativos de linha de comando do BLAST**. 2008. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK279670/>>. Acesso em: 15 de julho 2019.

NICOLAZZI, E. L. Software solutions for the livestock genomics SNP array revolution. **Animal Genetics**, v. 46, n. 4, p. 343–353, 2015.

OECD/FAO - ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT / FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Agricultural Outlook 2018-2027**, OECD Publishing, Paris/FAO, Rome, 2018. Disponível em: <https://doi.org/10.1787/agr_outlook-2018-en>. Acesso em: 26 de abril de 2019.

PIERCE, B. A. **Genética: um enfoque conceitual**. 3a edição. Ed. GUANABARA KOOGAN, 2016.

POOL, R., WADDELL, K. Exploring Horizons for Domestic Animal Genomics: Workshop Summary. Washington (DC): **National Academies Press** (US); 2002.

PROSDOCIMI, F. Introdução a Bioinformática. **Apostila**. 75 p. 2007. Disponível em: <<http://professor.pucgoias.edu.br/>> Acesso em: 15 de outubro de 2019.

ROTHSCHILD, M. F. et al. Advances in QTL Mapping in Pigs. **International Journal of Biological Sciences**, v. 3, p. 192-197, 2007.

ROTHSCHILD, M. F.; PLASTOW, G. S. Applications of genomics to improve livestock in the developing world. **Livestock Science**, v. 166, p. 76-83, 2014.

ROVADOSCKI, G. A. et al. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. **BMC Genomics**, v. 19, n. 375, p. 1-14, 2018.

ROWE, J. B. The Australian sheep industry - undergoing transformation. **Animal Production Science**, v. 50, p. 991–997, 2010.

RUPP, R. et al. Genomic application in sheep and goat breeding. **Animal Frontiers**, v. 6, n. 1, p. 39-44, 2016.

SANTOS, F. R.; ORTEGA, J. M. **Bioinformática aplicada à genômica. Belo Horizonte**, 2003. Disponível em: <<http://200.17.141.88/images/3/33/bio03.pdf>>. Acesso em: 10 de outubro de 2019.

SARMENTO, J.L.R.; SENA, L.S. Avanços no melhoramento genético na América Latina e perspectivas futuras. **Archivos Latinoamericanos de Produccion Animal**, v.25, n. 1-2, p.81-89, 2017.

SCHERER, A.; CHRISTENSEN, G.B. Concepts and relevance of genome-wide association studies. **Science Progress**, v. 99, n. 1, p. 59-67, 2016.

SCHMID, M.; BENNEWITZ, J. Invited review: Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs. **Archives of Animal Breeding**, v. 60, p. 335–346, 2017.

SCHMUTZ, J.; GRIMWOOD, J. Fowl Squence. **Nature**, v.432, p.679-80, 2004.

SHARMA, A. et al. Stories and Challenges of Genome Wide Association Studies in Livestock — A Review. **Asian Australasian Journal of Animal Science**, v. 28, n. 10, p. 1371-1379, 2015.

SHERMAN, B.T. et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. **BMC bioinformatics**, v. 8, n. 426, 2007.

SILVA, L. R. G. **Genômica comparativa para identificação de genes candidatos à resistência a endoparasitas gastrintestinais na espécie Ovis aries**. Dissertação (Mestrado) — Departamento de Zootecnia, Universidade Federal do Piauí, 2019.

SILVA, R.P. et al. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. **Journal of Animal Breeding and Genetics**, v. 136, p. 118-133, 2019.

SIMIANER, H. Genomic and other revolutions – why some technologies are quickly adopted and others are not. **Animal Frontiers**, v. 6, n. 1, p. 53-58, 2016.

SOUZA, S.F. et al. Aplicação da ultrassonografia para avaliação de condição corporal e acabamento de carcaça em pequenos ruminantes. **Ciência Veterinária nos Trópicos**, v. 19, n. 3, p. 34-42, 2016.

STOREY, J. D. A direct approach to false discovery rates. **Journal of the Royal Statistics Society Series**, v. 64, p. 479–498, 2002.

SUÁREZ-DÍAZ, E. Making room for new faces: evolution, genomics and the growth of bioinformatics. **History and Philosophy of the Life Sciences**, v. 32, n. 1, p. 65–89, 2010.

SUWANNASING, R. et al. The identification of novel regions for reproduction trait in Landrace and Large White pigs using a single step genome-wide association study. **Asian-Australas Journal of Animal Science**, v. 31, n. 12, p. 1852–1862, 2018.

SWAN, A. A. et al. **Genomic evaluations in the Australian sheep industry**. In: World Congress on Genetics Applied to Livestock Production, 10, 2014, Vancouver. Proceedings... 2014.

THOMAS, P.D.; MI, H.; LEWIS, S. Ontology annotation: mapping genomic regions to biological function. **Current Opinion in Chemical Biology**, v. 11, n. 1, p. 4–11, 2007.

UTSUNOMIYA, Y. T. et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC Genetics**, v. 14, n. 52, p. 1-12, 2013.

VALENTE, T. S. et al. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. **Plos One**, v. 11, n. 6, e0156956, 2016.

VENTER, J.C. et al. The human genome sequence. **Science**, v. 291, p. 1304–1351, 2001.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research (Cambridge)**, v. 94, n. 2, p. 73-83, 2012.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 94, p. 73–83, 2014.

WEBBER, C. Functional enrichment analysis with structural variants: Pitfalls and Strategies. **Cytogenet Genome Research**, v. 135, p. 277–285, 2011.

WEI, L. et al. Comparative genomics approaches to study organism similarities and differences. **Journal of Biomedical Informatics**, v. 35, n. 2, p. 142 – 150, 2002.

ZHANG, H. et al. Progress of genome wide association study in domestic animals. **Journal of Animal Science and Biotechnology**, v. 3, n. 26, p. 1-10, 2012.

ZHANG, X. et al. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. **Frontiers in Genetics**, v. 7, n. 151, p. 1-14, 2016.

CAPÍTULO I

APLICAÇÃO DE GENÔMICA COMPARATIVA PARA A IDENTIFICAÇÃO DE GENES RELACIONADOS A CARACTERÍSTICAS DE CARÇA NA ESPÉCIE *OVIS ARIES*

Aplicação de genômica comparativa para a identificação de genes relacionados a características de carcaça na espécie *Ovis aries*

Shirliane de Araújo Sousa^{a,*}, José Lindenberg Rocha Sarmiento^b Romuere Rodrigues Veloso e Silva^c

^aPrograma de Pós-Graduação em Ciência Animal, Centro de Ciências Agrárias, Universidade Federal do Piauí, Teresina, Piauí, Brasil.

^bDepartamento de Zootecnia, Centro de Ciências Agrárias, Universidade Federal do Piauí, Teresina, Piauí, Brasil. Pesquisador do CNPq.

^cProfessor do curso de Bacharelado em Sistemas de Informação, Universidade Federal do Piauí (Campus Senador Helvídio Nunes de Barros);

*Autor correspondente

RESUMO

Características economicamente importantes da carne, como por exemplo, o rendimento de carcaça, são complexas e raramente incluídas como critério de seleção em ovinos de corte. A identificação de genes e o conhecimento dos mecanismos moleculares envolvidos no crescimento muscular e metabolismo lipídico, podem auxiliar na seleção de animais com valor genético elevado para características como a área de olho-de-lombo (AOL) e espessura da gordura subcutânea (EGS). Contribuindo assim, para a melhoria da produção e qualidade da carne ovina brasileira. Este trabalho teve por objetivo identificar potenciais genes (ou regiões genômicas) candidatos relacionados a desenvolvimento muscular e deposição de gordura, em regiões específicas do genoma da espécie *Ovis aries*, com base em regiões associadas a características de carcaça nesta espécie e a partir da comparação com genes já descritos nos genomas das espécies *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Homo sapiens* e *Gallus gallus*. Para identificação dos genes candidatos, foram delimitadas janelas do genoma da

espécie *Ovis aries* associadas às características AOL e EGS. Posteriormente, foram realizadas análises comparativas de DNA entre as janelas e genes das espécies supracitadas. Para visualizar os genomas, verificar função de genes, comparar DNA e visualizar resultados, foram utilizados o navegador Ensembl, UniProt e tarefa BLASTn da ferramenta BLAST, respectivamente. Os alinhamentos gerados pelo BLASTn passaram por filtros de identidade, RAG (Razão entre o tamanho do Alinhamento e tamanho do Gene) e *e-value*, onde foram identificados 1.097 alinhamentos significativos para AOL e 1.067 para EGS, após a filtragem. Foram identificados no genoma ovino, 26 genes candidatos imediatos relacionados a AOL e 12 para EGS. A metodologia de investigação através da genômica comparativa possibilitou a identificação dos seguintes genes candidatos: ENSOARG00000005535, LOC101104530 e ENSOARG00000005104, para AOL; e CR2: 127924965-127925647 e CR2:127924874-127925647 para EGS. Essas informações poderão dar suporte a estudos genômicos em ovinos, assim como, à identificação e seleção de animais com maior rendimento e melhor qualidade de carcaça.

Palavras-chave: Deposição de gordura, Musculosidade, SNP, Melhoramento genético.

Application of comparative genomics for identification of genes related to carcass traits in the species *Ovis aries*

ABSTRACT

Economically important characteristics of meat, such as carcass yield, are complex and rarely included in the selection criteria. Identification of genes and the knowledge on the molecular mechanisms involved in muscle growth and lipid metabolism may assist in the selection of animals with higher breeding value for traits such as loin eye area (LEA) and subcutaneous fat thickness (SFT), thus contributing for the improvement of Brazilian sheep meat. This study aimed to identify potential candidate genes (or genomic regions) related to muscle development and fat deposition in specific regions of the *Ovis aries* genome, based on regions associated with carcass traits in this species and from comparison with genes already described in the genomes of the *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Homo sapiens* and *Gallus gallus* species. To identify candidate genes, windows of the *Ovis aries* genome associated with LEA and SFT traits were delimited. Subsequently, comparative DNA analyses were performed between the windows and genes of the species mentioned above. In order to visualize

genomes, verify gene function, compare DNA and visualize results, we used the browser Ensembl, UniProt and BLASTn task, respectively. The alignments generated by BLASTn passed through identity filters, RAG (Alignment Size to Gene Size Ratio) and e-value, where 1097 significant alignments for LEA and 1067 for EGS were identified after filtering. In the sheep genome, 26 immediate candidate genes related to LEA and 12 to SFT were identified. The research methodology through comparative genomics allowed the identification of the following candidate genes: ENSOARG00000005535, LOC101104530 and ENSOARG00000005104, for LEA; and CR2:127924874-127925647 and CR2:127924874-127925647 for SFT. This information may support genomic studies in sheep, as well as the identification and selection of animals with higher yield and better carcass quality.

Keywords: Fat deposition, Muscularity, SNP, Genetic improvement.

1 Introdução

A identificação de genes e o conhecimento dos mecanismos moleculares envolvidos no crescimento muscular e metabolismo lipídico podem auxiliar na seleção de animais com valor genético elevado para características como a área de olho de lombo (AOL) e espessura da gordura subcutânea (EGS), contribuindo assim, para a melhoria da carne ovina brasileira (Ouali et al., 2013; Picar et al., 2015).

Ainda são poucos os estudos na literatura sobre utilização de informação genômica em ovinos Santa Inês (Biagiotti, 2016; Berton et al., 2017; Rego Neto, 2017; Amorim et al., 2018; Alvarenga et al., 2018; Meira et al., 2019; Rovadoscki et al., 2018; Santos, 2018; Sena, 2019), ou sobre os potenciais genes candidatos associados a características de difícil mensuração (Dekkers, 2004; Biagiotti, 2016; Silva, 2019).

O maior conhecimento sobre a arquitetura genética, juntamente com a análise do genoma de espécies de produção mais estudadas são importantes para melhorar as práticas de seleção; além de permitirem maior compreensão da base genética de doenças em outros animais e humanos, através da comparação de genomas filogeneticamente relacionados.

O conhecimento sobre quais são os genes ou regiões do genoma dessas espécies responsáveis pela manifestação de determinada característica possibilita inferir as vias metabólicas evolutivamente conservadas dessas espécies (Pool; Waddell, 2002; O'Brien et al., 2001).

Até agora, os programas de melhoramento produziram resultados satisfatórios, mas dados suplementares sobre a base molecular e informações genômicas de uma determinada característica produtiva animal podem acelerar o progresso dos sistemas de produção em pouco tempo e aumentar a relação custo-benefício. No entanto, a escolha apropriada de genes candidatos (cujos polimorfismos podem influenciar a variabilidade das características produtivas) não é uma tarefa fácil (Knapik et al., 2017).

Com base nessa problemática, este trabalho teve por objetivo identificar potenciais genes (ou regiões genômicas) relacionados ao desenvolvimento muscular e deposição de gordura, em regiões específicas do genoma da espécie *Ovis aries*; com base em regiões associadas a características de carcaça nesta espécie e a partir da comparação filogenética com genes já descritos nos genomas das espécies *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Homo sapiens* e *Gallus gallus*.

2. Material e métodos

Os dados utilizados nesta pesquisa são oriundos do projeto intitulado “Estudo genômico aplicado ao melhoramento genético de ovinos Santa Inês para características de carcaça” (número 340/17), aprovado pela Comissão de Ética no Uso de Animais (CEUA) da Universidade Federal do Piauí, que integra o Instituto Nacional de Ciência e Tecnologia em Ciência Animal (INCT-CA) do CNPq, coordenado pela Universidade Federal de Viçosa.

2.1 Base de dados utilizada nos estudos genômicos

Os dados referentes ao genoma da espécie *Ovis aries* foram baseados em resultados de um estudo de associação genômica ampla para características de carcaça em ovinos da raça Santa Inês realizado por Sena (2019). As análises de genômica comparativa realizadas tiveram como base a metodologia proposta por Silva (2019).

No estudo de Sena (2019), para as análises dos dados genômicos, após o controle de qualidade, foram utilizadas informações de 45.465 SNPs e 388 animais genotipados com o painel *OvineSNP50 BeadChip* (Illumina Inc.). A metodologia de associação genômica ampla em passo único (ssGWAS) foi utilizada para estimar os efeitos de marcadores genéticos do tipo polimorfismo de nucleotídeo único (SNP) e sua associação com as características AOL e EGS. Como resultado, foram identificadas, no

genoma ovino, regiões cromossômicas (janelas) com 10 SNPs adjacentes, que explicaram pelo menos 1,00% da variância genética aditiva para AOL e EGS.

2.2 Delimitação das janelas para a busca de genes candidatos no genoma ovino e para os estudos de genômica comparativa

Para delimitar as janelas utilizadas dentro do genoma da espécie *Ovis aries*, foram calculadas as distâncias entre os SNPs adjacentes de cada janela encontrada por Sena (2019), com base nas posições de cada SNP presente no arquivo de mapa referente ao painel OvineSNP50 (Illumina, Inc.), de acordo com o assembly do genoma ovino Oar_v4.0. O parâmetro utilizado para aumentar, de ambos os lados, o tamanho das janelas utilizadas neste estudo foi a maior distância entre um par de SNPs adjacentes localizados dentro da região compreendida entre as posições extremas de cada janela de 10 SNPs relatada no trabalho de Sena (2019) (Figura 1).

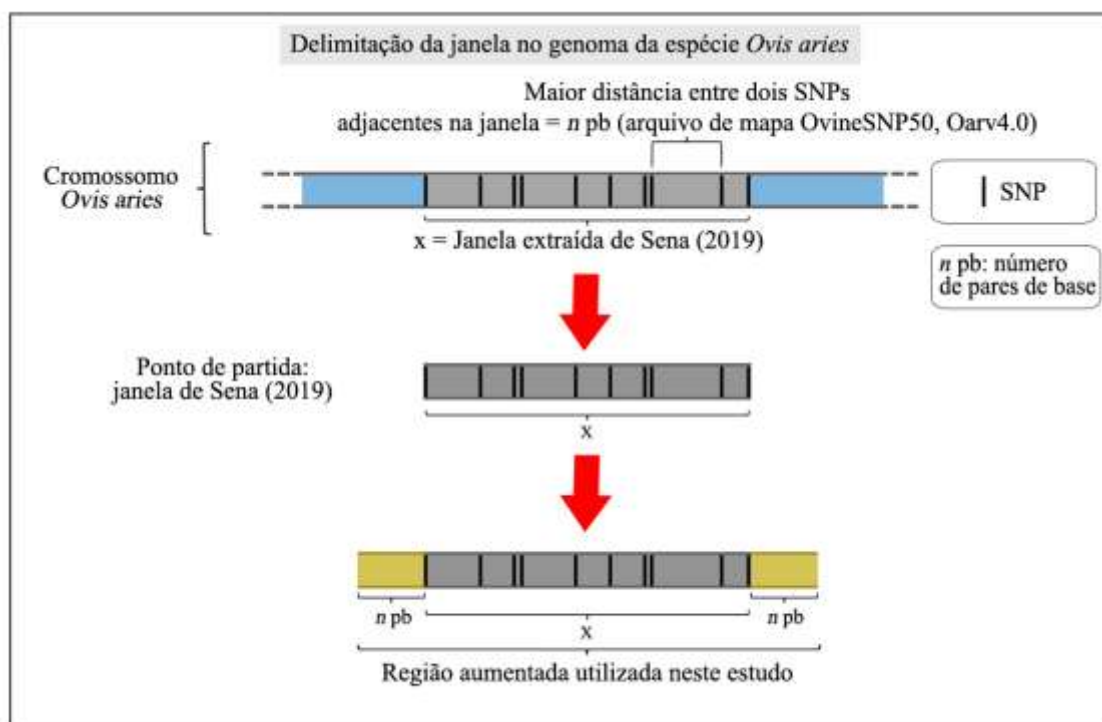


Figura 1 - Esquema que representa o processo utilizado para a delimitação das regiões (janelas) utilizadas para a busca de genes candidatos no genoma da espécie *Ovis aries*.

As regiões do genoma ovino utilizadas neste estudo encontram-se descritas na Tabela 1. Depois de delimitadas, cada região cromossômica foi visualizada no

navegador Ensembl (<https://www.ensembl.org>), com o objetivo de identificar quais são os genes candidatos já descritos e suas respectivas funções. Dentro de cada uma das regiões, buscou-se pelos genes codificantes de proteína, para identificar quais destes apresentam funções já descritas diretamente relacionadas a processos biológicos ligados à musculidade e acúmulo de gordura subcutânea. Além disso, foi verificado se existe algum relato na literatura sobre os mesmos nessas regiões cromossômicas na espécie ovina, ou em outras espécies de vertebrados.

Tabela 1 - Regiões genômicas utilizadas para a procura por genes ou regiões candidatos a características de carcaça em ovinos.

Característica	
AOL (Área de Olho de lombo)	EGS (Espessura de gordura subcutânea)
Cromossomo: posição (pb)*	Cromossomo: posição (pb)*
1:103513097-104228295	1:81071714-81913400
1:241330234-242537387	2:54188468-54975580
2:247415536-247936196	2:207470995-207913851
4:115937256-116408375	2:127905770-128537558
4:70322552-71006621	2:58322410-58860936
5:55257004-57402988	3:91300485-92021366
5:60444582-60966400	3:194994394-195730636
6:74322130-75061836	3:86150634-86987189
8:29781886-30245621	4:70720747-71174597
8:45108450-45799352	7:74774686-75252821
10:83092524-83510714	7:62601685-63816116
12:58323817-59108163	8:37820681-38326185
12:68852845-69254120	8:65187512-65848344
16:11355666-11850443	9:69199916-69711938
18:66190630-67431067	9:20187247-20652773
20:31270334-32157186	12:15959693-16781302
23:8958179-9713908	16:58656507-59047454
25:32190924-32585617	17:42338340-43046467
	23:30722503-31397888
	25:25163290-25811070

*Regiões aumentadas com base nas regiões identificadas por Sena (2019); Pb: pares de bases; AOL: Área de olho de lombo; EGS: Espessura de gordura subcutânea.

Aqueles genes que apresentaram funções descritas ou diretamente relacionadas com processos biológicos associados ao desenvolvimento muscular e deposição de gordura foram anotados e selecionados.

A função de cada gene do genoma ovino foi analisada na ferramenta Biomart, do Ensembl (<https://www.ensembl.org/biomart/martview/>), que disponibiliza termos de ontologia gênica (GO *terms*) oriundos da base de dados Uniprot (<https://www.uniprot.org/>).

2.3 Genômica comparativa para a identificação de genes candidatos a características de carcaça

A partir da análise prévia das regiões genômicas do ovino, foi realizado um estudo de genômica comparativa para verificar se algum dos genes que não foram propostos como potenciais candidatos, por não possuírem descrição de suas funções, ou por possuírem funções descritas relacionadas a uma proteína descaracterizada, poderiam ser considerados como potenciais genes candidatos. Este procedimento também foi utilizado para confirmar a função de cada gene com base na extrapolação de informação oriunda de outras espécies relacionadas via genômica comparativa.

Esta etapa consistiu em comparar as janelas do genoma ovino relacionadas a AOL e EGS de *Ovis aries* com os genes descritos nas seguintes espécies: *Capra hircus*, *Bos taurus*, *Sus scrofa*, *Mus musculus*, *Gallus gallus* e *Homo sapiens*. Estas foram denominadas neste estudo como “espécies referência”. Para a comparação das sequências de nucleotídeos, foi utilizada a ferramenta BLAST, com a tarefa BLASTn, que é específica para este tipo de análise e exige uma correspondência exata de 11 nucleotídeos para iniciar o alinhamento.

Os arquivos utilizados para as comparações estavam no formato FASTA, que é compatível com a entrada da ferramenta BLAST. No navegador genômico do Ensembl (<https://www.ensembl.org/Multi/Tools/Blast>) foram baixados as sequências correspondentes a cada janela, e de cada conjunto gênico das espécies referência a serem comparadas. Os arquivos foram submetidos à tarefa BLASTn, no qual o algoritmo realizou o alinhamento de duas sequências de DNA por vez. Então, o algoritmo realizou a busca

nas janelas do genoma ovino por regiões que tivessem similaridade com os genes das espécies referência.

2.3.1. Filtragem dos alinhamentos gerados pelo BLASTn

Os resultados dos alinhamentos gerados no BLASTn foram submetidos a três tipos de filtros (*e-value*, identidade e RAG - Razão entre o tamanho do Alinhamento e tamanho do Gene.), com a finalidade de utilizar nas análises seguintes os alinhamentos mais significativos.

Para encontrar similaridade entre as espécies avaliadas, utilizou-se a métrica de alinhamento *e-value* (*EV*), como primeiro parâmetro de filtragem dos alinhamentos. Este permite ao pesquisador ter uma ideia de quão significativos são os alinhamentos obtidos. Nesse trabalho foi utilizado $EV = 0$, ou seja, todos os alinhamentos não compatíveis com este parâmetro foram descartados. Então, foram retornados apenas os alinhamentos que não foram obtidos ao acaso.

Com o objetivo de usar apenas o melhor alinhamento de cada gene, utilizou-se mais um filtro, o valor da identidade do alinhamento. Esta métrica reflete a qualidade dos alinhamentos gerados, ou seja, dentre todos os alinhamentos gerados, foram escolhidos aqueles com os maiores valores (maiores *scores*) de identidade para prosseguir as análises.

Para a última etapa de filtragem foi utilizada a Equação 1, que representa o RAG, a razão entre os tamanhos do alinhamento e do gene. Este filtro consiste na verificação do alinhamento máximo entre o tamanho total do gene e o segmento analisado da janela, com o objetivo de identificar a proporção do gene que alinhou.

$$RAG = \frac{\text{tamanho do alinhamento}}{\text{tamanho do gene}}$$

Quanto maior o resultado desta razão, mais similares serão as duas sequências. O limiar $RAG = 0,99$ foi utilizado, ou seja, somente os alinhamentos que tiverem mais que 99% de similaridade com o gene que alinhou foram levados para a próxima etapa da análise comparativa.

2.3.2. Análise dos alinhamentos significativos e identificação de genes candidatos à características de carcaça via genômica comparativa

Após a filtragem, restaram apenas os alinhamentos mais significativos e, com estes, foram feitas as análises nas sequências de DNA. A análise dos genes presentes nas sequências das espécies referência, e sua posterior comparação com as regiões da espécie ovina, via genômica comparativa, seguiram as etapas descritas abaixo, propostas por Silva (2019).

Na primeira etapa, a função de cada gene das sequências de DNA das espécies referência foi analisada na ferramenta Biomart, do Ensembl. Se o gene da espécie referência, cujo transcrito alinhou com a região do genoma da espécie ovina, apresentou função descrita, aparente e associada com musculabilidade ou deposição de gordura, a posição do transcrito alinhado com o genoma ovino era anotada e selecionada para a etapa seguinte; caso contrário, se o gene da espécie referência não possuísse função descrita aparente e associada às características estudadas, o alinhamento era descartado nesta etapa.

A segunda etapa consistiu na visualização e avaliação, no Ensembl, da região do genoma da espécie ovina correspondente a posição que alinhou com o transcrito da espécie referência. Se a região da janela do genoma ovino é um gene relacionado com musculabilidade ou deposição de gordura, a função do gene candidato é confirmada pelo método de busca utilizado, e através disso, pode-se confirmar a eficiência do método na busca de genes candidatos.

Por outro lado, se a região do genoma ovino analisada correspondeu a um gene sem função descrita ou aparente relacionada com musculabilidade ou deposição de gordura, a informação da espécie referência foi extrapolada para o gene ovino. Assim, este gene passou a ser considerado como um potencial candidato para as características de carcaça em estudo. Entretanto, se a região do genoma da espécie *Ovis aries* ainda não era descrita como gene, esta passou a ser considerada como um potencial região candidata mediante a extrapolação de informação via genômica comparativa.

Todo o esquema do método utilizado, desde a delimitação das janelas que seriam analisadas até a extrapolação de informações dos genomas das sequências das espécies referência para o genoma ovino, está exemplificado na Figura 2.

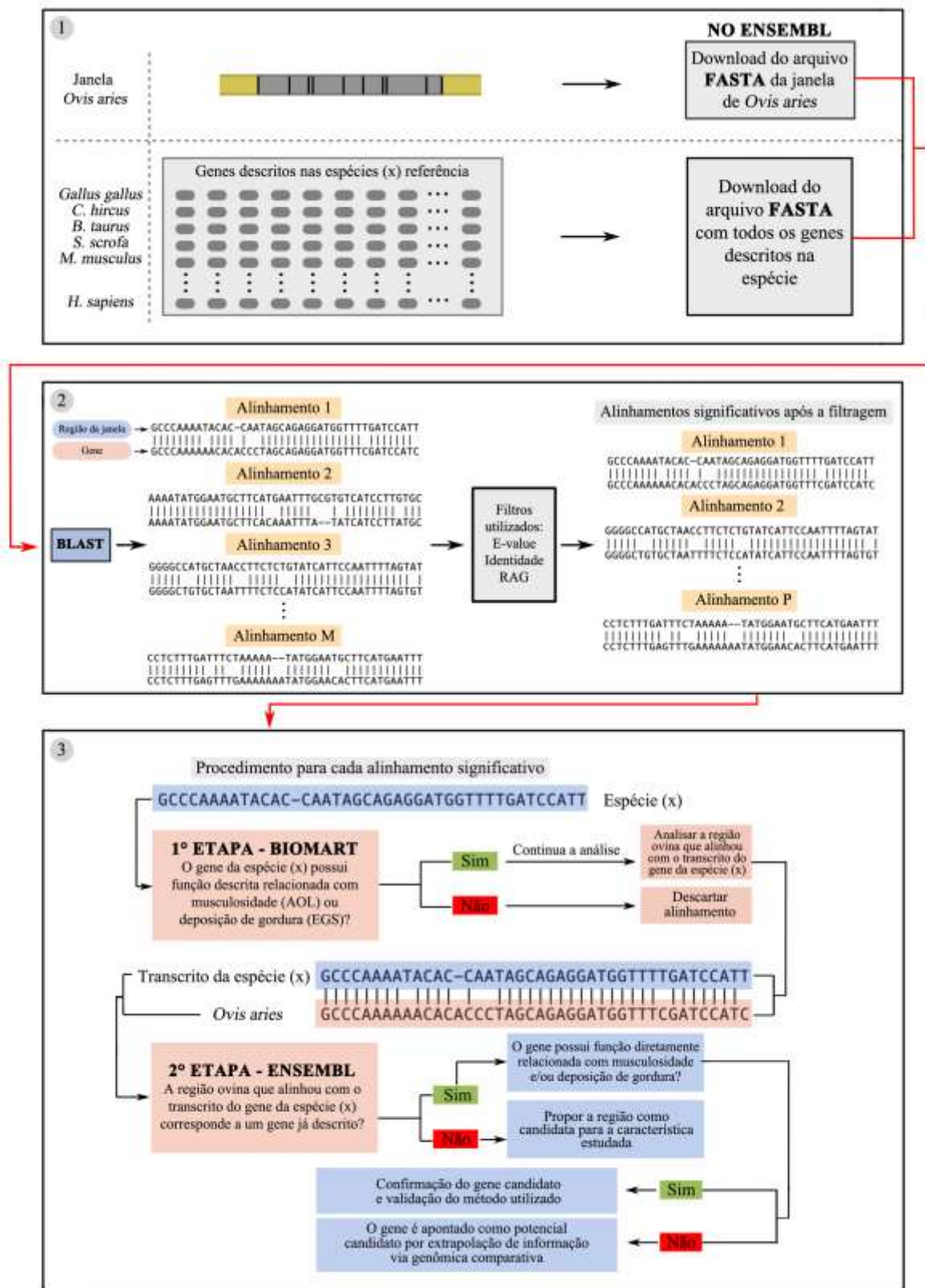


Figura 2 – Esquema que descreve o método de busca de genes utilizado neste trabalho, desde a análise dos alinhamentos significativos até a genômica comparativa para a identificação de genes candidatos às características de carcaça: 1- Etapa de seleção dos segmentos de DNA a serem alinhados. 2- Alinhamentos e filtragens gerados pelo

BLASTn. 3- Análise dos genes oriundos dos melhores alinhamentos, em busca de candidatos no genoma de *Ovis aries* via genômica comparativa.

Fonte: Figura adaptada de Silva (2019).

3 Resultados e discussão

3.1 Análise dos genes candidatos presentes nas janelas do genoma de Ovis aries

Foram analisadas regiões genômicas delimitadas (aumentadas) com base no estudo de Sena (2019), em que foram identificadas 18 regiões associadas a AOL (área de olho de lombo) e 20 janelas associadas a EGS (espessura de gordura subcutânea).

A partir da investigação das funções dos genes presentes nas regiões genômicas aumentadas, foram identificados 186 genes codificantes de proteína. Deste total, 131 genes foram identificados nas regiões relacionadas com AOL, e 55 genes estão presentes nas regiões relacionadas com EGS (Apêndice – Tabela S3 e S4).

Vários genes com funções biológicas já descritas foram identificados dentro dessas janelas. Algumas destas funções eram menos relacionadas e outras eram mais aparentemente relacionadas com as características estudadas. Estas últimas apresentaram termos biológicos (evidenciados pela ontologia gênica) bem associados com características de desenvolvimento muscular e deposição de gordura. O genes que apresentaram estas funções foram eleitos como potenciais “candidatos imediatos”, ou seja, genes que possuem função descrita, mas que não são relatadas na literatura como sendo relacionadas a musculabilidade e deposição de gordura. Todos os termos e vias *GO* mencionados neste estudo desempenham papéis importantes no metabolismo muscular e/ou lipídico e encontram-se descritos nas Tabelas S5 e S6, respectivamente.

3.1.1 Genes candidatos encontrados nas janelas relacionadas à área de olho de lombo (AOL)

Do total de 131 genes codificantes de proteína identificados nas janelas analisadas relacionadas a AOL, 26 destes apresentaram funções biológicas diretamente relacionadas com processos que influenciam na musculabilidade e no desenvolvimento muscular. Por isso, estes genes foram eleitos neste estudo como potenciais candidatos imediatos para a característica AOL (Apêndice - Tabela S5).

Importantes termos biológicos encontrados foram associados ao crescimento e proliferação celular, processo de regulação, contração e desenvolvimento de fibras musculares esqueléticas e células musculares, como por exemplo:

1. Reorganização do citoesqueleto da actina (GO:0031532);
2. Ligação do receptor tirosina quinase da (GO:0030971);
3. Ligação à ubiquitina proteína ligase (GO:0031625);
4. Atividade motora (GO:0003774);
5. Complexo miosina (GO:0016459);
6. Ligação à actina (GO:0003779);
7. Diferenciação de células musculares esqueléticas (GO:0035914);
8. Actina filamentosa (GO:0031941);
9. Regulação positiva da polimerização de filamentos de actina (GO:0030838);
10. Formação de rede de filamentos de actina (GO:0051639);
11. Desenvolvimento de fibras musculares esqueléticas (GO:0048741);
12. Desenvolvimento da fibra do músculo esquelético (GO:0048741);
13. Contração muscular (GO:0006936).

Os genes associados a esses processos foram *SHC1* (SHC adaptador de proteína 1), *SCAMP3* (Proteína 3 da membrana transportadora secretora), *MYO16* (Miosina XVI), *ATF3* (Fator de ativação da transcrição 3), *BAG5* (Atanogene 5 associado a BCL2), *CARMIL1* (regulador de proteína de nivelamento e ligador de miosina 1 1), *RIPOR2* (Regulador de polarização celular 2 que interage com a família RHO), *KCNMA1* (Subfamília M alfa 1 de canais de potássio ativados por cálcio) e *GLRA1* (Receptor de glicina alfa 1).

As proteínas actina e miosina desempenham um importante papel na contração muscular e são os principais componentes dos miofilamentos (organelas que constituem o citoesqueleto muscular) (Grant, 1993; Sgarbieri, 1996; Rottner et al., 2017). Foi demonstrado anteriormente que os genes pertencentes às famílias de actina e colágeno são expressos diferencialmente no tecido muscular de bovinos Nelore, divergindo fenotipicamente para a ingestão residual de ração (Tizioto et al., 2016). Cesar et al. (2015), encontraram que a expressão de genes da miosina no tecido muscular estava associada ao conteúdo de gordura intramuscular em bovinos Nelore.

Em contraste com o crescimento muscular, é muito importante entender quais os processos biológicos envolvidos na perda muscular, que é uma consequência debilitante de uma série de condições patológicas e metabólicas, como por exemplo, aumento na

degradação de proteínas. Encontrou-se os genes *SCAMP3* (Proteína 3 da membrana transportadora secretora) e *BAG5* (Atanogene 5 associados a BCL2) relacionados aos processos de ligação e ubiquitinação de proteínas envolvidas no processo catabólico de proteínas do músculo esquelético contrátil, que desempenha um papel importante na perda muscular.

Vários genes relacionados à ligação, transporte e resposta celular de importantes íons envolvidos no desenvolvimento muscular, como o potássio e o zinco foram identificados nessas janelas: *KCNN3* (Membro 3 da subfamília N do canal de potássio ativado por cálcio), *TRIM46* (*Tripartite motif containing 46*), *HCN3* (Canal de potássio 3 bloqueado por nucleotídeo cíclico ativado por hiperpolarização), *PKLR* (Piruvatoquinase L/R), *TRAF3* (Fator 3 associado ao receptor de TNF), *CDC42BPB* (Proteína quinase beta de ligação ao CDC42), *ENSOARG00000005104* (proteína descaracterizada), *KCNMA1* (Subfamília M alfa 1 de canais de potássio ativados por cálcio) e *GLRA1* (Receptor de glicina alfa 1).

O zinco é um elemento essencial para a nutrição, e necessário para o desenvolvimento dos animais (Suttle, 2010). Estudos demonstram que a presença do zinco aumenta o desempenho e rendimento de carcaça em bovinos e suínos (Genther-Schroeder et al., 2016; Patience e Chipman, 2011; Paulk et al., 2012). Esse íon atua no funcionamento de sistemas de enzimas e está envolvido na síntese de proteínas, metabolismo de carboidratos e muitas outras reações bioquímicas. A deficiência de zinco causa numerosas alterações patológicas, que afetam diretamente o crescimento do animal; em bovinos, caprinos e ovinos (Suttle, 2010; Miller, 1999; Genther-Schroeder; Branine; Hansen, 2016).

O potássio é o principal íon intracelular dos tecidos animais. Os gradientes estabelecidos por este íon criam um potencial elétrico essencial para a manutenção de capacidade de resposta a estímulos e tônus muscular. Todos os tecidos moles são muito mais ricos em potássio do que em sódio, tornando o potássio o terceiro mineral mais abundante no corpo animal. As maiores concentrações de potássio são encontradas em músculos. Muitas enzimas têm necessidades específicas ou facilitadoras de potássio e este elemento influencia muitas reações intracelulares envolvendo fosfato, com efeitos nas atividades enzimáticas, desenvolvimento e contração muscular (Arc, 1980; Suttle, 2010).

Os genes *PMVK* (Quinase fosfomevalonato), *GBA* (Glicosilceramidase beta), *SOAT1* (Esterol O-aciltransferase 1) e *GPLD1* (Fosfolipase D1 específica de

glicosilfosfatidilinositol) possuem termos de ontologia relacionados ao processo metabólico de lipídios e do colesterol. Alguns estudos fornecem evidências de uma relação entre o músculo esquelético e o metabolismo do colesterol. O colesterol é um constituinte importante de todas as membranas plasmáticas, incluindo o sarcolema do músculo esquelético, além de fornecer fluidez na membrana. Além disso, o colesterol é um componente de balsas lipídicas, que desempenha um papel na sinalização celular envolvida com o crescimento muscular (Simons; Ikonen, 2000; Riechman et al., 2009; Lee et al., 2019).

A massa muscular é determinada pelo equilíbrio entre síntese e degradação de proteínas que é controlado por várias vias de sinalização intracelular, como a via da insulina. A insulina é o hormônio mais importante na regulação do metabolismo energético, é essencial para a ingestão de carboidratos, síntese de proteínas e armazenamento de gordura (Dimitriadis et al., 2011; Ikeoka; Krusinova, 2009). Os genes *SHC1* (SHC adaptador de proteína 1) e *PIK3R1* (Subunidade reguladora 1 da fosfoinositida-3-quinase) possuem termo de ontologia relacionados com essas vias de sinalização intracelular.

3.1.2 Genes candidatos associados à espessura da gordura subcutânea (EGS)

Do total de 55 genes codificantes de proteína identificados nas janelas analisadas relacionadas a EGS, 12 destes apresentaram funções biológicas diretamente relacionadas com processos de deposição de gordura e metabolismo lipídico. Por isso, estes genes foram eleitos neste estudo como potenciais genes candidatos imediatos, ou seja, genes que apesar de possuírem as funções descritas, estas não são relacionadas a essa característica (Apêndice – Tabela S6).

Importantes termos biológicos encontrados foram associados aos processos biológicos e funções moleculares relacionados ao metabolismo lipídico e outros mecanismos fisiológicos que afetam diretamente a deposição de gordura, como por exemplo: Processo glicolítico (GO:0006096); Transporte de lipídios (GO:0006869); e ligação de receptores de colesterol (GO:0015485).

Os genes associados a esses processos foram *LOC101119348* (Gliceraldeído-3-fosfato desidrogenase), *OSBPL3* (Ligante de oxysterol semelhante a proteína 3),

NEUROG3 (Neurogenina 3), *HKDC1* (Hexocinase contendo domínio 1), *HK1* (Hexokinase 1) e *MDH1B* (Malato desidrogenase 1B) .

O gene *GNAQ* (Subunidade alfa q da protein G) atua na regulação da via de sinalização canônica Wnt. Esta via reprime a adipogênese ao bloquear os genes adipogênicos *PPARG* (receptor gama ativado por proliferador de peroxissomo) e *CEBPA* (CCAAT/enhancer binding protein C/EBP alfa) (Christodoulides et al., 2009).

A maioria dos genes aqui eleitos como candidatos imediatos não apresenta descrições na literatura relacionadas diretamente com as características investigadas (AOL e EGS) para a espécie ovina. Por esse motivo, faz-se necessário um melhor entendimento a respeito da caracterização das funções de genes candidatos para a característica de carcaça, com base na análise da descrição de termos de ontologia (GO terms) relacionados a cada um. Essa investigação proporciona um conhecimento mais estruturado com base nos papéis biológicos aos quais esses genes estão relacionados ou envolvidos, avaliando os mesmos de acordo com aspectos ligados à função molecular, processo biológico e componente celular.

As informações contidas na literatura podem qualificar um gene como candidato, e com esse objetivo, foram feitas buscas a respeito de registros de ocorrência dos genes aqui eleitos como candidatos imediatos para as características estudadas (AOL e EGS), na espécie *O. aries* e em outras espécies de vertebrados.

Nesse contexto, o gene *KCNN3* é relatado com uma presença abundante no músculo *Longissimus dorsi* de ovinos que apresentam o fenótipo *callipyge* (Fleming-Waddell et al., 2009). Esse fenótipo é oriundo de uma mutação caracterizada por hipertrofia muscular que proporciona efeitos benéficos na composição da carcaça (Teixeira et al., 2006).

O gene *ATF3* é relatado como regulador da expressão de vários genes envolvidos no controle da proliferação e diferenciação celular, e está envolvido na regulação das mudanças de expressão gênica no músculo *Longissimus dorsi* de indivíduos *callipyge* (Vuocolo et al., 2007). Em bovinos Angus, o gene *ATF3* também é relatado envolvido no metabolismo lipídico (Zhao et al., 2012).

O gene *EIF2AK2* (fator de iniciação da tradução eucariótica 2 alfa-cinase 2) não apresentou termos de ontologia gênica bem relacionados a processos biológicos ligados com EGS. Com base na literatura, o gene *EIF2AK2* está relacionado ao acúmulo de gordura subcutânea e é um indutor chave de inflamação, responsivo a interferons (IFN), que está associado ao stress do retículo endoplasmático e ao acúmulo de gordura

(Ramos-Lopez et al., 2018). Em estudo com bovinos Nelore, foi concluído que o gene *EIF2AK2* pode ser considerado como candidato para a seleção para espessura de gordura subcutânea (Silva-Vignato et al., 2019). Por esse motivo, este gene foi eleito como candidato imediato.

O gene *ADAM23* (Domínio ADAM metalopeptidase 23) também não apresentou termos de ontologia bem relacionados a processos biológicos ligados com EGS; mas em bovinos da raça Nelore, o ortólogo deste gene foi encontrado em uma região (BTA2: 95 Mpb) que explicou 2,84% da variância para perfil de ácidos graxos saturados na gordura intramuscular do músculo *Longissimus thoracis* (Lemos et al., 2016); e em camundongos, este gene age como um receptor de glioma inativado 3 rico em leucina, responsável pela supressão da adipogênese (Kim et al., 2012). Com base nessas descrições o gene foi candidatado também na espécie ovina.

O gene *OSBPL3* (Ligante de oxysterol semelhante a proteína 3), eleito como gene candidato imediato neste estudo, também foi relatado como relacionado a espessura de gordura em bovinos da raça Nelore (Santana et al., 2015).

3.2 Resultado gerado pela busca dos alinhamentos ótimos no Blastn

Após a filtragem (E-value, Identidade e RAG) dos alinhamentos gerados pelo BLASTn, restaram apenas os alinhamentos mais significativos. Do total de 2.145.285 alinhamentos relacionados a AOL, restaram 1.097 alinhamentos ótimos para serem analisados; e do total de 2.291.980 alinhamentos relacionados a EGS, restaram 1.067 alinhamentos ótimos para serem analisados (Tabelas 2 e 3).

Tabela 2 - Quantidade de alinhamentos (AOL) por janelas antes e após os filtros de *E-value*, Identidade e RAG.

Janela de <i>O. aries</i> (AOL) (Cromossomo:Região)	Quantidade de alinhamentos	Quantidade de Alinhamentos pós filtros		
		(<i>E-Value</i>)	(Identidade)	(RAG)
1:103513097-104228295	70019	3101	434	5
1:241330234-242537387	250508	23962	445	16
2:247415536-247936196	74473	2140	315	76
4:115937256-116408375	74867	2530	377	2
4:70322552-71006621	101618	5206	396	5
5:55257004-57402988	310171	26802	508	54
5:60444582-60966400	80504	5861	408	15
6:74322130-75061836	166913	15161	492	37
8:29781886-30245621	72031	5914	388	14
8:45108450-45799352	119390	9923	442	138
10:83092524-83510714	82643	3408	383	35
12:58323817-59108163	126653	6620	463	37
12:68852845-69254120	67743	5106	455	168
16:11355666-11850443	75935	4727	401	47
18:66190630-67431067	122705	8634	531	187
20:31270334-32157186	129122	7025	492	121
23:8958179-9713908	161368	9377	440	104
25:32190924-32585617	58622	3611	414	36
Total	2145285			1097

AOL: área de olho de lombo; RAG: Razão entre o tamanho do Alinhamento e tamanho do Gene.

Tabela 3 - Quantidade de alinhamentos (EGS) por janelas antes e após os filtros de *E-value*, Identidade e RAG.

Janela de <i>O. aries</i> (EGS) (Cromossomo :Região)	Quantidade de alinhamentos	Quantidade de Alinhamentos pós filtros		
		(<i>E-Value</i>)	(Identidade)	(RAG)
1:81071714-81913400	179874	21109	477	177
2:54188468-54975580	121613	12003	430	12
2:207470995-207913851	80929	5772	394	18
2:127905770-128537558	110662	11153	490	101
2:58322410-58860936	84743	3360	399	4
3:91300485-92021366	110384	7448	454	8
3:194994394-195730636	153248	6018	421	53
3:86150634-86987189	127595	10232	456	137
4:70720747-71174597	63706	3458	388	81
7:74774686-75252821	87899	4758	359	52
7:62601685-63816116	199033	16993	466	62
8:37820681-38326185	80261	3601	395	3
8:65187512-65848344	133479	15520	453	66
9:69199916-69711938	95493	8673	437	75
9:20187247-20652773	80173	2902	438	105
12:15959693-16781302	169634	13899	425	11
16:58656507-59047454	77996	2105	366	14
17:42338340-43046467	124267	12832	427	35
23:30722503-31397888	115897	6434	435	49
25:25163290-25811070	95094	5627	380	4
TOTAL	2291980			1067

EGS: espessura de gordura subcutânea; RAG: Razão entre o tamanho do Alinhamento e tamanho do Gene.

*3.3. Identificação de genes candidatos para características de carcaça na espécie *Ovis aries* via genômica comparativa*

Nete tópico, a partir das análises de genômica comparativa, foram encontrados alguns genes diferentes nas espécies referência, que apresentaram funções ligadas ao metabolismo muscular e à deposição de gordura.

3.3.1. Área de olho de lombo

A partir das análises de genômica comparativa, foram encontrados sete genes diferentes nas espécies referência que apresentaram funções ligadas ao metabolismo muscular. O alinhamento dos transcritos desses genes com a janela do genoma ovino possibilitou a confirmação de um gene candidato imediato identificado nas análises anteriores (subtópico 3.1.1) dentro das janelas referentes a AOL em ovinos; e a identificação de três genes candidatos para AOL na espécie *O. aries* (Tabela 4 e Figura S1).

Tabela 4 – Regiões de alinhamentos (alinham.) entre *Ovis aries* e as espécies referência para a característica área de olho de lombo.

Gene (espécie) / Transcrito (pos. ¹)	Gene ou região em <i>O. aries</i> (pos. ¹)	Alinham. do transcrito em <i>O. aries</i> (pos. ¹)
<i>HNRNPA2B1</i> (<i>Bos taurus</i>) ENSBTAT00000007527.4 (4:69735135-69744610)	<i>ENSOARG00000005535*</i> (18:66659730-66660789)	18:66659736-66660789
<i>HNRNPA2B1</i> (<i>Capra hircus</i>) ENSCHIT00000033734.1 (4:50471410-50479052)		18:66659736-66660789
<i>HNRNPA2B1</i> (<i>Mus musculus</i>) ENSMUST00000203220.2 (6:51462760-51469772)		18:66659736-66660785
<i>HNRNPA2B1</i> (<i>Sus scrofa</i>) ENSSSCT00000062290.1 (18:46205436-46215650)		18:66659736-66660789
<i>HNRNPA2B1</i> (<i>Homo sapiens</i>) ENST00000608362.2 (7:26173057-26201301)		18:66659736-66660789
<i>ENSCHIG00000015804</i> (<i>Capra hircus</i>) ENSCHIT00000022805.1 (2:5783728-5784837)		18:66659737-66660814
<i>BAG5</i> (<i>Capra hircus</i>) ENSCHIT00000016598.1 (21:67467948-67469291)		<i>BAG5</i> (18:66875079-66876452)
<i>BAG5</i> (<i>Mus musculus</i>) ENSMUST00000054636.6 (12:111709488-111713257)	18:66875079-66876422	
<i>BAG5</i> (<i>Gallus gallus</i>) ENSGALT00000075288.2 (5:50598446-50603397)	18:66875079-66876422	
<i>BAG5</i> (<i>Homo sapiens</i>) ENST00000445922.2 (14:103556545-103562318)	18: 66875079-66876422	
LOC108638534 (<i>Capra hircus</i>) ENSCHIT00000013701.1 (21:67154609-67155211)	LOC101104530* (18:66561182-66561784)	18:66561182-66561784
ENSBTAG00000047319 (<i>Bos taurus</i>) ENSBTAT00000063737.2 (21:67884902-67885504)		18:66561182-66561784
GM266 (<i>Mus musculus</i>) ENSMUST00000010673.6 (12:111484609-111485823)		18:66561184-66561784
(<i>Bos taurus</i>) ENSBTAT00000017850.5 (4:77284046-77335575)	ENSOARG00000005104* (20:31631834-31632688)	20:31631819-31632688

¹Posição no genoma da espécie em pares de base. *Genes candidatos via genômica comparativa.

Os transcritos do gene *BAG5* nas espécies *Capra hircus*, *Mus musculus*, *Gallus gallus* e *Homo sapiens* alinharam com o homólogo deste gene na espécie ovina. Nas diferentes espécies, este gene está ligado a atividades de ligação de chaperone e ligação de proteína quinase, que são importantes para o desenvolvimento muscular em vertebrados (Smith et al., 2014). Apesar de não possuir função descrita na literatura para a espécie ovina, relacionada a AOL, este gene foi eleito como candidato imediato quando analisados os termos de ontologia nas janelas do genoma ovino, e esse resultado comparativo confirma essa candidatura e a aplicabilidade funcional do método comparativo descrito por Silva (2019). A convergência de funções com as espécies referências dão indícios que esta se trata de uma região filogeneticamente conservada entre as espécies relacionadas (Wei et al., 2002).

O gene *HNRNPA2B1* (ribonucleoproteína nuclear heterogênea A2/B1) foi identificado nas espécies referência *Bos taurus*, *Capra hircus*, *Mus musculus*, *Sus scrofa* e *Homo sapiens*. Esse gene foi considerado importante devido a sua participação em processos ligados a mRNA (RNA mensageiro) e miRNA (micro RNA), nas diferentes espécies. O miRNA desempenha papel importante na miogênese e tem grande impacto na manutenção da massa muscular e tipo de fibra muscular desenvolvido (Wang et al., 2018).

O gene ovino *ENSOARG00000005535*, cuja posição alinhou com os transcritos dos homólogos do gene *HNRNPA2B1* nas demais espécies referência, é uma proteína não-caracterizada na espécie ovina. Desse modo, o gene *ENSOARG00000005535* passou a ser considerado candidato a AOL, por extrapolação de informação via genômica comparativa.

O gene *LOC101104530* da espécie ovina, não caracterizado, alinhou com os transcritos dos genes das espécies referência *Capra hircus*, *Bos taurus* e *Mus musculus*. Contudo, apenas o gene *GM266* (*Mus musculus*) é caracterizado e apresenta o transcrito *ENSMUST00000010673.6*. O gene *GM266* atua na transdução de sinal, que é um importante processo biológico que controla o crescimento muscular e hipertrofia em mamíferos (Wackerhage e Ratkevicius, 2008). Embora não caracterizados, os genes *ENSBTAG00000047319* (*Bos taurus*) e *LOC108638534* (*Capra hircus*), cujos transcritos alinharam com *LOC101104530*, também estão envolvidos na transdução de sinal (GO:0007165). Desse modo, o gene *LOC101104530* passa a ser considerado candidato a AOL, por extrapolação de informação via genômica comparativa.

O gene *BLVRA* da espécie referência *Bos taurus*, cujo transcrito alinhado com o gene *ENSOARG00000005104*, em ovinos, atua na ligação de um importante íon envolvido no desenvolvimento muscular, o zinco. Este mineral melhora o desempenho e composição de carcaça em animais de produção (Patience e Chipman, 2011). O gene ovino *ENSOARG00000005104*, cuja posição alinhada com o transcrito do gene *BLVRA*, é uma proteína não-caracterizada. Desse modo, o gene ovino mencionado acima passa a ser considerado candidato a AOL, por extrapolação de informação via genômica comparativa.

3.3.2. Espessura de gordura subcutânea

A partir das análises de genômica comparativa, foram encontrados oito genes diferentes nas espécies referência, que apresentaram funções ligadas à deposição de gordura. Os alinhamentos dos transcritos desses genes com a janela do genoma ovino possibilitaram a confirmação de três genes candidatos imediatos identificados nas análises anteriores (subtópico 3.1.2) dentro das janelas referente a EGS em *O. aries*; e a identificação de duas regiões candidatas para EGS na espécie *Ovis aries* (Tabela 5 e Figura S2).

Os genes identificados nas espécies referência relacionados à deposição de gordura foram escolhidos devido à participação desses em processos como desenvolvimento do pâncreas, processo metabólico da glicose e apoptose.

Os transcritos do gene *NEUROG3* nas espécies *Bos taurus*, *Capra hircus* e *Sus scrofa* alinharam com o homólogo deste gene na espécie ovina. Nas diferentes espécies, este gene está ligado à participação em processos no desenvolvimento do pâncreas. O pâncreas secreta insulina e glucagon (que se unem para manter a homeostase de glicose) diretamente na corrente sanguínea, que estão diretamente envolvidos no regulamento de carboidratos, proteínas e metabolismo de lipídio (Suckale e Solimena, 2008).

Os transcritos do gene *LOC101119348* nas espécies *Capra hircus*, *Gallus gallus* e *Sus scrofa* alinharam com o homólogo deste gene na espécie ovina. Nas diferentes espécies, este gene está ligado à participação no processo glicolítico e processo metabólico de glicose. A diminuição da atividade glicolítica no músculo resulta em menor utilização muscular de glicose, o que pode resultar em maior acúmulo de glicose em todo o corpo e aumentar a adiposidade da carcaça, pois a glicose é um forte estimulador da lipogênese (Bazile et al., 2019).

Tabela 5 – Regiões de alinhamentos (alinham.) entre *Ovis aries* e as espécies referência à característica espessura de gordura subcutânea.

Gene (espécie) / Transcrito (pos. ¹)	Gene ou região em <i>O. aries</i> (pos. ¹)	Alinham. c/ transcrito em <i>O. aries</i> (pos. ¹)
<i>NEUROG3</i> (<i>Bos taurus</i>) ENSBTAT00000008432.5 (28:25906029-25906824)	<i>NEUROG3</i> (25:25706463-25707128)	25:25706461-25707128
<i>NEUROG3</i> (<i>Capra hircus</i>) ENSCHIT00000030070.1 (28:19757099-19757746)		25:25706461-25707128
<i>NEUROG3</i> (<i>Sus scrofa</i>) ENSSSCT00000064994.1 (14:72604468:72606363)		25:25706461-25707128
<i>GAPDH</i> (<i>Bos taurus</i>) ENSBTAT00000037753.5 (5:103870384-103874667)	<i>LOC101119348</i> (1:81887323-81888323)	1:81887323-81888320
<i>GAPDH</i> (<i>Capra hircus</i>) ENSCHIT00000033836.1 (5:102789927-102794072)		1:81887323-81888320
<i>ENSCHIG00000026172</i> (<i>Capra hircus</i>) ENSCHIT00000039940.1 (11:43687157:43688110)		1:81887378-81888320
<i>ENSGALG00000014442</i> (<i>Gallus gallus</i>) ENSGALT00000023323.6 (1:77619214:77623236)		1:81887325-81888320
<i>ENSCHIG00000008960</i> (<i>Capra hircus</i>) ENSCHIT00000012351.1 (10:41638555:41639556)		1:81887323-81888320
<i>GAPDH</i> (<i>Sus scrofa</i>) ENSSSCT00000045248.1 (5:64129679:64133991)		1: 81887323-81888296
<i>CDC42EP3</i> (<i>Bos taurus</i>) ENSBTAT00000032505.4 (11:20086106:20111391)		<i>CDC42EP3</i> (3:86336269-86337033)
<i>CDC42EP3</i> (<i>Sus scrofa</i>) ENSSSCT00000023650.2 (3:102539223:102562389)	3:86336269-86337033	
<i>CDC42EP3</i> (<i>Mus musculus</i>) ENSMUST00000068958.8 (17:79333727:79355091)	3:86336269-86337032	
<i>LOC102183607</i> (<i>Capra hircus</i>) ENSCHIT00000013486.1(11:10159639:10160337)	2:127924965-127925647* (região candidata)	2:127924965-127925647
<i>ENSCHIG00000021399</i> (<i>Capra hircus</i>) ENSCHIT00000031934.1 (7:768791:769556)	2:127924874-127925647* (região candidata)	2:127924874-127925647

¹Posição no genoma da espécie em pares de base. *Genes candidatos via genômica comparativa.

Os transcritos do gene *CDC42EP3* (Proteína efetora CDC42 3) nas espécies *Bos taurus*, *Sus scrofa*, e *Mus musculus* alinharam com o homólogo deste gene na espécie ovina. Nas diferentes espécies, este gene está ligado à formação do citoesqueleto de actina e transporte de proteínas. Além das funções de contração das membranas celulares para motilidade, diferenciação ou divisão celular, o citoesqueleto de actina fornece forças para moldar e mover as membranas intracelulares de organelas e vesículas (Rottner et al., 2017).

Embora não apresentem função descrita na literatura para a espécie ovina, relacionada a EGS, os genes *NEUROG3*, *LOC101119348* e *CDC42EP3* foram eleitos como candidatos imediatos quando analisados os termos de ontologia nas janelas do genoma ovino. Esse resultado comparativo confirma essa candidatura e a aplicabilidade funcional do método comparativo.

Dois regiões genômicas ainda não descritas foram identificadas via genômica comparativa. Uma está localizada no cromossomo 2 (127924874-127925647) do genoma ovino, cuja posição alinhou com o transcrito do gene *LOC102183607* da espécie referência *Capra hircus*. Embora este gene não seja caracterizado, o mesmo apresenta funções relacionadas à regulação do processo apoptótico. A outra região identificada está localizada no cromossomo 2 (127924874-127925647) do genoma ovino, cuja posição alinhou com o transcrito do gene *ENSCHIG00000021399* da espécie referência *Capra hircus*, que também apresenta funções relacionadas à regulação do processo apoptótico.

As duas regiões da janela de *Ovis aries* ainda não descritas como gene passam a ser consideradas como um potenciais genes candidatos, mediante a extrapolação de informação dos respectivos transcritos dos genes das espécie referência via genômica comparativa.

4. Conclusões

As análises das janelas do genoma de *O. aries* possibilitaram a identificação de 38 genes candidatos imediatos, dos quais 26 são relacionadas à musculabilidade, e 12 à deposição de gordura subcutânea. As análises de genômica comparativa entre o genoma da espécie *O. aries* e os genomas das espécies utilizadas como referência confirmaram a aplicabilidade do método utilizado e a candidatura imediata dos genes selecionados para as características em estudo.

O fato de um gene ter sua função descrita não significa que a mesma esteja associada diretamente na literatura com determinadas características de interesse, como as de interesse zootécnico. Nesse estudo, foi possível a confirmação e associação direta de genes que já apresentavam funções descritas, mas que estas não se encontravam relacionadas na literatura com a musculosidade e deposição de gordura em ovinos, por isso estes receberam aqui a denominação de “candidatos imediatos”.

Novas regiões, ainda não descritas e relacionadas com as características investigadas nesse estudo foram identificadas e propostas como potenciais candidatas: três genes para AOL (ENSOARG00000005535, LOC101104530 e ENSOARG00000005104) e duas regiões ainda não descritas para EGS (CR2: 127924965-127925647 e CR2:127924874-127925647).

Essas informações poderão dar suporte a estudos genômicos em ovinos, assim como, à identificação e seleção de animais com maior rendimento e melhor qualidade de carcaça.

Referências

- Alvarenga, A.B., Rovadoscki, G. A., Petrini, J., Coutinho, L. L., Morota, G., Spangler, M. L., Pinto, L. F. B., Carvalho, G. G. P., Mourão, G. B., 2018. Linkage disequilibrium in Brazilian Santa Inês breed, *Ovis aries*. Scientific Reports, v.8, n.1.
- Amorim, S.T., Kluska, S., Berton, M. P., de Lemos, M. V. A., Peripolli, E., Stafuzza, N. B., Martin, J. F., Álvarez, M. S., Gaviña, B. V., Toro, M. A., Banchemo, G., Oliveira, P. S., Grigoletto, L., Eler, J. P., Baldi, F. Ferraz, J. B. S., 2018. Genomic study for maternal related traits in Santa Inês sheep breed. Livestock Science, v.217, p.76-84.
- Arc, 1990. The Nutrient Requirements of Ruminants. Common wealth Agricultural Bureaux, Farnham, Royal, UK, pp. 184–185.

- Bazile, J., Picard, B., Chambon, C., Valais, A., Bonnet, M., 2019. Pathways and biomarkers of marbling and carcass fat deposition in bovine revealed by a combination of gel-based and gel-free proteomic analyses. *Meat Science*, v. 156, p.146–155.
- Berton, M.P., de Oliveira Silva, R. M., Peripolli, E., Stafuzza, N. B., Martin, J. F., Álvarez, M. S., Gavinã, B. V., Toro, M. A., Banchemo, G., Oliveira, P. S., Eler, J. P., Baldi, F., Ferraz, J. B. S., 2017. Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. *Journal of Animal Science and Biotechnology*, v.8, n.73.
- Biagiotti, D., 2016. Associação e seleção genômica ampla em ovinos Santa Inês para características relacionadas a resistência à endoparasitas, Teresina: Centro de Ciências Agrárias, Universidade Federal do Piauí, p. 73, Tese de Doutorado.
- Cesar, A. S., Regitano, L.C., Koltés, J. E., Fritz-Waters, E. R., Lanna, D. P., Gasparin, G., 2015. Putative regulatory factors associated with intramuscular fat content. *PLoS One*, 10 (6), p. 1-21.
- Dekkers, J. C. M., 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science*, v.82, (E-SUPPL): E313–E328.
- Dimitriadis, G., Mitrou, P., Lambadiari, V., Maratou, E., Raptis, S. A., 2011. Insulin effects in muscle and adipose tissue. *Diabetes Research and Clinical Practice*, p. 93.
- Fleming-waddell, J. N., Olbricht, G. R., Taxis, T. M., White, J. D., Vuocolo, T., Craig, B. A., Tellam, R. L., Neary, M. K., Cockett, N. E., Bidwell, C. A., 2009. Effect of DLK1 and RTL1 but not MEG3 or MEG8 on muscle gene expression in Callipyge lambs. *PloS one*, v. 4, e7399.

- Genther-Schroeder, O. N., Branine, M. E., Hansen S. L., 2016. The effects of increasing supplementation of zinc-amino acid complex on growth performance, carcass characteristics, and inflammatory response of beef cattle fed ractopamine hydrochloride. *Journal of Animal Science*, 94:3389-3398.
- Grant, A. L., Skjaerlund, D. M., Helferich, W. G., Bergen, W. G., MERKEL, R. A., 1993. Skeletal muscle growth and expression of skeletal muscle α s-actin mRNA and insulin-like growth factor I mRNA in pigs during feeding and withdrawal of ractopamine. *Journal of Animal Science*, v. 71, p. 3319–3326.
- Ikeoka D, Krusinova E., 2009. Insulin resistance and lipid metabolism. *Revista da Associação Médica Brasileir*, 55(3):1.
- Kim, H. A., Park, Woo-Jae, Jeong, Hyo-Soon, Lee, Hyun-e, Lee, S., Kwon, N., Baek, K., Kim, Dong-Seok, Yun, Hye-Young., 2012. Leucine-rich glioma inactivated 3 regulates adipogenesis through ADAM23. *Biochim Biophys Acta*, v. 1821, n. 6, p.914–922.
- Knapik, J.; Ropka-Molik, K.; Pieszka, M., 2017. Genetic and nutritional factors determining the production and quality of sheep meat – a review. *Annals of Animal Science*, v. 17, n. 1, p. 23–40.
- Lee, T. V., Chang, L. W., Vicente, W. C., Steve, B., James, F. D., Steven, R. E., 2019. The effects of hindlimb unloading versus dietary cholesterol and resistance training on rat skeletal muscle responses. *Lipids in Health and Disease*, 18:3, p. 12.
- Lemos, M.V.A., Chiaia, H.L.J., Berton, M.P., Feitosa, F.L.B., Aboujaoud, C., Camargo, G.M.F., Pereira, A.S.C., Albuquerque, L.G., Ferrinho, A.M., Mueller, L.F., Mazalli, M.R., Furlan, J.J.M., Carvalheiro, R., Gordo, D.M., Tonussi, R., Espigolan, R., Silva, R.M.O., Oliveira, H.N., Duckett, S., Aguilar, I., Baldi, F.,

2016. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nelore cattle using the single step procedure. *BMC Genomics* 17, 213. <https://doi.org/10.1186/s12864-016-2511-y>.
- Meira, A. N., Montenegro, H., Coutinho, L. L., Mourão, G. B., Azevedo, H. C., Muniz, E. N., Machado, A. L., Sousa-Jr, L. P., Pedrosa, V. B., Pinto, L. F. B., 2019. Single nucleotide polymorphisms in the growth hormone and IGF type-1 (IGF1) genes associated with carcass traits in Santa Ines sheep. *Animal*, v. 13, p. 460-468.
- Miller, W. J., 1999. Zinc Nutrition of Cattle: A Review. *Journal of dairy science*. vol. 3. N.8. University of Georgia, Athens.
- O'brien, J. S., eizirik, E., William, J. M., 2001. On Choosing Mammalian Genomes for Sequencing. *Science, New Series*, v. 292, No. 5525, p. 2264-2266.
- Ouali, A., Gagaoua, M., Boudida, Y., Becila, S., Boudjellal, A., Herrera-Mendez, C. H., Sentandreu, M. A., 2013. Biomarkers of meat tenderness: present knowledge and perspectives. *Meat Science* 95(4):854-70.
- Patience, J. F., Chipman, A., 2011. Impact of zinc source and timing of implementation on grow-finish performance, carcass composition and locomotion score. *Animal Industry Report: AS 657*.
- Paulk, C. B., K. J. Prusa, M. D. Tokach, J. L. Nelssen, J. M. Derouchey, J. M. Goodband, D. Robert, AND S. S. Dritz., 2012. Effects of added Zn in diets with ractopamine HCl on growth performance and carcass quality of finishing pigs in a commercial environment. *Kansas Agricultural Experiment Station Research Reports*, v. 0, ed. 10.
- Ramos-Lopez, O., Ramos-Lopez, O., Riezu-Boj, J. I., Milagro, F. I., Martinez, J. A., 2018. DNA methylation signatures at endoplasmic reticulum stress genes are

- associated with adiposity and insulin resistance. *Molecular Genetics Metabolism*, v. 123, p. 50–58.
- Rego Neto, A., 2017. Estrutura genética e associação genômica ampla para características de tamanho corporal em ovinos da raça Santa Inês, Teresina: Centro de Ciências Agrárias, Universidade Federal do Piauí, 89 p. Tese de Doutorado.
- Rottner, K., faix, J., bogdan, S., linder, S., kerkhoff, E., 2017. Actin assembly mechanisms at a glance. *Journal of Cell Science*, p. 130, 3427-3435.
- Rovadoski, G.A., 2018. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics*, v.19, n.375.
- Santana, M.H.A., Ventura, R.V., Utsunomiya, Y.T., Neves, H.H.R., Alexandre, P.A., Oliveira Junior, G.A., Gomes, R.C., Bonin, M.N., Coutinho, L.L., Garcia, J.F., Silva, S.L., Fukumasu, H., Leme, P.R., Ferraz, J.B.S., 2015. A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. *Journal of Animal Breeding and Genetics*, v.132, p. 420-427.
- Santos, G. V., 2018. Estudo genômico aplicado ao melhoramento genético de ovinos tropicais para resistência à endoparasitas, Teresina: Centro de Ciências Agrárias, Universidade Federal do Piauí, 90 p. Tese de Doutorado.
- Sena, L. S., 2019. Estudo genômico aplicado ao melhoramento genético de ovinos Santa Inês para características de carcaça, Teresina: Centro de Ciências Agrárias, Universidade Federal do Piauí, 125 p. Tese de Doutorado.

- Silva, L. R. G. 2019. Genômica comparativa para identificação de genes candidatos à resistência a endoparasitas gastrintestinais na espécie *Ovis aries*. Dissertação (Mestrado) — Departamento de Zootecnia, Universidade Federal do Piauí.
- Silva-Vignato, B., Coutinho, Luiz L., Poleti, Mirele D., Cesar, Aline S. M., Moncau, Cristina T., Regitano, Luciana C. A., Balieiro, Júlio C. C., 2019. Gene co-expression networks associated with carcass traits reveal new pathways for muscle and fat deposition in Nelore cattle. *BMC Genomics*, v. 20, n. 32.
- Sgarbieri, V. C., 1996. Proteínas em alimentos protéicos. São Paulo: Varela.
- Smith, D. A., Carland, C.R., Guo, Y. and Bernstein, S.I., 2014. Getting Folded: Chaperone proteins in muscle development, maintenance and disease. *Anat Rec (Hoboken)*, v. 297, n. 9, p. 1637–1649.
- Suckale, J., Solimena, M., 2008. Pancreas islets in metabolic signaling-focus on the beta-cell. *Frontiers in Bioscience*, 13, p. 7156–7171.
- Suttle, N. F. ZINC., 2010. In: *The Mineral Nutrition of Livestock*. 4th ed. p. 426-458. CABI Publishing, New York.
- Teixeira, C.S., Oliveira, D.A.A., Quirino, C.R., 2006. Musculatura dupla: II - Determinação genética. *Arch. Latinoam. Prod. Anim.* 14, 17-23. <http://www.bioline.org.br/pdf?la06003>.
- Tizioto, P.C., Coutinho, L. L., Oliveira, P. S. N., Cesar, A. S, Diniz, W.J., Lima, A. O., 2016. Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake. *Scientific Reports*, 6:39493.
- Vuocolo, T., Byrne, K., White, J., McWilliam, S., Reverter, A., Cockett, N., Tellam, R., 2007. Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. *Physiology Genomics*, 12, 253-272.

- Wang, J., Yang, L.Z., Zhang, J.S., Gong, J.X., Wang, Y.H., Zhang, C.L., Chen, H., Fang, X.T., 2018. Effects of microRNAs on skeletal muscle development. *Gene*, v. 668, p.107–113.
- Wackerhage, H.; Ratkevicius, A., 2008. Signal transduction pathways that regulate muscle growth. *Essays Biochem*, v. 44, p. 99–108.
- Wei, L., Liu, Y., Dubchak, I., Shon, J., Park, J., 2002. Comparative genomics approaches to study organism similarities and differences. *J. Biomed. Inform.* 35, 142–150. [https://doi.org/10.1016/s1532-0464\(02\)00506-3](https://doi.org/10.1016/s1532-0464(02)00506-3).
- Zhao, C., Tian, F., Yu, Y., Luo, J., Mitra, A., Zhan, F., Hou, Y., Liu, G., Zan, L., Updike, M.S., Song, J., 2012. Functional genomic analysis of variation on beef tenderness induced by acute stress in Angus cattle. *Comp. Funct. Genomics*, 2012, 756284. <https://doi.org/10.1155/2012/756284>.

CAPÍTULO II

ANÁLISE DE ENRIQUECIMENTO GÊNICO PARA CARACTERÍSTICAS DE CARÇA EM OVINOS

Análise de enriquecimento gênico para características de carcaça em ovinos

Shirliane de Araújo Sousa^{a,*}, José Lindenberg Rocha Sarmiento^b, Romuere Rodrigues

Veloso e Silva^c

^aPrograma de Pós-Graduação em Ciência Animal, Centro de Ciências Agrárias, Universidade Federal do Piauí, Teresina, Piauí, Brasil.

^bDepartamento de Zootecnia, Centro de Ciências Agrárias, Universidade Federal do Piauí, Teresina, Piauí, Brasil. Pesquisador do CNPq.

^cProfessor do curso de Bacharelado em Sistemas de Informação, Universidade Federal do Piauí (Campus Senador Helvídio Nunes de Barros);

*Autor correspondente: shirliane.araujo@uece.br

RESUMO

Em ovinos, ainda são escassas informações sobre a arquitetura genética de características de carcaça e sobre os mecanismos fisiológicos em que os genes ligados a estas participam. Com base em ferramentas de anotação funcional, é possível realizar o agrupamento de genes para a obtenção de mapas para facilitar sua interpretação biológica. Portanto, o presente estudo teve o objetivo de realizar análises de anotações funcionais de genes presentes em regiões genômicas relacionadas aos fenótipos área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS), mensuradas por ultrassonografia em ovinos da raça Santa Inês. As análises de anotação funcional para averiguar quais as funções enriquecidas com base nas listas de genes utilizadas foram realizadas considerando o p-valor < 0,05 para termos de ontologia gênica (GO) e vias KEGG, com uso da ferramenta DAVID. As listas de genes utilizadas nas análises foram geradas por meio da busca em regiões genômicas aumentadas a partir de janelas de 10 SNPs adjacentes significativamente associadas a AOL e EGS, identificadas por meio de análise de associação genômica ampla com uso do método *single step* (ssGWAS).

Importantes termos de ontologia e vias KEGG relacionados ao metabolismo proteico e metabolismo lipídico foram identificados com utilização dos conjuntos gênicos analisados. A identificação de termos em comum, com a utilização de genes relacionados às diferentes características em estudo, sugere que há relação pleiotrópica entre desenvolvimento muscular e deposição de gordura em ovinos da raça Santa Inês.

Palavras-chave: anotação funcional, *longissimus dorsi*, DAVID, ontologia, *Ovis aries*

Gene enrichment analysis for carcass traits in sheep

ABSTRACT

In sheep, there is still a paucity of information about the genetic architecture of carcass traits and about the physiological mechanisms in which genes linked to these traits. Based on functional annotation tools, it is possible to cluster genes in order to use maps to facilitate biological interpretation. Therefore, the objective of this study was to perform a functional annotation analysis of genes present in the genetic regions related to loin eye (LEA) and subcutaneous fat thickness (SFT), measured by ultrasonography in Santa Inês sheep. The functional annotation analyses for checking which functions were enriched based on the lists of genes used were performed considering p-value <0.05 for gene ontology terms (GO) and KEGG pathways, using the DAVID tool. The gene lists used in the analyses were generated by searching in genetic regions augmented from windows of 10 adjacent SNPs associated to LEA and SFT, identified by single-step genomic association (ssGWAS) analysis. Important terms of ontology and KEGG pathways related to protein and lipid metabolism were identified using the analyzed gene sets. The identification of common terms using genes related to the different traits in study suggests that there is a pleiotropic relationship between muscle development and fat deposition in Santa Inês sheep.

Keywords: functional annotation, *longissimus dorsi*, DAVID, ontology, *Ovis aries*

1 Introdução

Alguns dos principais fatores que limitam a implantação de critérios para a seleção de animais de corte com maior rendimento de carcaça são: complexidade biológica do desenvolvimento muscular e deposição de gordura subcutânea; carência de informações sobre as bases fisiológicas dessas características; desenvolvimento tardio do músculo *longissimus dorsi* e do tecido adiposo; e necessidade da coleta em animais que já tenham atingido bom nível de desenvolvimento corporal.

Esses fatores sugerem que há necessidade de utilização de ferramentas genômicas para a obtenção de melhorias no processo de seleção desses animais. Neste sentido, a detecção de genes e *loci* de características quantitativas (QTLs) apresentam grande importância, pois é possível utilizar estas informações na seleção assistida por marcadores (Hay e Roberts, 2018).

Tecnologias como os Estudos de Associação Genômica Ampla (GWAS - *Genome Wide Association Studies*), permitem a identificação de variações no DNA que estão correlacionadas com uma determinada característica, apontando para a localização dos genes envolvidos com a expressão do fenótipo de interesse e auxiliando na compreensão dos mecanismos biológicos e fisiológicos dos genes (Zhang et al., 2012).

As análises de GWAS geram saídas de resultados referentes a regiões genômicas em que, normalmente, estão presentes vários genes, que resultam em grandes listas. As ferramentas e os métodos que têm sido utilizados para a interpretação de listas de genes, com base em informações disponíveis em bancos de dados biológicos, objetivam basicamente avaliar a frequência dos termos funcionais da lista de genes, aplicando testes estatísticos para determinar os termos significativamente representados ou enriquecidos (Webber, 2011).

Neste contexto, informações provenientes de diferentes fontes como *Gene Ontology* (GO) (Ashburner et al., 2000) e vias KEGG (*Kyoto Encyclopedia of Genes and Genomes*) são comumente utilizadas, possibilitando a geração de mapas e informações que agrupam padrões comuns de diferentes interações entre os componentes do sistema, por meio do enriquecimento funcional de genes (Kanehisa e Goto, 2000).

Alguns trabalhos de associação genômica com inclusão de análises de enriquecimento gênico em bovinos já possibilitaram a identificação de importantes processos biológicos e vias metabólicas associados a desenvolvimento muscular e

deposição de gordura, com base na área de olho de lombo e espessura de gordura subcutânea mensuradas *in vivo* (Silva et al., 2017; Silva et al., 2019).

Em ovinos, ainda não há relatos de estudos desse tipo com utilização das características mencionadas anteriormente. Assim, ainda há carência de informações sobre a arquitetura genética de fenótipos de carcaça nesta espécie e sobre os mecanismos fisiológicos em que os genes ligados a estas características participam. Portanto, o objetivo com esta pesquisa foi utilizar genes codificantes presentes em regiões relacionadas à área de olho de lombo e à espessura de gordura subcutânea, mensurados por ultrassonografia, para o enriquecimento funcional de termos de ontologia e vias metabólicas para estes fenótipos.

2. Material e métodos

Os dados utilizados nesta pesquisa são oriundos do projeto intitulado “Estudo genômico aplicado ao melhoramento genético de ovinos Santa Inês para características de carcaça” (número 340/17), aprovado pela Comissão de Ética no Uso de Animais (CEUA) da Universidade Federal do Piauí, que integra o Instituto Nacional de Ciências e Tecnologia em Ciência Animal (INCT-CA) do CNPq, coordenado pela Universidade Federal de Viçosa.

2.1. Base de dados genômicos utilizados nas análises de enriquecimento funcional

Os dados referentes ao genoma da espécie *Ovis aries* foram baseados em resultados de um estudo de associação genômica ampla para características de carcaça mensuradas por ultrassonografia em ovinos da raça Santa Inês realizado por Sena (2019).

No referido estudo, para as análises dos dados genômicos, após o controle de qualidade, foram utilizadas informações de 45.465 polimorfismo de nucleotídeo único (SNPs) e 388 animais genotipados com o painel OvineSNP50 BeadChip (Illumina Inc.). A metodologia de associação genômica ampla em passo único (ssGWAS) foi utilizada para estimar os efeitos de SNPs e sua associação com as características área de olho de lombo (AOL) e espessura de gordura subcutânea (EGS) mensuradas por ultrassonografia.

Como resultado, foram identificadas, no genoma ovino, 18 e 20 regiões cromossômicas (janelas) com 10 SNPs adjacentes, que explicaram pelo menos 1,00% da

variância genética aditiva para AOL e EGS, respectivamente. Essas regiões foram utilizadas como referência para a delimitação das posições das janelas utilizadas no presente trabalho (Tabelas S1 e S2).

2.2. Delimitação das janelas para a busca de genes candidatos no genoma ovino

Para delimitar as janelas utilizadas dentro do genoma da espécie *Ovis aries*, foram calculadas as distâncias entre os SNPs adjacentes de cada janela encontrada por Sena (2019), com base nas posições de cada SNP presente no arquivo de mapa referente ao painel OvineSNP50 (Illumina, Inc.), de acordo com o assembly do genoma ovino Oar_v4.0. O parâmetro utilizado para aumentar, de ambos os lados, o tamanho das janelas utilizadas neste estudo foi a maior distância entre um par de SNPs adjacentes localizados dentro da região compreendida entre as posições extremas de cada janela de 10 SNPs relatada no trabalho de Sena (2019) (Figura 1).

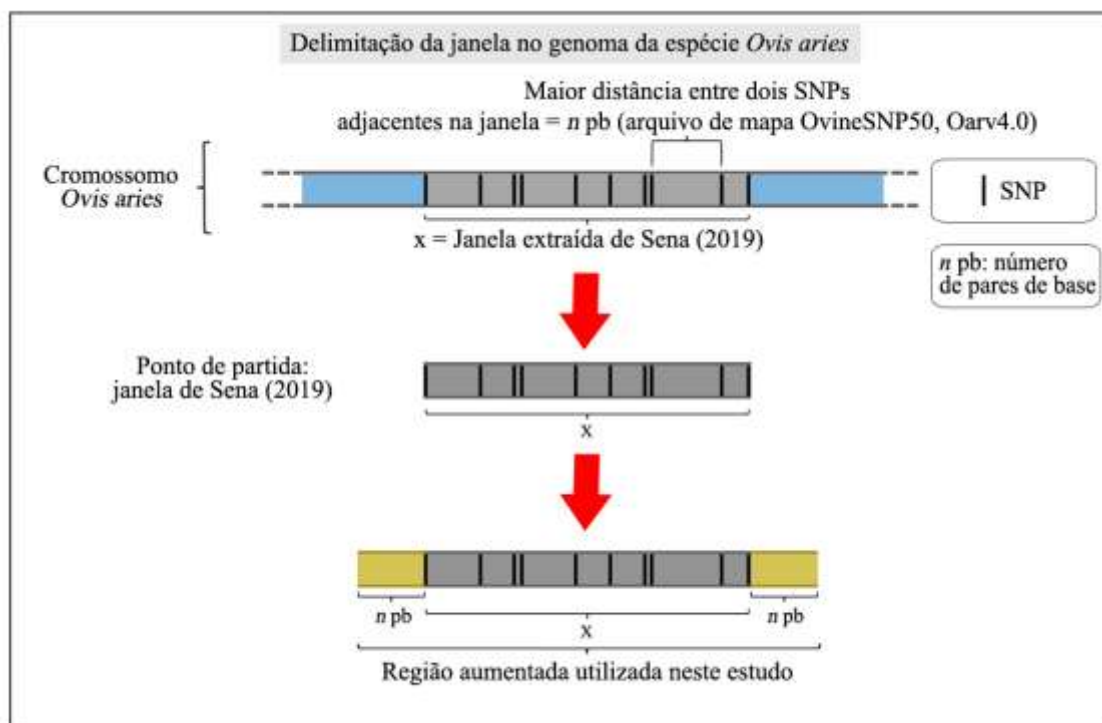


Figura 1 – Delimitação de cada janela da espécie *Ovis aries* utilizada.

As janelas correspondentes, com as regiões aumentadas do genoma ovino, utilizadas no presente estudo encontram-se descritas na Tabela 1.

Tabela 1 - Regiões do genoma da espécie *Ovis aries* utilizadas para a busca por genes.

Característica			
AOL		EGS	
Cromossomo	Região genômica (pb) *	Cromossomo	Região genômica (pb) *
1	103513097-104228295	1	81071714-81913400
1	241330234-242537387	2	54188468-54975580
2	247415536-247936196	2	207470995-207913851
4	115937256-116408375	2	127905770-128537558
4	70322552-71006621	2	58322410-58860936
5	55257004-57402988	3	91300485-92021366
5	60444582-60966400	3	194994394-195730636
6	74322130-75061836	3	86150634-86987189
8	29781886-30245621	4	70720747-71174597
8	45108450-45799352	7	74774686-75252821
10	83092524-83510714	7	62601685-63816116
12	58323817-59108163	8	37820681-38326185
12	68852845-69254120	8	65187512-65848344
16	11355666-11850443	9	69199916-69711938
18	66190630-67431067	9	20187247-20652773
20	31270334-32157186	12	15959693-16781302
23	8958179-9713908	16	58656507-59047454
25	32190924-32585617	17	42338340-43046467
		23	30722503-31397888
		25	25163290-25811070

*Regiões aumentadas com base nas regiões identificadas por Sena (2019); AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; pb: pares de bases.

Depois de delimitadas as regiões cromossômicas do genoma ovino a serem utilizadas, as posições de todas as regiões foram inseridas na ferramenta Biomart do navegador Ensembl (<https://www.ensembl.org/biomart/martview/>), que gera informações sobre as funções dos genes a partir de termos de ontologia gênica (GO terms) disponíveis na base de dados Uniprot (<https://www.uniprot.org/>).

Dentro de cada região, buscou-se por todos os genes codificantes de proteína, para gerar uma lista com os genes presentes nas regiões relacionadas a cada característica. Logo após, foram identificados quais destes genes apresentavam funções aparentes já descritas e associadas às características em estudo, e se existe algum relato na literatura, nesse sentido, sobre estes genes na espécie ovina, ou em outras espécies de vertebrados. Então, aqueles genes que apresentaram funções descritas ou relacionadas com desenvolvimento muscular e deposição de gordura foram anotados e selecionados como potenciais candidatos para AOL e EGS.

2.3. Análises de enriquecimento e anotação gênica

A análise de anotação funcional para o enriquecimento gênico de termos, de ontologia gênica (GO) ($p < 0,05$) e vias metabólicas da Enciclopédia de Genes e Genomas de Kyoto (*KEGG*; $P < 0,05$), foi realizada com auxílio da ferramenta DAVID (*Database for Annotation, Visualization and Integrated Discovery*) v6.8 (Huang et al., 2009a, 2009b). Para as análises, todos os genes anotados no genoma de *Ovis aries* foram considerados como base (*background*).

As identificações dos genes presentes nas listas analisadas foram baseadas na nomenclatura oficial de cada gene (*Official Gene Symbol*), de modo que aqueles que não estavam de acordo com esta codificação foram convertidos na ferramenta DAVID. Os genes não identificados pela ferramenta foram eliminados automaticamente e não entraram nas listas de genes utilizadas para as análises. Do total de 131 genes identificados nas regiões relacionadas a AOL, 98 IDs (identificações) da espécie *Ovis aries* foram mapeados pela ferramenta DAVID, a partir dos quais foram gerados os resultados da análise do conjunto gênico para AOL. Para EGS, 55 genes foram identificados nas regiões analisadas para esta característica, dos quais 49 foram mapeados pelo DAVID e utilizados nas análises.

As análises de anotação funcional dos genes identificados nas regiões analisadas do genoma ovino, para a identificação de termos de ontologia gênica (GO) e vias *KEGG*, foram realizadas, inicialmente, considerando todos os genes codificantes presentes nas regiões genômicas avaliadas para cada característica. Em seguida, foi conduzida análise com a utilização conjunta de todos os genes presentes nas regiões genômicas relacionadas às características AOL e EGS.

Após esta etapa, as análises foram realizadas apenas com os potenciais genes candidatos identificados com base na análise das funções na espécie ovina, para cada característica individualmente (AOL – Tabela S5 ou EGS – Tabela S6). Finalmente, foi utilizada uma análise com o conjunto gênico completo que incluía os potenciais genes candidatos com funções descritas e claramente relacionadas à característica de interesse, e os demais genes candidatos. Dos genes presentes em regiões relacionadas a AOL, 24 foram mapeados pelo DAVID e utilizados nas análises, enquanto que, para EGS, apenas 12 genes foram utilizados.

3 Resultados e discussão

3.1 Anotação funcional a partir do conjunto de genes por característica

Dos 131 genes codificadores de proteína presentes nas regiões analisadas para AOL (Apêndice – Lista de genes), 98 IDs da espécie *Ovis aries* foram mapeados pela ferramenta DAVID, a partir dos quais foram gerados os resultados da análise do conjunto gênico para AOL. Para EGS, 55 genes foram identificados nas regiões analisadas para esta característica, dos quais 49 foram utilizados nas análises (Tabela 2).

Tabela 2 - Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de genes presentes nas regiões selecionadas no genoma ovino.

Fenótipo	Termo GO	Tipo	p-valor	Genes	^a FDR
AOL	GO:0016477~ migração celular	PB	1,20 ⁻²	<i>ABL2, CDC42BPB, CARMIL1, EFNA1</i>	15,00
	GO:0005622~ Intracelular	CC	1,90 ⁻²	<i>HTR5A, ADAM15, G3BP1, RALGPS2, SHC1, WSB1, EFNA1, LOC101104530, TRIM46</i>	18,00
	GO:0010976~ regulação positiva do desenvolvimento da projeção de neurônio	PB	1,90 ⁻²	<i>ABL2, FBXO38, CD24,</i>	23,00
EGS	GO:0060113~ diferenciação de células receptoras do ouvido interno	PB	8,4 ⁻³	<i>DFNA5, BMP4</i>	10,00
	GO:0004396~ atividade da hexoquinase	FM	1,50 ⁻²	<i>HK1, HKDC1</i>	14,00
	GO:0060712~ desenvolvimento da camada espongiotrofoblasto	PB	2,00 ⁻²	<i>BIRC6, ZFAT</i>	23,00
	GO:0005536~ ligação de glicose	FM	2,40 ⁻²	<i>HK1, HKDC1</i>	21,00
	GO:0001678~ homeostase de glicose celular	PB	2,50 ⁻²	<i>HK1, HKDC1</i>	28,00
	GO:0000186~ ativação da atividade de MAPKK	PB	3,30 ⁻²	<i>BMP4, EIF2AK2</i>	36,00
	GO:0051260~ homooligomerização proteica	PB	4,20 ⁻²	<i>NLRC4, KCTD1, SPAST</i>	43,00

AOL:área de olho de lombo; EGS:espessura de gordura subcutânea; PB: processo biológico; FM: função molecular; CC: componente celular; ^aFDR:Taxa de falsa descoberta (%).

No termo de ontologia mais significativamente enriquecido (migração celular) a partir dos genes relacionados a AOL, estão presentes genes que participam diretamente da migração celular e de processos relacionados a este. A migração celular é um processo central no desenvolvimento e manutenção de organismos multicelulares, pois determina o movimento de uma célula de um local para um destino guiado por pistas moleculares. A atividade migratória celular é fundamental para a atividade de células precursoras da miogênese, como mioblastos e miócitos (Griffin et al., 2010).

Com a utilização do conjunto gênico relacionado a EGS, os genes *HK1* (hexoquinase 1) e *HKDC1* (hexoquinase contendo domínio 1) participaram de três termos de ontologia diretamente relacionados ao metabolismo de glicose. A glicose apresenta importância fundamental para todos os organismos vivos, principalmente para a geração de energia (ATP) e, em mamíferos, pode ser armazenada no organismo na forma de glicogênio, amido e gordura (Saltiel, 2016). A atividade da hexoquinase (GO:0004396) é essencial para o metabolismo da glicose, pois a reação catalizada pela hexoquinase representa a primeira etapa no metabolismo de glicose, por meio da fosforilação para formar glicose-6-fosfato (Wilson, 2003).

As análises de vias KEGG com os genes presentes nas regiões delimitadas para AOL possibilitaram a identificação das vias de sinalização Ras, PI3K-Akt e ErbB (Tabela 3 e Figuras 2 e 3), que atuam diretamente no desenvolvimento e homeostase muscular (Andrechek et al., 2002; Glass, 2010).

Tabela 3 - Vias KEGG significativamente enriquecidas ($P < 0,05$) para as características em estudo a partir de genes presentes nas regiões selecionadas no genoma ovino.

Característica	KEGG	p-valor	Genes	FDR
AOL	oas04014: via de sinalização Ras	$2,00^{-3}$	<i>ABL2, SHC1,</i> <i>EFNA1, EFNA3,</i> <i>EFNA4, PIK3R1</i>	2,30
	oas04151: via de sinalização PI3K-Akt	$1,40^{-2}$	<i>EFNA1, EFNA3,</i> <i>EFNA4, PIK3R1,</i> <i>PPP2R2B, THBS3</i>	15,00
	oas04012: via de sinalização ErbB	$4,99^{-2}$	<i>ABL2, SHC1,</i> <i>PIK3R1</i>	45,00
	oas05222: câncer de pulmão de pequenas células	$5,40^{-3}$	<i>CKS1B,</i> <i>LOC101107954,</i> <i>TRAF3, PIK3R1</i>	6,10
	oas00524: biosíntese de butirosina e neomicina	$1,00^{-2}$	<i>HK1, HKDC1</i>	10,00
EGS	oas04066: via de sinalização HIF-1	$1,70^{-2}$	<i>ANGPT1, HK1,</i> <i>HKDC1</i>	17,00

AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; PB: processo biológico; FM: função molecular; CC: componente celular; ^aFDR: Taxa de falsa descoberta (%).

Alguns componentes do sistema renina-angiotensina (RAS) desempenham função importante na regulação da regeneração muscular em condições fisiológicas, de modo que, a desregulação de RAS pode deteriorar a capacidade regenerativa do músculo (Delafontaine e Yoshida, 2016). A sinalização da regulação da proteína quinase B (PI3k-Akt) signaling (oas04151) é uma sinalização proteica fundamental nas vias metabólicas celulares que causa hipertrofia muscular e crescimento de tecidos em geral

(Fujita et al., 2007). A família ErbB de receptores de tirosina quinases (RTKs) acopla a ligação de ligantes do fator de crescimento extracelular a vias de sinalização intracelular que regulam diversas respostas biológicas, incluindo proliferação, diferenciação, motilidade celular e sobrevivência (Kanehisa et al., 2008). O receptor ErbB2 da família ErbB atua na diferenciação de mioblastos e na manutenção dos fusos musculares (Andrechek et al., 2002).

Com o conjunto de genes presentes nas regiões delimitadas para EGS, foi possível a identificação da via de sinalização HIF-1, que atua na regulação do metabolismo lipídico (Mylonis et al., 2019).

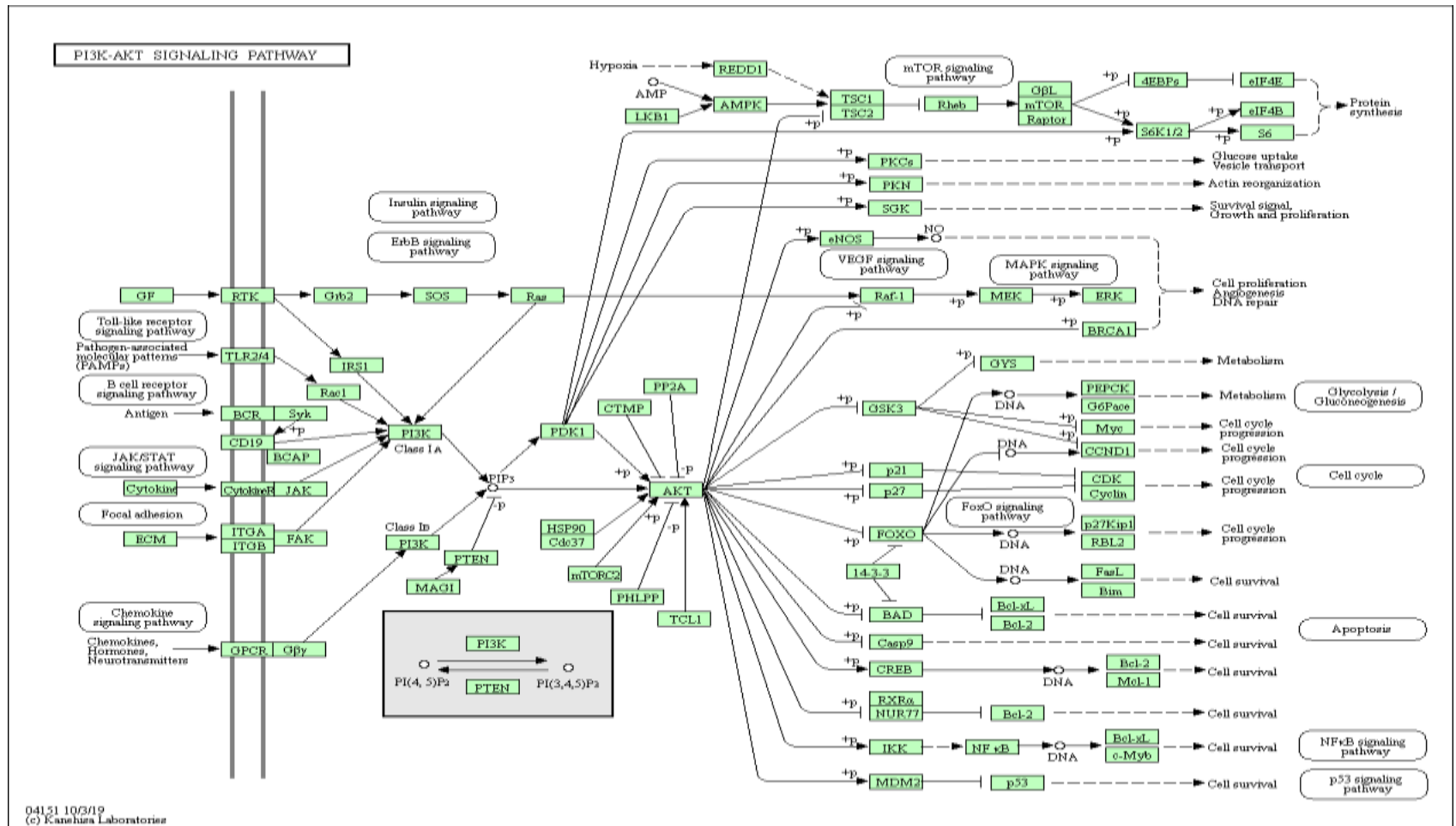


Figura 2 – Via de sinalização de PI3K-Akt (oas04151) enriquecida com genes presentes nas regiões delimitadas para área de olho de lombo em ovinos da raça Santa Inês.

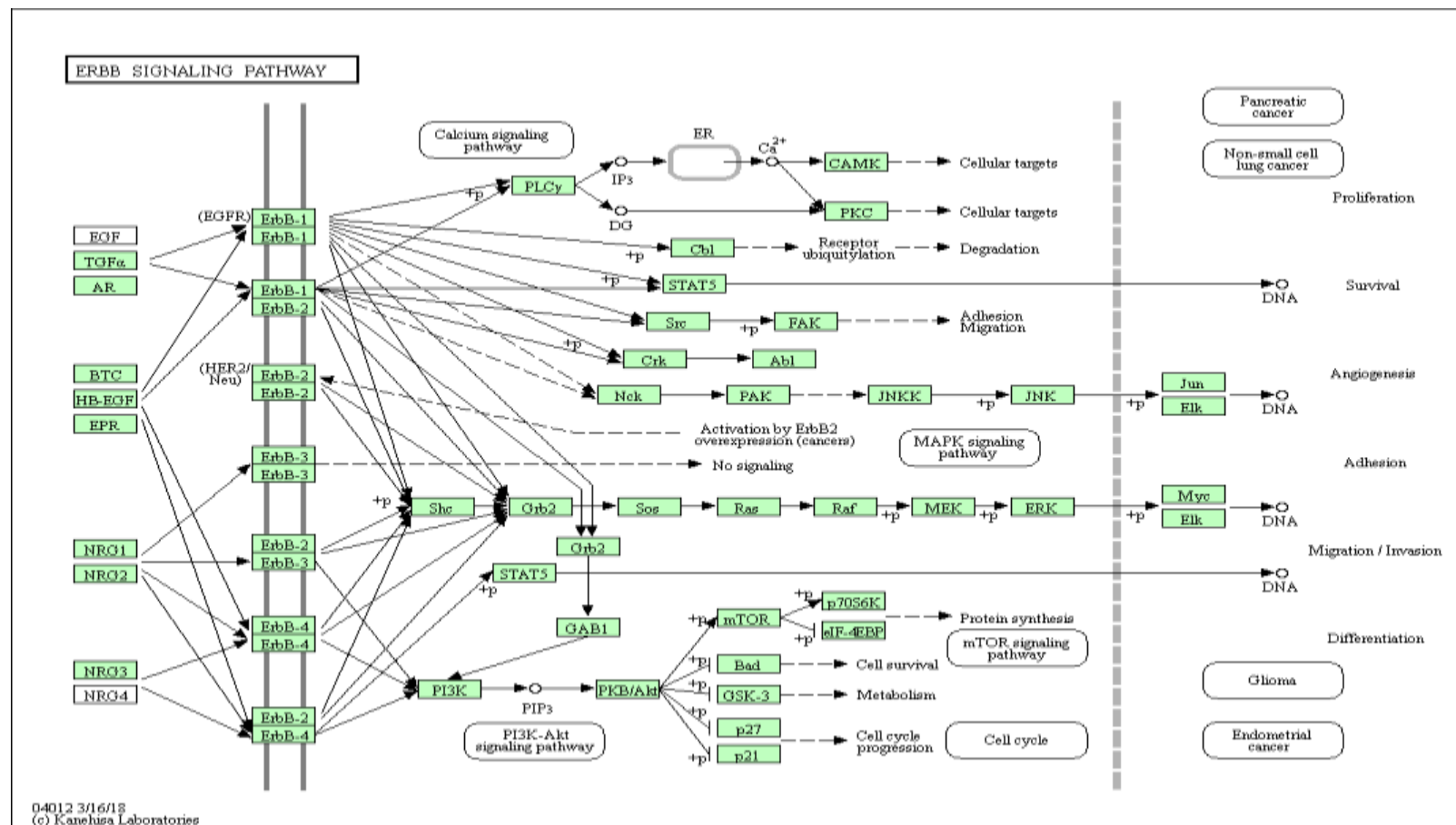


Figura 3 – Via de sinalização de ErbB (oas04012) enriquecida com genes presentes nas regiões delimitadas para área de olho de lombo em ovinos da raça Santa Inês.

3.2 Anotação funcional a partir do conjunto completo de genes

Com a utilização do conjunto gênico completo, ou seja, todos os genes presentes nas regiões relacionadas a AOL e EGS, foi possível realizar o enriquecimento de seis processos biológicos, dois componentes celulares e quatro funções moleculares (Tabela 4). O processo biológico mais significativamente enriquecido, neste caso, foi a homeostase de glicose celular (GO:0001678), que é essencial para a manutenção das atividades celulares e dos mecanismos envolvidos no desenvolvimento corporal, como mencionado anteriormente. Com uso do conjunto gênico completo, o processo GO:0001678 foi enriquecido a partir de genes presentes nas regiões relacionadas tanto a EGS quanto a AOL.

Tabela 4 - Análise de anotação de termos de ontologia gênica (GO) enriquecida (P<0,05) a partir de genes presentes nas regiões selecionadas no genoma ovino

Termo GO	Tipo	p-valor	Genes	³ FDR
GO:0001678~ homeostase de glicose celular	PB	2,6 ⁻³	<i>HK1¹, HKDC1¹, PIK3R1²</i>	3,70
GO:0048013~ via de sinalização de receptor de efrina	PB	8,2 ⁻³	<i>SS18¹, EFNA1², EFNA3²</i>	11,00
GO:0001227~atividade repressora da transcrição, ligação específica à sequência da região proximal promotora do núcleo da RNA polimerase II	FM	2,4 ⁻²	<i>AEBP2¹, ATF3², BATF3², NEUROG3¹</i>	24,00
GO:0060113~ diferenciação de células receptoras do ouvido interno	PB		<i>DFNA5¹, BMP4¹</i>	32,00
GO:0070062~exossomo extracelular	CC	2,8 ⁻²	<i>ADAM15², CDC42BPB², GNAQ¹, NPHS2², ANGPT1¹, CARMIL1², QPCT¹, ASH1L², GNA14¹, HEATR5B¹, VPS26A¹, ANGPTL1², EFNA1², GBA², GPLD1², MPP6¹,</i>	27,00

Continua

				<i>PMVK</i> ² , <i>KCNMA1</i> ² , <i>PSMA8</i> ¹ , <i>PKLR</i> ² , <i>SCAMP3</i> ² , <i>SPINK1</i> ² , <i>SPAST</i> ¹ , <i>TSPAN15</i> ¹ , <i>TOR3A</i> ²	
GO:0045121~ balsa de membrana	CC	3,1 ⁻²	<i>CD24</i> ² , <i>NPHS2</i> ² , <i>ANGPT1</i> ¹ , <i>HK1</i> ¹	30,00	
GO:0016477~migração celular	PB	3,3 ⁻²	<i>ABL2</i> ² , <i>CDC42BPB</i> ² , <i>CARMIL1</i> ² , <i>EFNA1</i> ²	38,00	
GO:0003714~atividade corepressora de transcrição	FM	3,8 ⁻²	<i>AEBP2</i> ¹ , <i>PBXIP1</i> ² , <i>GMNN</i> ² , <i>KCTD1</i> ¹	35,00	
GO:0010976~regulação positiva do desenvolvimento da projeção de neurônio	PB	3,8 ⁻²	<i>ABL2</i> ² , <i>CD24</i> ² , <i>FBXO38</i> ²	43,00	
GO:0004396~ atividade de hexoquinase	FM	4,1 ⁻²	<i>HK1</i> ¹ , <i>HKDC1</i> ¹	38,00	
GO:0070244~ regulação negativa do processo apoptótico de timócito	PB	4,2 ⁻²	<i>BMP4</i> ¹ , <i>EFNA1</i> ²	47,00	
GO:0000978~ ligação de DNA específica da sequência da região proximal do promotor do núcleo da RNA polimerase II	FM	4,3 ⁻²	<i>AEBP2</i> ¹ , <i>CEBPZ</i> ¹ , <i>ATF3</i> ² , <i>BATF3</i> ² , <i>NEUROG3</i> ¹ , <i>ZBTB7B</i> ²	39,00	

¹Gene presente em região relacionada a EGS; ²Gene presente em região relacionada a AOL; ³FDR:Taxa de falsa descoberta (%); PB: processo biológico; FM: função molecular; CC: componente celular.

Com exceção dos termos de ontologia GO:0060113, GO:0016477, GO:0010976 e GO:0004396, os demais termos foram enriquecidos com a participação de genes presentes nas regiões relacionadas às duas características em estudo. Este resultado é importante, pois corrobora alguns resultados mostrados em estudos com ovinos, que sugerem que há correlação genética entre desenvolvimento muscular e deposição de gordura. Por exemplo, em ovinos da raça Merino, Mortimer et al. (2017) observaram correlação genética de 0,80 entre espessura de gordura e profundidade do músculo *longissimus dorsi*, mensuradas por ultrassonografia. Em ovinos da raça Dorper, Kiya et al. (2019) encontraram correlação genética de 0,76 entre área de olho de lombo e

espessura de gordura subcutânea. Assim, é destacada a natureza pleiotrópica dos genes responsáveis pela expressão de musculabilidade e adiposidade em ovinos.

No total, 12 vias metabólicas foram significativamente enriquecidas ($P < 0,05$) com a utilização do conjunto gênico completo (Tabela 5). A principal via metabólica identificada neste caso foi a via de sinalização de insulina (Figura 4), que é essencial para a manutenção de diversos processos que estão associados ao desenvolvimento muscular e deposição de gordura em animais de produção (Qaid e Abdelrahman, 2016) e atividades celulares e dos mecanismos envolvidos no desenvolvimento corporal.

Tabela 5 - Vias KEGG significativamente enriquecidas (P<0,05) a partir do conjunto de genes presentes nas regiões selecionadas no genoma ovino

KEGG	p-valor	Genes	³ FDR
oas04015: Via de sinalização de Rap1	1,9 ⁻³	<i>GNAQ</i> ¹ , <i>ANGPT1</i> ¹ , <i>EFNA1</i> ² , <i>EFNA3</i> ² , <i>EFNA4</i> ² , <i>PIK3R1</i> ² , <i>PRKD3</i> ¹	2,20
oas04014: Via de sinalização de Ras	2,7 ⁻³	<i>ABL2</i> ² , <i>SHC1</i> ² , <i>ANGPT1</i> ¹ , <i>EFNA1</i> ² , <i>EFNA3</i> ² , <i>EFNA4</i> ² , <i>PIK3R1</i> ²	3,20
oas04930: Diabetes mellitus tipo II	3,2 ⁻³	<i>HK1</i> ¹ , <i>HKDC1</i> ¹ , <i>PIK3R1</i> ² , <i>PKLR</i> ²	3,80
oas04910: via de sinalização de insulina	1,1 ⁻²	<i>SHC1</i> ² , <i>HK1</i> ¹ , <i>HKDC1</i> ¹ , <i>PIK3R1</i> ² , <i>PKLR</i> ²	13,00
oas05222: câncer de pulmão de pequenas células	1,7 ⁻²	<i>CKS1B</i> ² , <i>TRAF3</i> ² , <i>LOC101107954</i> ² , <i>PIK3R1</i> ²	19,00
oas04151: via de sinalização PI3K-Akt	2,3 ⁻²	<i>ANGPT1</i> ¹ , <i>EFNA1</i> ² , <i>EFNA3</i> ² , <i>EFNA4</i> ² , <i>PIK3R1</i> ² , <i>PPP2R2B</i> ² , <i>THBS3</i> ²	24,00
oas04066: Via de sinalização de HIF-1	2,5 ⁻²	<i>ANGPT1</i> ¹ , <i>HK1</i> ¹ , <i>HKDC1</i> ¹ , <i>PIK3R1</i> ²	26,00
oas04973: digestão e absorção de carboidratos	2,6 ⁻²	<i>HK1</i> ¹ , <i>HKDC1</i> ¹ , <i>PIK3R1</i> ²	27,00
oas04020: via de sinalização do cálcio	2,8 ⁻²	<i>HTR4</i> ² , <i>HTR5A</i> ² , <i>GNA14</i> ¹ , <i>GNAQ</i> ¹ , <i>TACR2</i> ¹	29,00
oas00524: biosíntese de butirósina e neomicina	3,1 ⁻²	<i>HK1</i> ¹ , <i>HKDC1</i> ¹	32,00
oas05142: doença de Chagas (trypanosomíase americana)	3,5 ⁻²	<i>GNA14</i> ¹ , <i>GNAQ</i> ¹ , <i>PIK3R1</i> ² , <i>PPP2R2B</i> ²	35,00
oas00520: metabolismo de açúcar Amino e açúcar nucleotídeo	4,1 ⁻²	<i>LOC101106806</i> ² , <i>HK1</i> ¹ , <i>HKDC1</i> ¹	40,00

¹Gene presente em região relacionada a EGS; ²Gene presente em região relacionada a AOL; ³FDR:Taxa de falsa descoberta (%).

A insulina é o hormônio mais importante na regulação do metabolismo energético (Dimitriadis et al., 2011), pois é essencial para manter o equilíbrio entre a utilização e a produção de glicose, além de atuar como o principal agente anabólico fisiológico,

promovendo a síntese e armazenagem de glicose, lipídios e proteínas e inibindo sua degradação e liberação para a corrente sanguínea (Saltiel, 2016).

As vias KEGG oas04151 (PI3K-Akt) e oas04066 (HIF-1) haviam sido enriquecidas a partir dos genes presentes nas regiões relacionadas a AOL e EGS, respectivamente. Com a utilização do conjunto gênico completo, foi possível a obtenção do enriquecimento dessas vias metabólicas com genes presentes nas regiões relacionadas a ambas as características em estudo.

Apenas as vias oas05222 (câncer de pulmão de pequenas células) e oas00524 (biosíntese de butirósina e neomicina) apresentaram apenas genes localizados em regiões relacionadas a AOL ou EGS. Em todas as demais vias KEGG significativamente enriquecidas ($p < 0,05$) estão presentes genes que constam nas listas de AOL e EGS. Este resultado reforça o que foi ilustrado nas análises dos termos de ontologia gênica (Tabela 4), em relação a ação pleiotrópica de genes relacionados à expressão de musculosidade e deposição de gordura em ovinos.

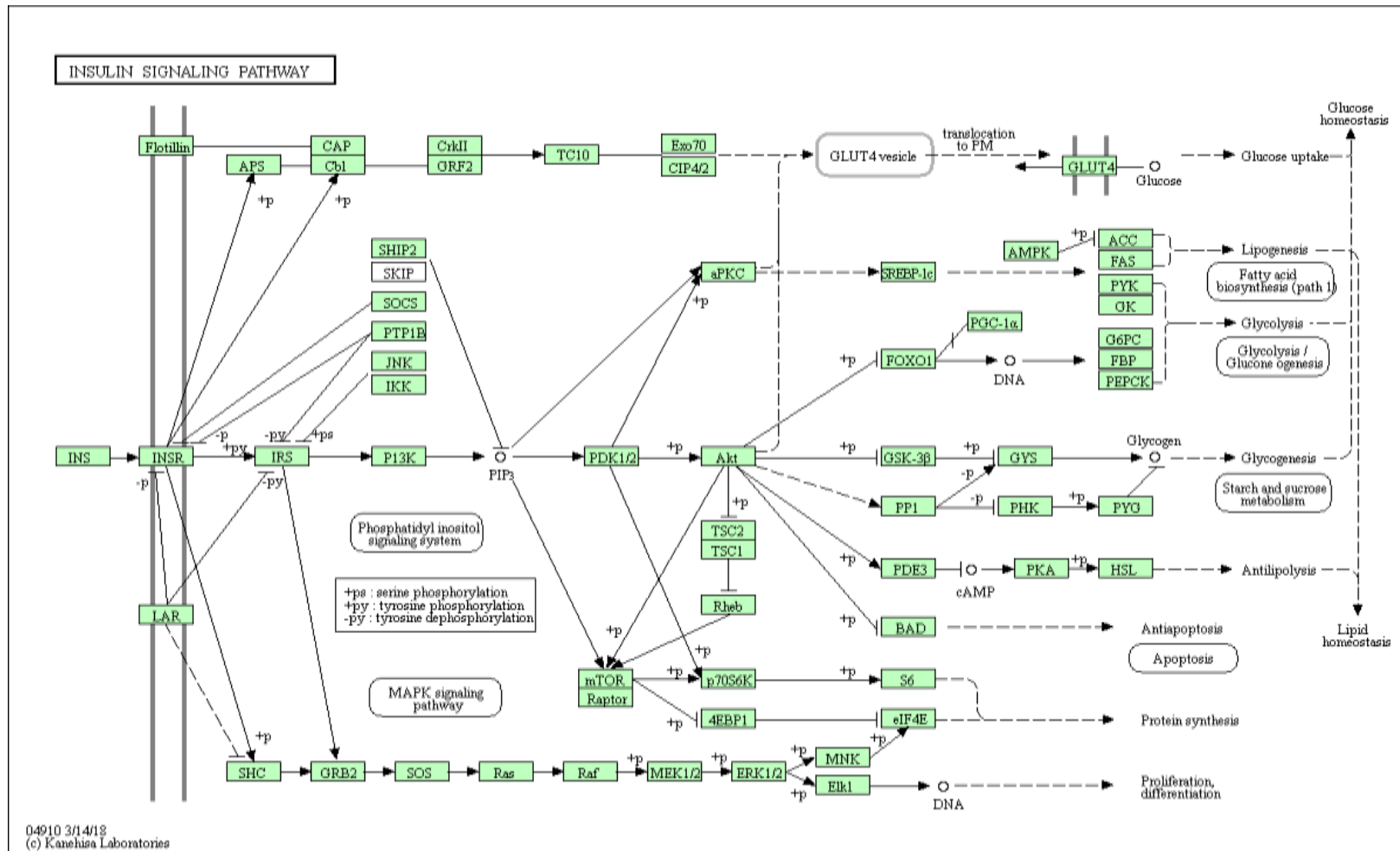


Figura 4 - Via de sinalização de insulina (oas04910) enriquecida com genes presentes nas regiões delimitadas para área de olho de lombo e espessura de gordura subcutânea em ovinos da raça Santa Inês.

3.3 Anotação funcional a partir de potenciais genes candidatos por característica

As análises com os conjuntos gênicos constituídos apenas por potenciais genes considerados como candidatos para AOL e EGS possibilitaram a identificação de alguns processos biológicos, funções moleculares e vias metabólicas semelhantes às aquelas identificadas com o conjunto de genes presentes nas janelas avaliadas para cada característica (Tabelas 6 e 7).

Tabela 6 - Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de potenciais genes candidatos presentes no genoma ovino, por característica

Fenótipo	Termo GO	Tipo	p-valor	Genes	^a FDR
AOL	GO:0016477~ migração celular	PB	8,1 ⁻³	<i>ABL2</i> , <i>CDC42BPB</i> , <i>CARMIL1</i>	9,30
	GO:0070062~exossomo extracelular	CC	3,0 ⁻²	<i>CDC42BPB</i> , <i>CARMIL1</i> , <i>GBA</i> , <i>GPLD1</i> , <i>PMVK</i> , <i>KCNMA1</i> , <i>PKLR</i> , <i>SCAMP3</i>	25,00
	GO:0032526~resposta a ácido retinoico	PB	3,3 ⁻²	<i>CLK2</i> , <i>SCAMP3</i>	33,00
	GO:0005524~ligação de ATP	FM	3,8 ⁻²	<i>ABL2</i> , <i>CLK2</i> , <i>CDC42BPB</i> , <i>MARK3</i> , <i>MYO16</i> , <i>PMVK</i>	29,00
	GO:0004674~Atividade de proteína serina/treonina quinase	FM	4,0 ⁻²	<i>CLK2</i> , <i>CDC42BPB</i> , <i>MARK3</i>	30,00
EGS	GO:0004396~atividade da hexoquinase	FM	4,3 ⁻³	<i>HK1</i> , <i>HKDC1</i>	3,40
	GO:0001678~homeostase de glicose celular	PB	5,9 ⁻³	<i>HK1</i> , <i>HKDC1</i>	5,70
	GO:0005536~ligação de glicose	FM	6,8 ⁻³	<i>HK1</i> , <i>HKDC1</i>	5,40
	GO:0006096~processo glicolítico	PB	1,4 ⁻²	<i>HK1</i> , <i>HKDC1</i>	13,00
	GO:0005623~Célula	CC	4,4 ⁻²	<i>HK1</i> , <i>HKDC1</i>	28,00

AOL:área de olho de lombo; EGS:espessura de gordura subcutânea; PB: processo biológico; FM: função molecular; CC: componente celular; ^aFDR:Taxa de falsa descoberta (%).

Os três potenciais genes candidatos (*ABL2*, *CDC42BPB* e *CARMIL1*) que possibilitaram o enriquecimento da migração celular também estão presentes na lista de quatro genes relacionados a este processo biológico, que foram enriquecidos quando foram utilizados todos os genes presentes em regiões relacionadas a AOL (Tabela 2).

Apesar do menor número de genes na lista de potenciais candidatos a EGS (n=12), os genes *HK1* e *HKDC1* possibilitaram a identificação de termos significativamente relacionados ao metabolismo da glicose. Isto se deve ao fato de que estes genes estão envolvidos em termos de ontologia em comum relacionados ao metabolismo de glicose, como processo glicolítico (GO:0006096) e homeostase de glicose celular (GO:0001678) e viabilizam a ligação de glicose (GO:0005536).

A via de sinalização ErbB foi significativamente enriquecida (Tabela 7) com os mesmos genes presentes nesta via quando foi utilizado o conjunto de todos os genes presentes nas regiões relacionadas a EGS. Devido à menor quantidade de genes na lista de potenciais candidatos a AOL, os valores de significância (p-valor) e FDR foram menores.

Tabela 7 - Vias KEGG significativamente enriquecidas ($P < 0,05$) para as características em estudo a partir da plataforma DAVID, a partir de potenciais genes candidatos presentes no genoma ovino, por característica.

Fenótipo	KEGG	p-valor	Genes	FDR
AOL	oas04012: via de sinalização ErbB	$9,0^{-3}$	<i>ABL2, SHC1, PIK3R1</i>	9,50
	oas04910: via de sinalização de insulina	$2,2^{-2}$	<i>SHC1, PIK3R1, PKLR</i>	21,00
EGS	oas00524: biosíntese de butirósina e neomicina	$2,6^{-3}$	<i>HK1, HKDC1</i>	2,50
	oas00500: Metabolismo de amido e sacarose	$1,6^{-2}$	<i>HK1, HKDC1</i>	15,00
	oas00052: Metabolismo de galactose	$1,6^{-2}$	<i>HK1, HKDC1</i>	15,00
	oas00051: Metabolismo de frutose e manose	$1,6^{-2}$	<i>HK1, HKDC1</i>	15,00
	oas04973~digestão e absorção de carboidrato	$2,1^{-2}$	<i>HK1, HKDC1</i>	19,00
	oas04930: Diabetes mellitus tipo II	$2,4^{-2}$	<i>HK1, HKDC1</i>	22,00
	oas00520: metabolismo de açúcar amino e açúcar nucleotídeo	$2,6^{-2}$	<i>HK1, HKDC1</i>	23,00
	oas05230: metabolismo central do carbono no câncer	$3,3^{-2}$	<i>HK1, HKDC1</i>	28,00
	oas00010: glicólise/gliconeogênese	$3,4^{-2}$	<i>HK1, HKDC1</i>	29,00

AOL: área de olho de lombo; EGS: espessura de gordura subcutânea; ^aFDR: Taxa de falsa descoberta (%).

Apenas os genes *HK1* e *HKDC1* possibilitaram o enriquecimento de vias KEGG, quando utilizado o conjunto de potenciais genes candidatos para EGS. Isto pode ser explicado devido à maior semelhança entre as funções moleculares e processos biológicos aos quais estes genes estão relacionados, em comparação aos demais genes presentes nesta lista. Do mesmo modo que observado para os termos de ontologia, a maior parte das vias metabólicas enriquecidas a partir de *HK1* e *HKDC1* é relacionada ao metabolismo de açúcares.

3.4 Anotação funcional a partir do conjunto completo de potenciais genes candidatos

No total, 13 termos de ontologia gênica foram enriquecidos com utilização do conjunto completo de potenciais genes candidatos (Tabela 8). Deste total, nove termos também foram identificados quando a análise foi realizada com os conjuntos de prováveis candidatos por característica. Nos termos atividade da hexoquinase (GO:0004396), migração celular (GO:0016477) e resposta a ácido retinoico (GO:0032526) estão presentes apenas potenciais genes candidatos relacionados a apenas uma das características.

Tabela 8 - Termos de ontologia gênica (GO) significativamente ($P < 0,05$) enriquecidos a partir de potenciais genes candidatos presentes no genoma ovino.

Termo GO	Tipo	p-valor	Genes	^a FDR
GO:0001678~ homeostase de glicose celular	PB	2,0 ⁻⁴	<i>HK1¹, HKDC1¹, PIK3R1²</i>	0,25
GO:0005524~ligação de ATP	FM	8,0 ⁻³	<i>ABL2², CLK2², CDC42BPB², EIF2AK2¹, HK1¹, HKDC1¹, MARK3², MYO16², PMVK²</i>	7,40
GO:0005623~ Célula	CC	1,2 ⁻²	<i>ABL2², HK1¹, HKDC1¹</i>	11,00
GO:0004674 ~ Atividade de proteína serina/treonina quinase	FM	1,3 ⁻²	<i>CLK2¹, CDC42BPB², EIF2AK2¹, MARK3²</i>	12,00
GO:0004396~ atividade da hexoquinase	FM	1,3 ⁻²	<i>HK1¹, HKDC1¹</i>	12,00
GO:0016477~ migração celular	PB	1,6 ⁻²	<i>ABL2², CDC42BPB², CARMIL1²</i>	19,00
GO:0005536~ ligação de glicose	FM	2,1 ⁻²	<i>HK1¹, HKDC1¹</i>	19,00
GO:0001508~ potencial de ação	PB	2,2 ⁻²	<i>GNAQ¹, GLRA1²</i>	24,00

Continua

GO:0070062~exossomo extracelular	CC	2,9 ⁻²	<i>CDC42BPB</i> ² , <i>GNAQ</i> ¹ , <i>CARMIL1</i> ² , <i>GBA</i> ² , <i>QPCT</i> ¹ , <i>GPLDI</i> ² , <i>PMVK</i> ² , <i>KCNMA1</i> ² , <i>PKLR</i> ² , <i>SCAMP3</i> ²	25,00
GO:0034198~Resposta celular à privação de aminoácido	PB	3,1 ⁻²	<i>ATF3</i> ² , <i>EIF2AK2</i> ¹	33,00
GO:0005765~Membrana lisossômica	CC	3,9 ⁻²	<i>GNAQ</i> ¹ , <i>GBA</i> ² , <i>GPLDI</i> ²	32,00
GO:0032526~resposta a ácido retinoico	PB	4,8 ⁻²	<i>CLK2</i> ² , <i>SCAMP3</i> ²	46,00
GO:0015485~ligação de colesterol	FM	4,9 ⁻²	<i>OSBPL3</i> ¹ , <i>SOAT1</i> ²	39,00

¹Gene presente em região relacionada a EGS; ²Gene presente em região relacionada a AOL; ³FDR:Taxa de falsa descoberta (%). AOL:área de olho de lombo; EGS:espessura de gordura subcutânea; PB: processo biológico; FM: função molecular; CC: componente celular.

Do total de 14 vias KEGG significativamente enriquecidas com uso do conjunto de potenciais genes candidatos relacionados a AOL e EGS (Tabela 9), foram identificadas oito vias metabólicas diferentes daquelas que já haviam sido identificadas com uso dos conjuntos gênicos individuais por características, separadamente.

Tabela 9 - Vias KEGG significativamente enriquecidas ($P < 0,05$) a partir de potenciais genes candidatos para as características em estudo.

KEGG	p-valor	Genes	³ FDR
oas04930: Diabetes mellitus tipo II	1,6 ⁻⁴	<i>HK1¹, HKDC1¹, PIK3R1², PKLR²</i>	1,90
oas04910: via de sinalização de insulina	2,5 ⁻⁴	<i>SHC1², HK1¹, HKDC1¹, PIK3R1², PKLR²</i>	2,80
oas04973~digestão e absorção de carboidrato	3,8 ⁻³	<i>HK1¹, HKDC1¹, PIK3R1²</i>	4,30
oas05230: metabolismo central do carbono no câncer	9,5 ⁻³	<i>HK1¹, HKDC1¹, PIK3R1²</i>	11,00
oas00010: glicólise/gliconeogênese	1,0 ⁻²	<i>HK1¹, HKDC1¹, PKLR²</i>	11,00
oas00524: biosíntese de butirósina e neomicina	1,2 ⁻²	<i>HK1¹, HKDC1¹</i>	13,00
oas04911: secreção de insulina	1,6 ⁻²	<i>GNAQ¹, KCNMA1², KCNN3²</i>	17,00
oas04012: via de sinalização ErbB	1,7 ⁻²	<i>ABL2², SHC1², PIK3R1²</i>	18,00
oas04915: via de sinalização de estrógeno	2,2 ⁻²	<i>GNAQ¹, SHC1², PIK3R1²</i>	22,00
oas04066: via de sinalização HIF-1	2,2 ⁻²	<i>HK1¹, HKDC1¹, PIK3R1²</i>	22,00
oas01200: metabolismo de carbono	2,9 ⁻²	<i>HK1¹, HKDC1¹, PKLR²</i>	29,00
oas05169: infecção pelo vírus Epstein-Barr	3,3 ⁻²	<i>TRAF3², EIF2AK2¹, PIK3R1²</i>	32,00
oas05160: hepatite C	3,8 ⁻²	<i>TRAF3², EIF2AK2¹, PIK3R1²</i>	36,00
oas04950: diabetes com início da maturidade dos jovens	4,8 ⁻²	<i>NEUROG3¹, PKLR²</i>	43,00

¹Gene presente em região relacionada a EGS; ²Gene presente em região relacionada a AOL; ³FDR: Taxa de falsa descoberta (%).

4. Conclusões

A utilização de genes presentes em regiões relacionadas aos fenótipos área de olho de lombo e espessura de gordura subcutânea possibilitou o enriquecimento de importantes termos de ontologia e vias metabólicas relacionados aos metabolismos

proteicos, glicolítico e lipídico, que influenciam diretamente a musculosidade e deposição de gordura em ovinos.

Algumas conclusões importantes podem ser destacadas:

- Os termos de ontologias mais significativamente enriquecido a partir dos genes relacionados a AOL, estão presentes genes que participam diretamente da migração celular e de processos relacionados a este. Com a utilização do conjunto gênico relacionado a EGS, os genes *HK1* (hexoquinase 1) e *HKDC1* (hexoquinase contendo domínio 1) participaram de três termos de ontologia diretamente relacionados ao metabolismo de glicose.
- As análises de vias KEGG com os genes presentes nas regiões delimitadas para AOL possibilitaram a identificação das vias de sinalização Ras, PI3K-Akt e ErbB, que atuam diretamente no desenvolvimento e homeostase muscular.
- Com o conjunto de genes presentes nas regiões delimitadas para EGS, foi possível a identificação da via de sinalização HIF-1, que atua na regulação do metabolismo lipídico.
- Com todos os genes presentes nas regiões relacionadas a AOL e EGS, foi possível realizar o enriquecimento de seis processos biológicos, dois componentes celulares e quatro funções moleculares. O processo biológico mais significativamente enriquecido foi a homeostase de glicose celular (GO:0001678), que é essencial para a manutenção das atividades celulares e dos mecanismos envolvidos no desenvolvimento corporal, como mencionado anteriormente.
- Com exceção dos termos de ontologia GO:0060113, GO:0016477, GO:0010976 e GO:0004396, os demais termos foram enriquecidos com a participação de genes presentes nas regiões relacionadas às duas características em estudo, o que faz este resultado corroborar com alguns estudos com ovinos, que sugerem que há correlação genética entre desenvolvimento muscular e deposição de gordura.
- Com a utilização do conjunto gênico completo, ou seja, todos os genes presentes nas regiões relacionadas a AOL e EGS, foi possível realizar o enriquecimento de seis processos biológicos, dois componentes celulares e quatro funções moleculares e o processo biológico mais significativamente enriquecido, neste caso, foi a homeostase de glicose celular (GO:0001678), que é essencial para a

manutenção das atividades celulares e dos mecanismos envolvidos no desenvolvimento corporal, como mencionado anteriormente.

- 12 vias metabólicas foram significativamente enriquecidas ($P < 0,05$) com a utilização do conjunto gênico completo, no qual a principal via metabólica identificada foi a via de sinalização de insulina, essencial para a manutenção de diversos processos que estão associados ao desenvolvimento muscular e deposição de gordura em animais de produção e atividades celulares e dos mecanismos envolvidos no desenvolvimento corporal.

Referências

- Andrechek, E.R., Hardy, W.R., Girgis-Gabardo, A.A., Perry, R.L., Butler, R., Graham, F.L., Muller, W.J., 2002. ErbB2 is required for muscle spindle and myoblast cell survival. *Mol. Cell. Biol.* 22, 4714–4722. doi:10.1128/mcb.22.13.4714-4722.2002
- Ashburner et al., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25-29. <https://doi.org/10.1038/75556>.
- Delafontaine, P., Yoshida, T., 2016. The renin-angiotensin system and the biology of skeletal muscle: mechanisms of muscle wasting in chronic disease states. *Trans. Am. Clin. Climatol. Assoc.* 127, 245–258.
- Dimitriadis, G., Mitrou, P., Lambadiari, V., Maratou, E., Raptis, S. A., 2011. Insulin effects in muscle and adipose tissue. *Diabetes Research and Clinical Practice*, p. 93.
- Fujita, S., Dreyer, H.C., Drummond, M.J., Glynn, E.L., Cadenas, J.G., Yoshizawa, F., Rasmussen, B.B., 2007. Nutrient signalling in the regulation of human muscle protein synthesis. *J. Physiol.* 582, 813–823. doi:10.1113/jphysiol.2007.134593
- Glass, D.J., 2010. PI3 kinase regulation of skeletal muscle hypertrophy and atrophy. *Curr. Top. Microbiol. Immunol.* 346, p. 267–278, 2010. doi: 10.1007/82_2010_78.

- Griffin, C. A., Apponi, L. H., Long, K. K., & Pavlath, G. K. 2010. Chemokine expression and control of muscle cell migration during myogenesis. *Journal of cell science*, 123(Pt 18), 3052–3060. doi:10.1242/jcs.066241
- Hay, E.H., Roberts, A., 2018. Genome-wide association study for carcass traits in a composite beef cattle breed. *Livest. Sci.* 213, 35-43. <https://doi.org/10.1016/j.livsci.2018.04.018>.
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1-13. <https://doi.org/10.1093/nar/gkn923>.
- Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 4, 44-57. <https://doi.org/10.1038/nprot.2008.211>.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28, 27-30.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Yamanishi, Y., 2008. KEGG for linking genomes to life and the environment. *Nucleic acids research*, 36(Database issue), D480–D484. doi:10.1093/nar/gkm882
- Kiya, C.K., Pedrosa, V.B., Muniz, K.F.A., Gusmão, A.L., Pinto, L.F.B., 2019. Estimates of the genetic parameters of a Dorper flock in Brazil. *Small Rumin. Res.* 171, 57-62. <https://doi.org/10.1016/j.smallrumres.2018.12.007>.
- Mortimer, S.I., Hatcher, S., Fogarty, N.M., van der Werf, J.H.J., Brown, D.J., Swan, A.A., Greeff, J.C., Refshauge, G., Edwards, J.E.H., Gaunt, G.M., 2017. Genetic parameters for wool traits, live weight, and ultrasound carcass traits in Merino sheep. *J. Anim. Sci.* 95, 1879-1891. <https://doi.org/10.2527/jas.2016.1234>.

- Mylonis, I., Simos, G., Paraskeva, E., 2019. Hypoxia-Inducible Factors and the Regulation of Lipid Metabolism. *Cells* 8, 3.
- Qaid, M.M.; Abdelrahman, M. M., 2016. Role of insulin and other related hormones in energy metabolism—A review. *Cogent Food & Agricult.* 2, 1, 1267691.
- Saltiel, A.R., 2016. Insulin signaling in the control of glucose and lipid homeostasis. *Handb. Exp. Pharmacol.* 233: 51–71. doi: 10.1007/164_2015_14.
- Sena, L. S., 2019. Estudo genômico aplicado ao melhoramento genético de ovinos Santa Inês para características de carcaça, Teresina: Centro de Ciências Agrárias, Universidade Federal do Piauí, 125 p. Tese de Doutorado.
- Silva, L. R. G. 2019. Genômica comparativa para identificação de genes candidatos à resistência a endoparasitas gastrintestinais na espécie *Ovis aries*. Dissertação (Mestrado) — Departamento de Zootecnia, Universidade Federal do Piauí.
- Silva, R.M.O., Stafuzza, N.B., Fragomeni, B.O, Camargo, G.M.F., Ceacero, T.M, Cyrillo, N.S.G.J., Albuquerque, L.G., 2017. Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. *PloS one*, 12, 1, e0169860. doi:10.1371/journal.pone.0169860.
- Silva, R.P., Berton, M.P, Grigoletto, L, Carvalho, F.E., Silva, R.M.O., Peripolli, E., Castro, L.M., Ferraz, J.B.S., Eler, J.P., Lôbo, R.B., Baldi, F., 2019. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. *Journal of Animal Breeding and Genetics*, 136, 118-133.
- Webber, C., 2011. Functional enrichment analysis with structural variants: Pitfalls and Strategies. *Cytogenet Genome Res.* 135, 277–285.
- Wilson, J.E., 2003. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J. Exp. Biol.* 206(Pt 12): 2049–2057.

Zhang, H., Wang, Z., Wang, S., Li., 2012. Progress of genome wide association study in domestic animals. *J. Anim. Sci. Biotechnol.* 3, 26. <https://doi.org/10.1186/2049-1891-3-26>.

CONSIDERAÇÕES FINAIS

A utilização de informação de SNPs associados com características de interesse econômico, como o rendimento de carcaça e a qualidade de carne, aliados com estudos de comparação de genomas de espécies filogeneticamente relacionadas, representam excelentes ferramentas para a identificação de genes candidatos importantes.

A quantidade de informações utilizada no presente estudo foi relativamente pequena, comparada a outros trabalhos de genômica comparativa, mas ainda assim, esses resultados mostraram-se satisfatórios e servirão como aporte para mais estudos relacionados aos fatores genéticos que influenciam as características associadas à carcaça em pequenos ruminantes.

Nesta pesquisa, importantes genes e regiões relacionados com os processos de desenvolvimento muscular e deposição de gordura foram evidenciados: 38 genes candidatos imediatos, dos quais 26 são relacionadas à musculabilidade e 12 à deposição de gordura subcutânea; foram identificadas e propostas como potenciais candidatas três genes para AOL (ENSOARG00000005535, LOC101104530 e ENSOARG00000005104) e duas regiões ainda não descritas para EGS (CR2:127924874-127925647 e CR2:127924874-127925647). Isto contribui para o entendimento sobre a função e expressão desses genes em outras espécies diretamente relacionadas filogeneticamente, através da genômica comparativa. Entretanto, para ser confirmada a candidatura de um determinado gene, estudos mais detalhados da estrutura e expressão gênica são necessários.

Estudos mais aprofundados sobre a expressão desses genes poderão confirmar a candidatura dos mesmos e recomendar seu uso nos critérios de seleção em programas de melhoramento genético de ovinos de corte. Essas informações poderão dar suporte a estudos genômicos em ovinos, assim como, à identificação e seleção de animais com maior rendimento e melhor qualidade de carcaça.

APÊNDICES

Tabela Suplementar 1 - Regiões genômicas relacionadas a área de olho de lombo (AOL) em ovinos Santa Inês* e regiões extrapoladas para análise no presente estudo.

*Sena (2019)					Presente estudo				
OAR	Início (pb)	Fim (pb)	Tamanho (pb)	Tamanho (Mb)	¹ Aumento (pb)	² Início (pb)	³ Fim (pb)	Tamanho em PB	Tamanho em Mb
1	103.626.285	104.115.107	488.822	0,4888	113.188	103.513.097	104.228.295	715.198	0,7152
1	241.611.252	242.256.369	645.117	0,6451	281.018	241.330.234	242.537.387	1.207.153	1,2072
2	247.483.890	247.867.842	383.952	0,3839	68.354	247.415.536	247.936.196	520.660	0,5207
4	116.002.333	116.343.298	340.965	0,3410	65.077	115.937.256	116.408.375	471.119	0,4711
4	70.440.497	70.888.676	448.179	0,4482	117.945	70.322.552	71.006.621	684.069	0,6841
5	55.647.041	57.012.951	1.365.910	1,3659	390.037	55.257.004	57.402.988	2.145.984	2,1460
5	60.500.082	60.910.900	410.818	0,4108	55.500	60.444.582	60.966.400	521.818	0,5218
6	74.422.550	74.961.416	538.866	0,5389	100.420	74.322.130	75.061.836	739.706	0,7397
8	29.854.902	30.172.605	317.703	0,3177	73.016	29.781.886	30.245.621	463.735	0,4637
8	45.222.581	45.685.221	462.640	0,4626	114.131	45.108.450	45.799.352	690.902	0,6909
10	83.151.438	83.451.800	300.362	0,3004	58.914	83.092.524	83.510.714	418.190	0,4182
12	58.410.632	59.021.348	610.716	0,6107	86.815	58.323.817	59.108.163	784.346	0,7843
12	68.910.684	69.196.281	285.597	0,2856	57.839	68.852.845	69.254.120	401.275	0,4013
16	11.417.839	11.788.270	370.431	0,3704	62.173	11.355.666	11.850.443	494.777	0,4948
18	66.371.408	67.250.289	878.881	0,8789	180.778	66.190.630	67.431.067	1.240.437	1,2404

Continua

Continuação da Tabela Suplementar 1

20	31.443.849	31.983.671	539.822	0,5398	173.515	31.270.334	32.157.186	886.852	0,8869
23	9.073.407	9.598.680	525.273	0,5253	115.228	8.958.179	9.713.908	755.729	0,7557
25	32.244.874	32.531.667	286.793	0,2868	53.950	32.190.924	32.585.617	394.693	0,3947

OAR:cromossomo autossômico da espécie *Ovis aries*; pb: pares de base; Mb: mega pares de base; ¹Tamanho da **região adicionada a cada lado** das janelas associadas a AOL identificadas por Sena (2019), com base na maior distância entre pares de SNPs adjacentes presentes nas janelas mostradas no estudo referência; ²Início da região genômica utilizada no presente estudo após a adição da extrapolação; ³Final da região genômica após adição da extrapolação.

Tabela Suplementar 2 – Regiões genômicas relacionadas a espessura de gordura subcutânea (EGS) em ovinos Santa Inês* e regiões extrapoladas para análise no presente estudo.

*Sena (2019)					Presente estudo				
OAR	Início (pb)	Fim (pb)	Tamanho (pb)	Tamanho (Mb)	¹ Extrapolação (pb)	² Início (pb)	³ Fim (pb)	Tamanho em PB	Tamanho em Mb
1	81.223.040	81.762.074	539.034	0,5390	151.326	81.071.714	81.913.400	841.686	0,8417
2	54.304.983	54.859.065	554.082	0,5541	116.515	54.188.468	54.975.580	787.112	0,7871
2	207.523.564	207.861.282	337.718	0,3377	52.569	207.470.995	207.913.851	442.856	0,4429
2	127.994.859	128.448.469	453.610	0,4536	89.089	127.905.770	128.537.558	631.788	0,6318
2	58.377.511	58.805.835	428.324	0,4283	55.101	58.322.410	58.860.936	538.526	0,5385
3	91.413.691	91.908.160	494.469	0,4945	113.206	91.300.485	92.021.366	720.881	0,7209
3	195.104.765	195.620.265	515.500	0,5155	110.371	194.994.394	195.730.636	736.242	0,7362
3	86.273.231	86.864.592	591.361	0,5914	122.597	86.150.634	86.987.189	836.555	0,8366
4	70.777.595	71.117.749	340.154	0,3402	56.848	70.720.747	71.174.597	453.850	0,4539
7	74.829.782	75.197.725	367.943	0,3679	55.096	74.774.686	75.252.821	478.135	0,4781
7	62.785.817	63.631.984	846.167	0,8462	184.132	62.601.685	63.816.116	1.214.431	1,2144
8	37.883.088	38.263.778	380.690	0,3807	62.407	37.820.681	38.326.185	505.504	0,5055
8	65.291.460	65.744.396	452.936	0,4529	103.948	65.187.512	65.848.344	660.832	0,6608
9	69.276.855	69.634.999	358.144	0,3581	76.939	69.199.916	69.711.938	512.022	0,5120
9	20.247.556	20.592.464	344.908	0,3449	60.309	20.187.247	20.652.773	465.526	0,4655

Continua

Continuação da Tabela Suplementar 2

12	16.132.429	16.608.566	476.137	0,4761	172.736	15.959.693	16.781.302	821.609	0,8216
16	58.712.294	58.991.667	279.373	0,2794	55.787	58.656.507	59.047.454	390.947	0,3909
17	42.416.982	42.967.825	550.843	0,5508	78.642	42.338.340	43.046.467	708.127	0,7081
23	30.815.656	31.304.735	489.079	0,4891	93.153	30.722.503	31.397.888	675.385	0,6754
25	25.257.437	25.716.923	459.486	0,4595	94.147	25.163.290	25.811.070	647.780	0,6478

OAR:cromossomo autossômico da espécie *Ovis aries*; pb: pares de base; Mb: mega pares de base; ¹Tamanho da **região adicionada a cada lado** das janelas associadas a EGS identificadas por Sena (2019), com base na maior distância entre pares de SNPs adjacentes presentes nas janelas mostradas no estudo referência; ²Início da região genômica utilizada no presente estudo após a adição da extrapolação; ³Final da região genômica após adição da extrapolação.

Tabela Suplementar 3. Regiões genômicas relacionadas à área de olho de lombo mensurada por ultrassonografia em tempo real em ovinos Santa Inês.

Cromossomo: Região ^a	Genes
1:103513097-104228295	<i>KCNN3, PMVK, PBXIP1, SHC1, DCST2, KRTCAP2, THBS3, GBA, FAM189B, SCAMP3, CLK2, PKLR, ASH1L, DPM3, CKS1B, FLAD1, ZBTB7B, DCST1, ADAM15, EFNA1, SLC50A1, TRIM46, MTX1, HCN3, LENE, EFNA4, EFNA3, MUC1, ENSOARG00000004341, ENSOARG00000023246</i>
1:241330234-242537387	<i>ENSOARG00000025624</i>
2:247415536-247936196	<i>ACTL8, ARHGEF10L, ENSOARG00000025866, ENSOARG00000025867, ENSOARG00000025868, ENSOARG00000025869</i>
4:115937256-116408375	<i>HTR5A, ENSOARG00000006061, ENSOARG00000006181, PAXIP1, ENSOARG00000006565</i>
4:70322552-71006621	<i>NPVF, C7orf31, WSB1, OSBPL3, ENSOARG00000024588, ENSOARG00000024707, LOC101107954</i>
5:55257004-57402988	<i>DPYSL3, JAKMIP2, SPINK1, C5H5orf46, STK32A, SCGB3A2, SPINK5, SPINK6, PPP2R2B, SPINK7, FBXO38, HTR4, ENSOARG00000003500, ENSOARG00000004408, ENSOARG00000004453, LOC101116828, ENSOARG00000004524, ENSOARG00000004867, ENSOARG00000014332, ENSOARG00000021234</i>
5:60444582-60966400	<i>GLRA1, G3BP1, ATP5MPL, ENSOARG00000014409, C6H4orf19</i>
6:74322130-75061836	-
8:29781886-30245621	<i>PDSS2, BEND3, C8H6orf203, ENSOARG00000011078, CD24</i>
8:45108450-45799352	<i>ENSOARG00000012275, ENSOARG00000027023, ENSOARG00000027024</i>
10:83092524-83510714	<i>MYO16</i>
12:58323817-59108163	<i>ABL2, NPHS2, FAM20B, TOR3A, SOAT1, AXDND1, TDRD5, RALGPS2, ANGPTL1, FAM163A</i>
12:68852845-69254120	<i>ATF3, NSL1, BATF3, SPATA45, TATDN3, ENSOARG00000005299, ENSOARG00000025465</i>
16:11355666-11850443	<i>PIK3R1</i>
18:66190630-67431067	<i>CDC42BPB, XRCC3, PPP1R13B, ATP5MPL, RD3L, MARK3, TRMT61A, BAG5, TDRD9, TRAF3, APOPT1, AMN, ENSOARG00000005499, TNFAIP2, ENSOARG00000005535, EIF5, ENSOARG00000005800, KLC1, ENSOARG00000006331, ENSOARG00000006510, ENSOARG00000007159, ENSOARG00000007166, ENSOARG00000007170, ENSOARG00000007197, ENSOARG00000007222, LOC101104530, ENSOARG00000026466,</i>

Continua

Continuação da Tabela Suplementar 3

	<i>ENSOARG00000026467, ENSOARG00000026468, ENSOARG00000026469, ENSOARG00000026470, ENSOARG00000026471</i>
20:31270334-32157186	<i>CARMIL1, ARMH2, GMNN, KIAA0319, TDP2, ALDH5A1, GPLD1, LOC101106806, ENSOARG00000005945, ENSOARG00000005997, ENSOARG00000006103, ENSOARG00000005104, ENSOARG00000005436, FAM65B</i>
23:8958179-9713908	<i>DSEL, ENSOARG00000026154</i>
25:32190924-32585617	<i>KCNMA1, ENSOARG00000009394</i>

^aPosição em pares de bases (assembly Oar_v4.0); ^bPorcentagem de variância genética explicada por janelas de 10 SNPs adjacentes.

Tabela Suplementar 4. Regiões genômicas relacionadas à espessura de gordura subcutânea mensurada por ultrassonografia em tempo real em ovinos Santa Inês.

Cromossomo: Região ^a	Genes
1:81071714-81913400	<i>ENSOARG00000011145, LOC101119348)</i>
2:54188468-54975580	<i>TLE1, LOC101110483</i>
2:207470995-207913851	<i>ADAM23, DYTN, FASTKD2, MDH1B, ENSOARG00000018828, ENSOARG00000025842, ENSOARG00000025843</i>
2:127905770-128537558	<i>ENSOARG00000016891</i>
2:58322410-58860936	<i>GNA14, GNAQ,</i>
3:91300485-92021366	<i>BIRC6, NLRC4, SLC30A6, SPAST, YIPF4, MEMO1, ENSOARG00000024388</i>
3:194994394-195730636	<i>AEBP2, PLEKHA5, ENSOARG00000020236, ENSOARG00000024921</i>
3:86150634-86987189	<i>CEBPZ, EIF2AK2, NDUFAF7, PRKD3, QPCT, SULT6B1, GPATCH11, HEATR5B, ENSOARG00000025996, ENSOARG00000025997</i>
4:70720747-71174597	<i>OSBPL3, WSB1, DFNA5, MPP6, ENSOARG00000010705, ENSOARG00000024588, ENSOARG00000024707</i>
7:74774686-75252821	<i>ENSOARG00000024177, ENSOARG00000026717</i>
7:62601685-63816116	<i>BMP4, DDHD1</i>
8:37820681-38326185	<i>FBXL4, LOC101112574</i>
8:65187512-65848344	<i>ENSOARG00000024104, ENSOARG00000027037</i>
9:69199916-69711938	<i>ANGPT1,</i>
9:20187247-20652773	<i>ZFAT</i>
12:15959693-16781302	<i>ENSOARG00000024013, ENSOARG00000025433</i>
16:58656507-59047454	<i>TRIO</i>
17:42338340-43046467	<i>CTSO, ENSOARG00000025674</i>
23:30722503-31397888	<i>PSMA8, SS18, TAF4B, KCTD1, ENSOARG00000026180, ENSOARG00000026181</i>
25:25163290-25811070	<i>HK1, HKDC1, NEUROG3, SRGN, SUPV3L1, TACR2, TSPAN15, VPS26A, C25H10orf35, ENSOARG00000023294, ENSOARG00000024696</i>

^aPosição em pares de bases (assembly Oar_v4.0); ^bPorcentagem de variância genética explicada por janelas de 10 SNPs adjacentes.

Tabela Suplementar 5 – Genes candidatos imediatos presentes nas regiões relacionadas a AOL na espécie *Ovis aries* e funções biológicas mais relevantes.

Gene	Função disponível no Biomart (Uniprot)
<i>KCNN3</i> (Membro 3 da subfamília N do canal de potássio ativado por cálcio)	Transporte de íons de potássio; transporte de íons de potássio pela transmembrana; atividade do canal de potássio ativado por cálcio
<i>PMVK</i> (Quinase fosfomevalonato)	Resposta ao colesterol; processo biossintético de esteroide; processo biossintético de colesterol; processo metabólico lipídico; processo biossintético esteroide; atividade de fosfomevalonato quinase
<i>SHC1</i> (SHC adaptador de proteína 1)	Transdução de sinal; ligação ao receptor do fator de crescimento semelhante à insulina; ligação ao receptor de insulina; resposta celular ao estímulo do fator de crescimento; adesão célula-célula; reorganização do citoesqueleto de actina; receptor de ligação tirosina quinase; regulação da proliferação da população celular
<i>TRIM46</i> (Tripartite motif containing 46)	Ligação de íons de zinco; ligação de íons metálicos; organização do citoesqueleto de microtúbulos.
<i>GBA</i> (Glicosilceramidase beta)	Autofagia; processo metabólico do colesterol; regulação do processo metabólico das proteínas celulares; regulação positiva da lipidação das proteínas.
<i>SCAMP3</i> (Proteína 3 da membrana transportadora secretora)	Transporte de proteínas; ligação à ubiquitina proteína ligase.
<i>CLK2</i> (CDC semelhante a quinase 2)	Regulação negativa da gliconeogênese; atividade da proteína tirosina quinase; atividade da proteína serina/treonina quinase; ligação de proteínas idênticas; atividade da proteína quinase
<i>HCN3</i> (Canal de potássio 3 bloqueado por nucleotídeo cíclico ativado por hiperpolarização)	Transporte transmembranar de íons de potássio; transporte transmembranar de íons de sódio; transporte de íons potássio

Continua

Continuação da Tabela Suplementar 5

<i>PKLR</i> (Piruvato-quinase L/R)	Ligação de íons magnésio; processo glicolítico; atividade de quinase; fosforilação; ligação de íons de potássio; atividade de piruvato cinase
<i>DPM3</i> (subunidade 3 de dolicolil-fosfato manossiltransferase, reguladora)	Retículo endoplasmático; glicosilação de proteínas; regulação de estabilidade protéica
<i>MYO16</i> (Miosina XVI)	Atividade motora; complexo de miosina; ligação de actina; sinalização de fosfatidilinositol 3-quinase
<i>ATF3</i> (Fator de ativação da transcrição 3)	Diferenciação de células do músculo esquelético; gliconeogênese
<i>ABL2</i> (Proto-oncogene 2 ABL, tirosina-quinase não receptora)	Atividade da proteína tirosina quinase; ligação de íons magnésio; atividade da quinase; fosforilação
<i>SOAT1</i> (Esterol O-aciltransferase 1)	Atividade da O-aciltransferase de colesterol; ligação ao colesterol; processo metabólico do colesterol; retículo endoplasmático.
<i>PIK3R1</i> (Subunidade reguladora 1 da fosfoinositida-3-quinase)	Transdução de sinal; resposta celular ao estímulo à insulina; ligação à insulina; ligação ao receptor do fator de crescimento semelhante à insulina.
<i>TRAF3</i> (Fator 3 associado ao receptor de TNF)	Transdução de sinal; ligação de íons de zinco; regulação do processo apoptótico; ligação de íons metálicos; ligação de íons de zinco; ligação de proteína quinase
<i>CDC42BPB</i> (Proteína quinase beta de ligação ao CDC42)	Ligação de íons metálicos; atividade da proteína serina / treonina-quinase; atividade da proteína quinase; organização do citoesqueleto de actina; atividade da quinase; fosforilação da proteína.
<i>MARK3</i> (Quinase 3 reguladora da afinidade de microtúbulos)	Atividade da proteína quinase; atividade da proteína serina / treonina quinase; atividade da quinase; fosforilação da proteína.
ENSOARG00000005800	Atividade da creatina quinase; atividade da quinase; fosforilação; processo biossintético da fosfocreatina; atividade catalítica.

Continua

Continuação da Tabela Suplementar 5

BAG5 (Atanogene 5 associado a BCL2)	Ligação da ubiquitina à proteína ligase; ligação de chaperona; ligação de proteína quinase; organização de Golgi; actina filamentosa; regulação positiva da polimerização de filamentos de actina; formação de redes de filamentos de actina
CARMIL1 (Regulador de proteína de nivelamento e ligador 1 de miosina 1)	Actina filamentosa; regulação positiva da polimerização de filamentos de actina; formação de redes de filamentos de actina
ENSOARG00000005104	Ligação de íons de zinco; atividade da oxidoreductase; processo catabólico do heme; processo de redução da oxidação
<i>RIPOR2</i> (Regulador de polarização celular 2 que interage com a família RHO)	Desenvolvimento de fibras do músculo esquelético
<i>GPLD1</i> (Fosfolipase D1 específica de glicosilfosfatidilinositol)	Resposta celular ao triglicérido; resposta celular ao colesterol; resposta celular ao íon cálcio; regulação positiva do processo apoptótico; resposta celular ao estímulo à insulina.
KCNMA1 (Subfamília M alfa 1 de canais de potássio ativados por cálcio)	Transporte de íons potássio; resposta ao íon cálcio; atividade do canal de potássio ativado por cálcio; ligação de actina
GLRA1 (Receptor de glicina alfa 1)	Ligação de íons de zinco; contração muscular; ligação de taurina

Tabela Suplementar 6 - Genes candidatos imediatos presentes nas regiões relacionadas a EGS na espécie *Ovis aries* e funções biológicas mais relevantes.

Gene	Função disponível no Biomart (Uniprot)
<i>LOC101119348</i> (Gliceraldeído-3-fosfato desidrogenase)	Processo glicolítico; processo metabólico de glicose
<i>ADAM23</i> (Domínio ADAM metalopeptidase 23)	*Não foram identificados termos relevantes neste gene em ovinos, mas este apresenta funções ligadas a adipogênese em outras espécies
<i>EIF2AK2</i> (Fator de iniciação da tradução eucariótica 2 alfa-cinase 2)	Resposta ao interferon alfa; regulação negativa da proliferação de osteoblastos, atividade da proteína cinase, fosforilação da proteína
<i>OSBPL3</i> (Ligante de oxysterol semelhante a proteína 3)	Transporte de lipídios; ligação de colesterol
<i>GNAQ</i> (Subunidade alfa q da protein G)	Atividade ativadora de GTPase; Via sinalização do receptor acoplado à proteína G; Regulação da via de sinalização Wnt canônica
<i>DYTN</i> (Distrotelina)	Ligação de íons metálicos; ligação de íons zinco; interação seletiva e não covalente com íons (Zn).
<i>NEUROG3</i> (Neurogenina 3)	Desenvolvimento do pâncreas endócrino; Transdiferenciação
<i>QPCT</i> (Ciclotransferase de glutaminil-peptídeo)	Ligação de íons zinco.
<i>HKDC1</i> (Hexocinase contendo domínio 1)	Processo glicolítico; Atividade da glucocinase; Processo metabólico de carboidratos
<i>HK1</i> (Hexokinase 1)	Processo glicolítico; Atividade quinase; Ligação de glicose
<i>CDC42EP3</i> (Proteína efetora CDC42 3)	Citoesqueleto de actina
<i>MDH1B</i> (Malato desidrogenase 1B)	Atividade de malato desidrogenase; processo de redução de oxidação; atividade catalítica

**LISTA DE GENES CODIFICANTES PRESENTES NAS REGIÕES
ANALISADAS PARA ÁREA DE OLHO DE LOMBO**

<i>ABL2</i>	<i>EFNA1</i>	<i>ENSOARG00000014332</i>
<i>ACTL8</i>	<i>EFNA3</i>	<i>ENSOARG00000014409</i>
<i>ADAM15</i>	<i>EFNA4</i>	<i>FAM163A</i>
<i>ALDH5A1</i>	<i>EIF5</i>	<i>FAM189B</i>
<i>AMN</i>	<i>ENSOARG00000003500</i>	<i>FAM20B</i>
<i>ANGPTL1</i>	<i>ENSOARG00000004341</i>	<i>FAM65B</i>
<i>APOPT1</i>	<i>ENSOARG00000004408</i>	<i>FBXO38</i>
<i>ARHGEF10L</i>	<i>ENSOARG00000004453</i>	<i>FLAD1</i>
<i>ASH1L</i>	<i>ENSOARG00000004524</i>	<i>G3BP1</i>
<i>ATF3</i>	<i>ENSOARG00000004867</i>	<i>GBA</i>
<i>ATOX1</i>	<i>ENSOARG00000005104</i>	<i>GLRA1</i>
<i>ATP5MPL</i>	<i>ENSOARG00000005299</i>	<i>GMNN</i>
<i>AXDND1</i>	<i>ENSOARG00000005499</i>	<i>GPLD1</i>
<i>BAG5</i>	<i>ENSOARG00000005535</i>	<i>HCN3</i>
<i>BATF3</i>	<i>ENSOARG00000005800</i>	<i>HTR4</i>
<i>BEND3</i>	<i>ENSOARG00000005945</i>	<i>HTR5A</i>
<i>C20H6orf229</i>	<i>ENSOARG00000006061</i>	<i>JAKMIP2</i>
<i>C5orf46</i>	<i>ENSOARG00000006103</i>	<i>KCNMA1</i>
<i>C6H4orf19</i>	<i>ENSOARG00000006127</i>	<i>KCNN3</i>
<i>C7orf31</i>	<i>ENSOARG00000006181</i>	<i>KIAA0319</i>
<i>C8H6orf203</i>	<i>ENSOARG00000006331</i>	<i>KLC1</i>
<i>CARMIL1</i>	<i>ENSOARG00000006510</i>	<i>KRTCAP2</i>
<i>CD24</i>	<i>ENSOARG00000006565</i>	<i>LENEP</i>
<i>CDC42BPB</i>	<i>ENSOARG00000007159</i>	<i>LOC101104530</i>
<i>CKS1B</i>	<i>ENSOARG00000007166</i>	<i>LOC101106806</i>
<i>CLK2</i>	<i>ENSOARG00000007170</i>	<i>LOC101107954</i>
<i>DCST1</i>	<i>ENSOARG00000007197</i>	<i>LOC101116828</i>
<i>DCST2</i>	<i>ENSOARG00000007222</i>	<i>MARK3</i>
<i>DPM3</i>	<i>ENSOARG00000009394</i>	<i>MTX1</i>
<i>DPYSL3</i>	<i>ENSOARG00000011078</i>	<i>MUC1</i>
<i>DSEL</i>	<i>ENSOARG00000012275</i>	<i>MYO16</i>

<i>NPHS2</i>	<i>RD3L</i>	<i>TDP2</i>
<i>NPVF</i>	<i>SCAMP3</i>	<i>TDRD5</i>
<i>NSL1</i>	<i>SCGB3A2</i>	<i>TDRD9</i>
<i>OSBPL3</i>	<i>SHC1</i>	<i>THBS3</i>
<i>PAXIP1</i>	<i>SLC50A1</i>	<i>TNFAIP2</i>
<i>PBXIP1</i>	<i>SOAT1</i>	<i>TOR3A</i>
<i>PDSS2</i>	<i>SPATA45</i>	<i>TRAF3</i>
<i>PIK3R1</i>	<i>SPINK1</i>	<i>TRIM46</i>
<i>PKLR</i>	<i>SPINK5</i>	<i>TRMT61A</i>
<i>PMVK</i>	<i>SPINK6</i>	<i>WSB1</i>
<i>PPP1R13B</i>	<i>SPINK7</i>	<i>XRCC3</i>
<i>PPP2R2B</i>	<i>STK32A</i>	<i>ZBTB7B</i>
<i>RALGPS2</i>	<i>TATDN3</i>	

**LISTA DE GENES CODIFICANTES PRESENTES NAS REGIÕES
ANALISADAS PARA ESPESSURA DE GORDURA SUBCUTÂNEA**

<i>ADAM23</i>	<i>LOC101112574</i>
<i>AEBP2</i>	<i>LOC101119348</i>
<i>ANGPT1</i>	<i>MDH1B</i>
<i>BIRC6</i>	<i>MEMO1</i>
<i>BMP4</i>	<i>MPP6</i>
<i>C25H10orf35</i>	<i>NDUFAF7</i>
<i>CDC42EP3</i>	<i>NEUROG3</i>
<i>CEBPZ</i>	<i>NLRC4</i>
<i>CTSO</i>	<i>OSBPL3</i>
<i>DDHD1</i>	<i>PLEKHA5</i>
<i>DFNA5</i>	<i>PRKD3</i>
<i>DYTN</i>	<i>PSMA8</i>
<i>EIF2AK2</i>	<i>QPCT</i>
<i>ENSOARG00000010705</i>	<i>SLC30A6</i>
<i>ENSOARG00000011145</i>	<i>SPAST</i>
<i>ENSOARG00000018828</i>	<i>SRGN</i>
<i>ENSOARG00000020236</i>	<i>SS18</i>
<i>FASTKD2</i>	<i>SULT6B1</i>
<i>FBXL4</i>	<i>SUPV3L1</i>
<i>GNA14</i>	<i>TACR2</i>
<i>GNAQ</i>	<i>TAF4B</i>
<i>GPATCH11</i>	<i>TLE1</i>
<i>HEATR5B</i>	<i>TRIO</i>
<i>HK1</i>	<i>TSPAN15</i>
<i>HKDC1</i>	<i>VPS26A</i>
<i>KCTD1</i>	<i>WSB1</i>
<i>LOC101110483</i>	<i>YIPF4</i>
<i>ZFAT</i>	

