



LUÍS AUGUSTO EIJY NAGAI

IDENTIFICAÇÃO DE GENES RELACIONADOS À MACIEZ DA CARNE DE BOVINOS
DAS RAÇAS ANGUS E NELORE POR ANÁLISE DA EXPRESSÃO DIFERENCIAL

CAMPINAS

2015



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

LUÍS AUGUSTO EIY NAGAI

IDENTIFICAÇÃO DE GENES RELACIONADOS À MACIEZ DA CARNE DE BOVINOS
DAS RAÇAS ANGUS E NELORE POR ANÁLISE DA EXPRESSÃO DIFERENCIAL

Dissertação apresentada ao Instituto de
Biologia da Universidade Estadual de
Campinas como parte dos requisitos
exigidos para a obtenção do título de
Mestre em Genética e Biologia Molecular
na Área de Bioinformática.

Orientadora: Dr^a Paula Regina Kuser-Falcão

Co-orientador: Dr. Adhemar Zerlotini Neto

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA DISSERTAÇÃO DEFENDIDA PELO ALUNO LUÍS
AUGUSTO EIY NAGAI, E ORIENTADO PELA PROF.
DRA PAULA REGINA KUSER-FALCÃO

PROFA. DRA. PAULA REGINA KUSER-FALCÃO

CAMPINAS

2015

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

N131i Nagai, Luís Augusto Eijy, 1987-
Identificação de genes relacionados à maciez da carne de bovinos das raças Angus e Nelore por análise de expressão diferencial / Luís Augusto Eijy Nagai. – Campinas, SP : [s.n.], 2015.

Orientador: Paula Regina Kuser Falcão.
Coorientador: Adhemar Zerlotini Neto.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Expressão gênica. 2. Aberdeen-Angus (Bovino). 3. Nelore (Zebu). 4. Maciez da carne. I. Kuser-Falcão, Paula Regina. II. Zerlotini Neto, Adhemar. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Identification of genes involved in meat tenderness in Angus and Nellore breeds using differential expression analysis

Palavras-chave em inglês:

Gene expression

Aberdeen-Angus cattle

Nellore (Zebu)

Meat tenderness

Área de concentração: Bioinformática

Titulação: Mestre em Genética e Biologia Molecular

Banca examinadora:

Paula Regina Kuser Falcão [Orientador]

Luciana Correia de Almeida Regitano

Benilton de Sá Carvalho

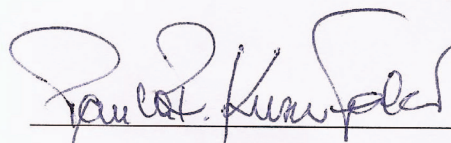
Data de defesa: 11-03-2015

Programa de Pós-Graduação: Genética e Biologia Molecular

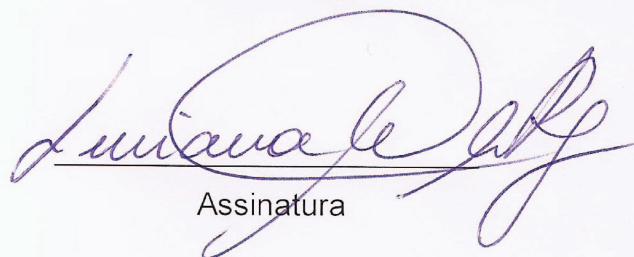
Campinas, 11 de Março de 2015.

Banca Examinadora

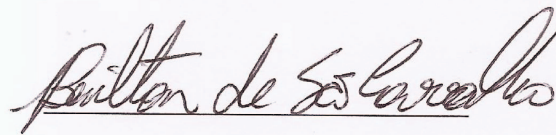
Profa. Dra. Paula Regina Kuser-Falcão


Assinatura

Profa. Dra. Luciana Correia de Almeida Regitano


Assinatura

Prof. Dr. Benilton de Sá Carvalho


Assinatura

Prof. Dr. Marcelo Falsarella Carazzolle

Assinatura

Prof. Dr. Michel Eduardo Beleza Yamagishi

Assinatura

RESUMO

A carne bovina é um dos principais produtos que o Brasil comercializa no mercado internacional. Atualmente, o Brasil é líder nas exportações e tem o objetivo de melhorar sua produção para aumentar a competitividade. A carne bovina é consumida em várias partes do mundo e é uma das principais fontes de proteínas e vitaminas para o homem. A publicação em 2009 do genoma bovino, proveniente de um animal da subespécie *Bos taurus taurus*, beneficiou os estudos de caracterização de mecanismos moleculares responsáveis por características de interesse. Dentre várias características, a maciez da carne é considerada o atributo mais importante. As raças zebuínas, principalmente Nelore (*Bos taurus indicus*), possuem grande aceitação na pecuária de corte brasileira e adaptabilidade ao território, porém apresentam menor produtividade e qualidade da carne, em relação às raças taurinas como Angus (*B. taurus taurus*). Assim, o interesse pela identificação de genes relacionados com a qualidade da carne bovina justifica o investimento em pesquisas genéticas, moleculares e bioquímicas. As novas tecnologias de sequenciamento de alto rendimento promoveram um grande avanço nas pesquisas de avaliação da expressão de genes, aliada às novas ferramentas computacionais e o poder de processamento de dados. O objetivo deste trabalho é identificar genes diferencialmente expressos no tecido muscular de animais das raças Angus e Nelore, para elucidar os mecanismos genéticos envolvidos com maciez da carne. Para realizar este estudo foram utilizados 10 bovinos da raça Angus e 8 bovinos Nelore, a partir de um rebanho de 50 touros de cada raça, selecionados pelos extremos no resultado do teste de força de cisalhamento. Ao todo, 313 milhões de fragmentos de 100 pares de bases foram gerados em Illumina HiScanSQ *paired end*. Ao longo desse estudo, foi também avaliada a utilização de sequências originais e filtradas por qualidade para estimar a eficácia das ferramentas atuais ao trabalhar com sequências que comumente são trimadas. As sequências foram mapeadas no genoma bovino utilizando Tophat2 e a biblioteca edgeR foi utilizada na análise estatística para a identificação de expressão diferencial. Os resultados encontrados nos animais Angus, com sequências não filtradas, se mostraram mais

semelhantes com o que tem sido observado em trabalhos relacionados à maciez da carne, os quais indicam o envolvimento da via de metabolismo de lipídeos no mecanismo de maciez. Contudo, os dados de Angus e Nelore possibilitaram a identificação de processos de resposta imune e inflamatórias, que poderiam ter sido influenciados por fatores externos pré abate.

Palavras chaves: RNA-Seq, maciez da carne, expressão diferencial.

ABSTRACT

Brazil is one of the largest beef exporters in the world, and beef quality is a major subject of research. The goal of this research is to improve the quality and productivity of Brazilian livestock. Beef is also an important source of protein and vitamins in human nutrition. Publication of The *Bos taurus* genome in 2009 has contributed to studies on the characterization of molecular mechanisms responsible for traits of interest, however, much remains to be determined. Among the many traits of interest, meat tenderness is the most important measured attribute. The Zebu breeds, particularly Nelore (*Bos taurus indicus*), have wide acceptance in the Brazilian market, and they are well adapted to tropical conditions. Nevertheless, Zebu cattle have lower productivity and beef quality when compared to Taurine breeds, such as Angus (*Bos taurus taurus*). The global market interest in the identification of genes involved in beef quality justifies the importance of research of genetic, molecular and biochemical mechanisms that determine this trait. The objective of this study is to identify differentially expressed genes found in the muscle tissue of Angus and Nelore breeds. To obtain the data to do this gene expression study, ten Angus steers and eight Nelore steers were used, they were chosen from a group of fifty animals of each breed, and were selected for their extreme tenderness and toughness scores on the shear force test. In total, 313 million fragments of a 100 base pairs were generated using Illumina HiScanSQ *paired end*. In parallel with the gene expression study, the use of trimmed or non-trimmed sequences was determined to evaluate the efficiency of the tools that have been used to filter this background noise. All sequences were mapped to the bovine genome using Tophat2, and the edgeR library provided the statistical analysis used to identify differential expression. Results obtained from the analysis of the data from the Angus breed animals with non filtered sequences are consistent with results observed in studies of beef tenderness. This finding implies the involvement of the lipid metabolism pathway in the biochemical mechanism underlying the improved tenderness. However, Nelore data analysis identified the existence of processes involved in inflammatory/immune response; they may have been influenced by external factors prior to the slaughter.

Keywords: RNA-Seq, beef tenderness, differential expression.

Sumário

1 INTRODUÇÃO GERAL.....	1
2 REVISÃO DE LITERATURA.....	3
2.1 A carne bovina brasileira.....	3
2.1.1 Nelore.....	4
2.1.2 Angus.....	5
2.2 Maciez da carne.....	5
2.3 Genoma bovino.....	6
2.4 NGS e RNA-Seq.....	7
2.5 Análise de expressão diferencial e funcional de genes.....	8
3 OBJETIVOS.....	11
3.1 Objetivo geral.....	11
3.2 Objetivos específicos.....	11
4 MATERIAIS E MÉTODOS.....	12
4.1 Animais.....	12
4.2 Coleta e seleção das amostras.....	12
4.3 Preparo das bibliotecas de mRNA e sequenciamento.....	13
4.4 Plataforma Galaxy e utilização dos dados.....	14
4.5 Tratamento das sequências e filtro de qualidade.....	15
4.6 Mapeamento das sequências no genoma de referência.....	15
4.7 Identificação dos transcritos.....	16
4.8 Matriz de contagens.....	16
4.9 Análise de expressão diferencial com edgeR.....	16
4.10 Identificação de vias enriquecidas.....	17
5 RESULTADOS.....	18
5.1 Seleção dos animais.....	18
5.2 Sequenciamento das amostras.....	19
5.3 Controle de qualidade.....	21
5.4 Mapeamento no genoma.....	26
5.5 Matriz de contagens.....	31
5.6 Expressão diferencial de genes.....	34
5.7 Identificação de vias enriquecidas.....	37
6 DISCUSSÃO.....	53
7 CONCLUSÃO.....	58
8 REFERÊNCIAS.....	59

Dedico este trabalho
aos meus queridos pais, Quioji e Iraci
que fizeram o possível e o impossível para garantir todas as
condições necessárias para o meu desenvolvimento até o presente momento.
Dedico este trabalho também a meu mestre da vida, Dr. Daisaku Ikeda, que me
incentivou a cultivar grandes sonhos e a superar meus limites.

AGRADECIMENTOS

À Dra. Paula R. Kuser-Falcão pela orientação, amizade, conselhos e confiança depositada em mim durante todo o período em que estive como aluno de mestrado no Programa da Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp e estagiário da Embrapa Informática Agropecuária.

Ao Dr. Adhemar Zerlotini pela orientação, confiança e insistência para com a minha pessoa durante todo o período que estive na Embrapa Informática Agropecuária. Agradeço por transmitir todo o conhecimento em bioinformática necessários para a realização deste trabalho e pela disponibilidade de sempre poder tirar minhas dúvidas e fazer significativas sugestões. Tanto Paula como Adhemar, tiveram a mesma importância, independente da posição de orientador e co-orientador, oferecendo em igual parte, cada um com sua habilidade, sua contribuição no decorrer do projeto e da minha formação.

À Dra. Poliana Giachetto pela grande contribuição em todos os momentos e pela grande preocupação, sempre, com os mínimos detalhes.

Aos demais pesquisadores e colaboradores do Laboratório de Bioinformática Aplicada, Dr. Michel Yamagishi, Dr. Felipe Silva, Dr. Francisco Lobo, Dr. Leandro Cintra, Ms. Joaquim Manoel da Silva, Alan Romaniuc, Fernanda Stringassi.

Ao Dr. Fernando Flores Cardoso por ter cedido os dados de bovinos utilizados para a realização deste trabalho.

Aos colegas e amigos de laboratório Giovanni, Edmar, Lucas, Danilo, Luiz, Jorge, André, Fernanda, Adriano, Renata, Rodrigo, Renato, Gustavo, enfim, todos que fizeram parte em algum momento nesta maravilhosa jornada na salinha dos estagiários do Laboratório Multiusuário de Bioinformática.

À Embrapa pela concessão da bolsa durante todo o período do mestrado e pela utilização de todo equipamento necessário.

À Dra. Luciana Regitano e FAPESP por permitir utilizar a ferramenta de bioinformática, IPA, dando grande contribuição na análise dos resultados.

À minhas irmãs que sempre estiveram ao meu lado nos bons e maus momentos.

Elas são um dos motivos por eu ter me dedicado ao máximo, para poder abrir caminhos e cumprir o papel de um bom irmão mais velho.

À Thaís Shirai, minha querida companheira, amiga, amante, faxineira, cozinheira, que me aguenta nos momentos de ânimo e desânimo desde o primeiro mês que cheguei nesta significativa cidade de Campinas.

À minha querida professora de japonês, Ayako Akamine, que abriu as portas para uma nova cultura e proporcionou a oportunidade de dar o prosseguimento em meus sonhos.

Aos meus companheiros da Associação Brasil Soka Gakkai Internacional, pela oportunidade de desenvolvimento, por darem apoio e incentivos nos momentos de dificuldades e pelo companheirismo, sempre.

A todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho, meu muito obrigado.

LISTA DE ILUSTRAÇÕES

Figura 1: Imagem ilustrativa do relatório do FASTQC, exemplificando uma sequência de boa qualidade (esquerda) e uma de má qualidade (direita).....	22
Figura 2: Imagem ilustrativa do relatório geral do FASTQC da qualidade por base, comparando as mesmas amostras antes (superior) e após (inferior) trimagem.....	23
Figura 3: Imagem ilustrativa do relatório geral do FASTQC da qualidade média na biblioteca (esquerda) e do tamanho médio das sequências na biblioteca (direita) após trimagem.....	26
Figura 4: Gráfico comparativo entre os grupos de sequências.....	33
Figura 5: Diagrama de Venn mostrando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne nas raças Angus e Nelore, utilizando as sequências originais.....	35
Figura 6: Diagrama de Venn mostrando a interseção entre Angus e Nelore com sequências filtradas.....	35
Figura 7: Diagrama de Venn apresentando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne da raça Angus, entre sequências originais e filtradas.....	36
Figura 8: Diagrama de Venn apresentando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne da raça Nelore, entre sequências originais e filtradas.....	36
Figura 9: Via do metabolismo de lipídeos. Via enriquecida nos GDEs regulados positivamente em Angus, utilizando-se as sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais resistente.....	39
Figura 10: Via de resposta imune, anti inflamatória. Via enriquecida nos GDEs regulados negativamente em Angus utilizando-se sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais resistente.....	41
Figura 11: Via de resposta antimicrobiana, anti inflamatória. Via enriquecida nos GDEs regulados positivamente em Nelore, utilizando-se as sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais resistente.....	43
Figura 12: Via relacionada a problemas dermatológicos, imunológicos e inflamatórios. Via enriquecida nos GDEs regulados negativamente em Nelore, utilizando-se as sequências originais. A maior intensidade da cor representa o maior nível de expressão no grupo com carne mais resistente.....	45
Figura 13: Via de desenvolvimento muscular e esquelético. Via enriquecida nos GDEs regulados negativamente em Nelore. utilizando-se as sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais resistente.....	46
Figura 14: Via de crescimento celular e desenvolvimento de tecido. Via enriquecida nos GDEs regulados negativamente em Nelore, utilizando-se as sequências originais. A maior intensidade da cor representa um maior expressão no grupo com carne mais resistente.....	47

LISTA DE ABREVIATURAS

ANAN – Amostra referente a animais Angus

cDNA – DNA complementar

CPM – *Counts Per Million*

Ct – *Cycle threshold*

DNA – Ácido Desoxirribonucleico (ADN)

Embrapa – Empresa Brasileira de Pesquisa Agropecuária

ESALQ – Escola Superior de Agricultura Luiz de Queiroz

FAO – *Food and Agriculture Organization of the United Nations*

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo

FC – *Fold Change*

FDR – *False Discovery Rate*

FTP – *File Transfer Protocol*

Gb – Giga bases

GDE – Genes Diferencialmente Expressos

GLM – *Generalized Linear Model*

IBGE – Instituto Brasileiro de Geografia e Estatística

IPA – *Ingenuity Pathway Analysis*

logFC – *Log of Fold Change*

MAPA – Ministério da Agricultura, Pecuária e Abastecimento

NCBI – *National Center for Biotechnology Information*

NENE – Amostra referente a animais Nelore

NGS – *Next-Generation Sequencing*

pb – Pares de bases

pM – pico Molar

QTL – *Quantitative Trait Loci*

RIN – *RNA Integrity Number*

RNA – Ácido Ribonucleico (ARN)

RNA-Seq – *RNA Sequencing*

RT-qPCR - PCR quantitativo em tempo real

R1 – Sequência *forward* da biblioteca

R2 – Sequência *reverse* da biblioteca

SNP – *Single Nucleotide Polymorphism*

TEC – Tonelada Equivalente Carcaça

UNICAMP – Universidade Estadual de Campinas

USDA – *United States Department of Agriculture*

USP – Universidade de São Paulo

WBS – *Warner-Bratzler shear force*

1 INTRODUÇÃO GERAL

O Brasil detém uma posição de destaque no mercado mundial de carne bovina. Segundo levantamento feito pelo IBGE (2013), o país possui cerca de 211 milhões de cabeças de gado, sendo considerado o maior rebanho comercial do mundo. Desde 2008, o país lidera as exportações de carne bovina no mercado mundial e, em 2013, exportou cerca de 2 milhões de toneladas de equivalente carcaça (TEC) para diversos países. Atrás do Brasil seguem: Índia, Austrália e Estados Unidos. De acordo com o Departamento de Agricultura dos Estados Unidos (USDA), em 2014, o consumo mundial de carne bovina atingiu 57,8 milhões de TEC, no qual o Brasil contribuiu com 9,9 milhões de TEC, o que corresponde a 17% do total consumido. Conforme a projeção do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) em conjunto com a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a produção de carne bovina nacional até 2023 terá um crescimento de 22,8%.

A produção por animal no Brasil tem o custo significativamente menor quando comparado a grandes produtores como a Austrália e os Estados Unidos, onde o custo chega a ser 60% e 50% maior para os respectivos países (FERRAZ; FELÍCIO, 2010). Entretanto, a produção por animal precisa ser mais eficiente para ser capaz de suprir a demanda no futuro.

De 1975 a 2011, o Brasil reduziu sua área com pastagens em 8% e duplicou o seu rebanho. Um dos principais fatores que explicam o sucesso é a utilização de animais da raça Zebu (*Bos taurus indicus*) (MAPA, 2012).

Os animais da raça Zebu possuem comprovada rusticidade e resistência ao calor devido a sua descendência indiana e têm sido amplamente utilizados em programas de melhoramento para o cruzamento e otimização das características da prole (CROUSE et al., 1989). Dentre as raças de corte, o Nelore possui grande destaque no rebanho nacional, ocupando 80% da parcela.

Assim, com a grande oferta de pastagens e um rebanho composto em sua maioria por raças melhor adaptadas, a pecuária extensiva, de um modo geral, é a

principal forma de criação de bovinos (FERRAZ; FELÍCIO, 2010).

Contudo, se por um lado os animais da raça Zebu possuem comprovada resistência ao calor e a ecto- e endoparasitas, por outro lado são conhecidos por apresentarem menor produtividade, carne mais rígida e menos marmorizada que os animais das raças Taurinas (CROUSE et al., 1989).

Entre os animais de origem europeia, o Angus é uma raça que tem se destacado no Brasil por contribuir com o aumento de qualidade e produtividade em programas de melhoramento genético.

Além da técnica de cruzamento entre as raças, os programas de melhoramento genético animal passaram a utilizar tecnologias genômicas, proporcionando uma grande contribuição com o setor produtivo nos últimos anos (ROLF et al., 2014). A busca por novas formas de melhorar a qualidade e a maciez da carne tem possibilitado a identificação de vários genes candidatos importantes, *Quantitative Trait Loci* (QTL) e *Single Nucleotide Polimorphism* (SNP). Além disso, as recentes tecnologias de sequenciamento de nova geração (NGS) e a sequência do genoma bovino publicado possibilitam mais ampla exploração dos diversos mecanismos biológicos, dentre eles, os relacionados à maciez e qualidade da carne em bovinos (ZHAO et al., 2012).

Assim, utilizando a técnica de sequenciamento de RNA (RNA-Seq), é possível estudar o perfil de expressão do transcriptoma bovino, identificando genes diferencialmente expressos e contribuir para a compreensão dos possíveis mecanismos moleculares relacionados à maciez da carne de bovinos.

2 REVISÃO DE LITERATURA

2.1 A carne bovina brasileira

A carne bovina é uma importante fonte de nutrição e sustento para cerca de 6,6 bilhões de pessoas no mundo (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al., 2009). Faz parte de uma dieta balanceada de nutrientes em uma alimentação saudável, pois contém proteínas, vitaminas, minerais e micronutrientes em níveis essenciais para o crescimento e desenvolvimento humano (FAO, 2015).

O Brasil ocupa uma posição de destaque no mercado mundial de carne bovina. Segundo levantamento feito pelo IBGE (2013), o país possui cerca de 211 milhões de cabeças e é considerado o segundo maior rebanho do mundo, ficando atrás da Índia, que possui aproximadamente 301 milhões de cabeças (USDA, 2014). Porém, na Índia, por influência religiosa, a maioria dos bovinos não sofre abate. Sendo assim, o maior rebanho utilizado para produção comercial de carne é o brasileiro, seguido pela China (103 milhões) e Estados Unidos (87,7 milhões) .

Dados do Departamento de Agricultura dos Estados Unidos (USDA) mostram que o Brasil lidera as exportações de carne bovina no mercado mundial desde 2008 e a previsão do MAPA é que a exportação de carne bovina cresça 2,15% ao ano. Em 2013, foram exportadas cerca de 2 milhões de toneladas de equivalente carcaça (TEC) para diversos países. Outros grandes países exportadores de carne bovina para o mercado mundial são Índia, Austrália e Estados Unidos.

Em 2014, o consumo mundial de carne bovina atingiu 57,8 milhões de TEC, sendo que o Brasil contribuiu com 9,9 milhões de TEC, que corresponde a 17% do total consumido no ano (USDA, 2014).

A produção por animal no Brasil tem custo significativamente menor quando comparado a grandes produtores como a Austrália e os Estados Unidos, onde o custo chega a ser 60% e 50% maior para os respectivos países (FERRAZ; FELÍCIO, 2010).

O Brasil possui um território de 8,5 milhões de km² de extensão e estima-se que 1,72 milhão de km² são utilizados como pastagens. De 1975 a 2011, o Brasil reduziu sua área com pastagens em 8% e duplicou o seu rebanho. Um dos principais fatores que explicam o sucesso é a utilização de animais da sub-espécie *Bos taurus indicus* (MAPA, 2012).

Estima-se que 80% do rebanho bovino brasileiro possui constituição genética de origem zebuína, entre animais puros ou provenientes de cruzamentos (ABIEC, 2013).

Entretanto, embora o Brasil possua maior quantidade de cabeças e menor custo de produção, ao se comparar a produtividade por animal com países como Estados Unidos e Austrália, observa-se considerável inferioridade (USDA, 2014). A produtividade animal no Brasil precisa ser melhorada para acomodar a demandas futuras, evitando a ampliação da utilização de áreas com pastagens.

2.1.1 Nelore

Os animais zebuínos possuem comprovada rusticidade e resistência ao calor devido a sua descendência indiana. São utilizados amplamente em programas de melhoramento genético para o cruzamento e otimização da prole devido à sua característica de resistência (CROUSE et al., 1989; PAGE et al., 2004; PRADO et al., 2012; SHACKELFORD et al., 1991). Dentre as raças zebuínas de corte, o Nelore possui grande destaque no rebanho nacional ocupando uma parcela de 80%.

Entretanto, os animais *Bos taurus indicus* possuem carne mais dura e com menor quantidade de gordura intramuscular (marmorização) quando comparados aos animais *Bos taurus taurus*. Mesmo em cruzamentos entre taurinos e zebus, uma contribuição genética maior de *Bos taurus indicus* já confere resultados parecidos aos de animais zebuínos puros à prole (CROUSE et al., 1989; JOHNSON et al., 1990).

2.1.2 Angus

Entre as várias raças de origem europeia, o Angus recebe grande destaque por reunir uma grande quantidade de qualidades positivas em uma única raça. Uma das suas qualidades é a gordura intramuscular (marmorização) que colabora com a maciez da carne e com o sabor. Mesmo em criação extensiva possui grande rendimento comparado às outras raças taurinas (CHAMBAZ et al., 2003).

Além disso, atrai grande interesse devido a sua precocidade, podendo ser levado ao abate mais cedo e tendo uma produção maior (IGARASI et al., 2008).

Embora muitos pontos positivos, os animais Angus são menos produtivos em climas tropicais, devido a menor adaptação ao calor e por possuírem menor resistência a ecto- e endoparasitas. Dessa forma, os bovinos de origem europeia puros têm sido criados na região sul do país. Os cruzamentos, que possuem maior porcentagem de contribuição de *Bos taurus taurus*, são encontrados em todas as regiões e trazem bons resultados (FERRAZ; FELÍCIO, 2010).

2.2 Maciez da carne

A satisfação ao comer carne é resultado da interação entre maciez, suculência e sabor (CROUSE et al., 1989). Entre os vários fatores que influenciam o consumidor na hora da compra do produto, a maciez da carne é o principal fator (MILLER et al., 2001; MORGAN et al., 1991; SAVELL et al., 1991).

Lehmann (1907) foi um dos primeiros pesquisadores na literatura a fazer contribuições científicas sobre estudos com maciez da carne. Ele descreve que o aumento da maciez da carne de carcaças armazenadas em frigorífico ocorre durante o processo de *pos mortem*. Embora se tenham registros de longa data, segundo Koohmaraie (1994), o mecanismo por trás da variação da maciez da carne permaneceu uma incógnita e apresentava controvérsias.

Durante o processo de produção da carne, há poucos fatores que provocam

variação na suculência e no sabor. Entretanto, para a maciez, vários fatores podem influenciar o processo de amaciamento, entre eles idade, sexo, nível de estresse pré abate (ZHAO et al., 2012), força iônica, temperamento (TITTO et al., 2010), degradação de proteínas miofibrilares (KOOHMARAIE, 1994), tamanho do sarcômero (SMULDERS et al., 1990), pH, alimentação, raça (CROUSE et al., 1989; SHACKELFORD et al., 1991) entre outros. Portanto, obter carne com maior maciez requer maiores conhecimentos e cuidados pela indústria da carne durante toda a cadeia produtiva (KOOHMARAIE et al., 2002).

A genética e os fatores de criação afetam o tipo de fibra, quantidade e tipo de colágeno, gordura intramuscular e atividade proteolíticas, que por sua vez, regulam a maciez e o sabor. Assim, as características biológicas do músculo são controladas pela expressão dos genes. Como consequência, diversos trabalhos revelaram diferentes genes e proteínas que possuem papel no mecanismo da maciez e sabor da carne (BERNARD et al., 2007).

Um dos mecanismos mais conhecidos e estudados é o da calpaína na ação proteolítica das fibras musculares (KOOHMARAIE, 1996), e da calpastatina que atua inibindo a calpaína no período *post mortem* (após abate) (CORVA et al., 2007). Vários outros trabalhos elucidaram diferentes genes candidatos como o da miostatina (WHEELER et al., 2001), leptina (SCHENKEL et al., 2005), DNAJA1 (BERNARD et al., 2007), KCNJ11 (TIZIOTO et al., 2014).

Com o advento das novas tecnologias de sequenciamento e a publicação da sequência do genoma bovino, pesquisadores podem dar mais um passo na caracterização molecular dos mecanismos responsáveis pela maciez da carne.

2.3 Genoma bovino

Dentre os animais domésticos, os bovinos recebem destaque devido às grandes possibilidades de melhoramento genético e aumento de produtividade pela indústria de carnes e leite (KOOHMARAIE et al., 2002; ROLF et al., 2014).

Nesse sentido, o genoma bovino publicado tem contribuído para o estudo da evolução em mamíferos e para o melhor entendimento dos mecanismos moleculares, contribuindo para o melhoramento genético de bovinos de corte e de leite (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al., 2009).

O projeto foi concluído em 2009 e teve duração de 6 anos, tendo sido liderado pelo *Bovine Genome Sequencing and Analysis Consortium* e pelo Departamento de Agricultura dos Estados Unidos (USDA), com o envolvimento de mais de 300 cientistas de 25 países.

O genoma bovino possui 2,87 Gb (Giga bases, 2,87 bilhões de pares de bases) de tamanho e contém aproximadamente 22 mil genes, dos quais 14.345 são comuns a sete espécies de mamíferos e 16.177 compartilhados com humanos. Verificou-se que os rearranjos no genoma estão relacionados a genes envolvidos em processos de reprodução, imunidade, lactação e digestão (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al., 2009).

Depois do genoma bovino sequenciado, novas formas de análises tiveram um suporte maior para serem realizadas como, por exemplo, a detecção de alterações pontuais no genoma, conhecidas por polimorfismo de base única (SNP – *Single Nucleotide Polymorphism*) (LIU et al., 2010; THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al., 2009) e o estudo de genes diferencialmente expressos, utilizando-se o transcriptoma para caracterização molecular de mudanças fenotípicas (CHITWOOD et al., 2013; MCCABE et al., 2012; SCHOLEY et al., 2013).

2.4 NGS e RNA-Seq

A tecnologia de sequenciamento em larga escala ou sequenciamento de próxima geração (NGS – *Next Generation Sequencing*) trouxe grandes mudanças na genômica e nas diversas variações das “ômicas” existentes atualmente. Caracterizada por produzir milhões de pequenas sequências por um custo reduzido por base sequenciada, o sequenciamento de próxima geração é um dos principais responsáveis

pela grande quantidade de informações geradas e utilizadas atualmente em pesquisas e diagnósticos (BAI; SARTOR; CAVALCOLI, 2012). O NGS emergiu como uma grande ferramenta para estudos de transcriptoma, dada sua precisão na quantificação de transcritos (WANG; GERSTEIN; SNYDER, 2009).

Dentre as técnicas que utilizam NGS, o sequenciamento de moléculas de RNA (RNA-Seq) demonstra grande versatilidade em comparação com o microarranjo, ferramenta de análise de transcriptomas largamente utilizada até então (MORTAZAVI et al., 2008). O RNA-Seq possibilita a obtenção do perfil de expressão dos genes a partir de uma amostra de RNA e a identificação de novos transcritos, regiões regulatórias e eventos de *splicing* alternativo com considerável sensibilidade e precisão. Além disso, comparativamente, possui uma maior cobertura e alto rendimento para geração de dados, o que torna de grande importância para a busca de novos candidatos a genes com expressão diferencial entre diferentes condições (NOOKAEW et al., 2012; OSHLACK; ROBINSON; YOUNG, 2010).

Compreender o processo de transcrição dos genes é essencial para interpretação dos elementos funcionais do genoma (WANG; GERSTEIN; SNYDER, 2009). Porém, a maior dificuldade atualmente em lidar com uma grande quantidade de dados é extrair o significado biológico (ROBINSON; OSHLACK, 2010).

2.5 Análise de expressão diferencial e funcional de genes

Com o avanço das tecnologias de sequenciamento e o crescente aumento na geração de dados, novos programas estão sendo criados, atualizados ou adaptados para analisar grandes conjuntos de dados.

Há uma grande quantidade de formatos para os diversos tipos de dados utilizados em bioinformática. Os dados gerados pelo sequenciamento são comumente convertidos para o formato FASTQ. Este formato, que surgiu na área de sequenciamento de DNA, inclui uma pontuação de acurácia estimada do sequenciamento, conhecida por Phred Quality Score, determinada para cada

nucleotídeo da sequência. Essa pontuação indica a possibilidade da base ter sido sequenciada incorretamente.

Atualmente existem diversas metodologias e ferramentas para análise de expressão diferencial de genes. Os programas para mapeamento das sequências a um genoma de referência são parametrizados para lidar com variação na qualidade das bases (MCCABE et al., 2012; TRAPNELL et al., 2012). Desta forma, se eventualmente a sequência estiver errada, ela possivelmente não será mapeada no genoma de referência, ou se mapeada, a qualidade de mapeamento refletirá a incerteza associada.

Os arquivos provenientes do mapeamento, com extensão SAM (*Sequence Alignment Map*) ou BAM (*Binary version of SAM*) possuem 11 colunas que contém informações sobre o mapeamento, tais como posição de mapeamento e qualidade da sequência (SAMTOOLS, 2014). A informação apresentada na quinta coluna traz detalhes do mapeamento utilizados para se produzir a matriz de contagens. Através desta coluna, o programa pode eliminar sequências não mapeadas ou mapeadas em várias posições diferentes do genoma de referência.

Existem vários programas para realizar a análise estatística de expressão diferencial (SONESON; DELORENZI, 2013). Até o presente momento, não existe um consenso sobre qual a melhor ferramenta para se utilizar nas análises (ZHANG et al., 2014). DESeq e EdgeR são ferramentas amplamente utilizadas em pesquisas recentes (ANDERS et al., 2013; DILLIES et al., 2013; ZHANG et al., 2014), assim como o Cuffdiff2 (TRAPNELL et al., 2013). De acordo com os resultados do trabalho de Zhang et al (2014), o DESeq não é recomendado quando as bibliotecas apresentam muita variação na profundidade do sequenciamento, o Cuffdiff2 é a ferramenta recomendada para análise de bibliotecas com maior profundidade de sequenciamento, ou seja, bibliotecas com mais de 10 milhões de sequências e o edgeR é a ferramenta mais adequada para o caso de bibliotecas com maior variação de quantidade de sequências.

Identificar genes diferencialmente expressos (GDEs) que estejam envolvidos na maciez da carne em bovinos através das novas tecnologias de sequenciamento é um importante passo no caminho para decifrar os mecanismos genéticos que controlam a maciez.

A utilização da análise de expressão diferencial tem por objetivo principal identificar genes cujo nível de expressão se altera entre diferentes condições (ANDERS et al., 2013). Em outras palavras, utiliza-se uma tabela de contagens para cada biblioteca para realizar o teste estatístico entre as amostras de interesse e identificar genes com expressão diferencial entre as condições experimentais. Depois de identificados os genes, estes são usados para confirmação da hipótese da pesquisa com subsequentes análises (ANDERS et al., 2013).

O EdgeR é um pacote do R/Bioconductor (ROBINSON; MCCARTHY; SMYTH, 2010). Esta ferramenta utiliza informações de uma matriz de contagens para análise de expressão diferencial. Foi idealizada para utilizar dados de bibliotecas com réplicas para fazer comparações entre uma ou mais condições experimentais. Essa ferramenta se baseia na distribuição binomial negativa e utiliza a normalização TMM (ROBINSON; MCCARTHY; SMYTH, 2010; ROBINSON; OSHLACK, 2010). Segundo o autor, a normalização dos dados é uma etapa crucial durante a análise, que possibilita remover o viés técnico e melhorar a acurácia nos resultados da análise de expressão diferencial.

A análise de enriquecimento funcional, por sua vez, pode auxiliar na interpretação biológica dos dados, bem como a caracterização funcional dos genes. Os genes diferencialmente expressos são relacionados em um banco de dados de informações públicas curadas manualmente no Ingenuity Pathway Analysis (IPA). De acordo com o enriquecimento funcional atribuído a uma lista de genes, vias metabólicas podem ser inferidas, o que contribui para a compreensão do papel de um grupo de genes na condição indicada pelo trabalho (JIMÉNEZ-MARÍN et al., 2009).

3 OBJETIVOS

3.1 Objetivo geral

- Este trabalho tem por objetivo geral identificar genes envolvidos com a maciez da carne, no tecido muscular de bovinos das raças Angus e Nelore, por meio de análise de expressão diferencial e análise funcional dos genes.

3.2 Objetivos específicos

1. Selecionar genes diferencialmente expressos, relacionados à maciez da carne;
2. Identificar categorias funcionais dos genes a partir da literatura e bases de dados;
3. Realizar a comparação do resultado final ao se utilizar sequências trimadas e originais.

4 MATERIAIS E MÉTODOS

4.1 Animais

Foram utilizados animais experimentais da Embrapa Pecuária Sul, Bagé, Rio Grande do Sul, gerados no âmbito do projeto “BifeQuali - Bife de Qualidade”. Foram criados 50 animais machos da raça Nelore e 50 machos da raça Angus, de diferentes pais, mantidos após a desmama em pastagem cultivada de azevém, trevo branco e cornichão durante o inverno e em pastagem melhorada durante a primavera-verão-outono, com oferta de forragem de, em média, 12% do peso vivo e com suplementação mineral à vontade. Os animais foram abatidos quando atingiram cobertura de gordura superior a 3 mm e peso próximo de 480 kg, entre 22 e 26 meses de idade. O abate ocorreu no frigorífico Mercosul S/A, em Bagé, Rio Grande do Sul, e as carcaças foram avaliadas e tipificadas pelo Serviço de Inspeção Federal do Ministério da Agricultura.

4.2 Coleta e seleção das amostras

Imediatamente após o abate, foram coletadas amostras de músculo esquelético dos 100 animais envolvidos no projeto inicialmente e congeladas em nitrogênio líquido para posterior extração do RNA total a ser utilizado na construção das bibliotecas. Realizou-se, posteriormente, em outras amostras coletadas, a avaliação instrumental da textura da carne (análise de maciez) no *Longissimus dorsi* (contrafilé) e a determinação da força de cisalhamento, com a utilização de um “Texture Analyser TA-500”, equipado com célula de carne “Warner-Bratzler”. Foram determinados, dentro da raça Angus, os cinco indivíduos de carne mais macia e os cinco de carne mais dura, e na raça Nelore, os quatro indivíduos de carne mais macia e os quatro de carne mais dura. A seleção das amostras para este trabalho foi efetuada pelo Dr. Fernando Flores Cardoso da Embrapa Pecuária Sul, através do resíduo de um modelo linear envolvendo raça, data de abate e idade do animal.

Amostras de RNA total do tecido muscular dos indivíduos com valores extremos

para força de cisalhamento foram obtidas. A extração do RNA total, bem como a construção e sequenciamento das bibliotecas, foram realizados pelo Laboratório de Biotecnologia Animal, na Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP), a partir de 0,05 a 0,10 g de tecido muscular, utilizando o protocolo do reagente Trizol (Invitrogen, Carlsbad, CA), conforme as instruções do fabricante.

A pureza e a quantidade do RNA total extraído foram avaliadas por meio de eletroforese em gel de agarose 1% corado com GelRed e por espectrofotometria a 260 e 280 nm, através do espectrofotômetro Nanodrop 1000 (Thermo Scientific). A qualidade e a integridade do RNA total foram analisadas com o Agilent 2100 Bioanalyser (Agilent Technologies). Apenas amostras com RNA total apresentando *RNA Integrity Number* (RIN) maior ou igual a 8,0 foram utilizadas para as análises de RNA-Seq.

4.3 Preparo das bibliotecas de mRNA e sequenciamento

Cada amostra, contendo de 1 a 4 µg de RNA total, foi purificada de acordo com o protocolo descrito pela Illumina, utilizando microesferas magnéticas ligadas a oligos dT para separação do RNA mensageiro. O mRNA purificado foi fragmentado em tampão específico (Illumina). A síntese da primeira fita de cDNA foi realizada com a enzima Superscript II (Invitrogen). A síntese da segunda fita de cDNA foi realizada com as enzimas RNaseH e DNA Pol I e as extremidades das moléculas foram tratadas com as enzimas T4 DNA Polymerase e Klenow DNA Polymerase (Illumina) para a geração de extremidades abruptas. A extremidade 5' foi fosforilada com a enzima T4 PNK (e a adenilação da extremidade 3' foi realizada com a enzima Klenow exo, ambas da Illumina). As moléculas de cDNA foram então ligadas à adaptadores com uso da enzima T4 DNA Ligase (Illumina). As bibliotecas foram em seguida submetidas a eletroforese em gel de agarose, purificadas (fração correspondente a 200 pb ±30 pb), e amplificadas por meio de PCR, com primers específicos para os adaptadores (Illumina).

Após o término da preparação, a qualidade das bibliotecas foi validada no Agilent 2100 Bioanalyser (Agilent Technologies) usando o chip DNA 1000, com intuito de

verificar se o procedimento foi bem sucedido e se as amostras apresentaram a maior parte dos fragmentos próximos do tamanho de 260 pb.

A seguir, as bibliotecas foram quantificadas individualmente via PCR em Tempo Real (RT-qPCR), com o uso do Kit KAPA Library Quantification (KAPA Biosystems), o qual possui seis amostras padrões de concentrações (entre 20 a 0,0002 pM) e tamanho de fragmento conhecido (452 pb) presentes no kit. A partir dos valores de *Cycle threshold* (Ct) dos padrões foram determinadas as concentrações das amostras por meio de uma regressão linear. Depois de calculadas essas concentrações, elas foram diluídas novamente para uma concentração padronizada, combinadas e então inseridas na lâmina de sequenciamento para a clusterização.

O sequenciamento do mRNA foi realizado com a tecnologia de sequenciamento de nova geração por meio do equipamento HiScanSQ (Illumina – San Diego, EUA), de acordo com as instruções do fabricante, utilizando-se o protocolo para sequências *pair-end*. Com uso do kit TruSeq PE Cluster Kit v3-cBot-HS (Illumina – San Diego, EUA), as amostras ligaram-se a oligos complementares aos adaptadores posicionados na superfície das canaletas da lâmina, e foram amplificadas, resultando em *clusters* (agrupamentos) de fragmentos iguais. A sequência dos agrupamentos foi determinada com uso do kit TruSeq SBS kit v3-HS (Illumina – San Diego, EUA), (200 cycles), conforme recomendações do fabricante. Cada canaleta conteve seis amostras e cada amostra foi sequenciada até gerar cerca de 10 milhões de sequências por biblioteca.

Através do programa CASAVA, versão 1.8.2, os dados gerados pelo sequenciador HiScanSQ (Illumina, Inc., San Diego, CA), foram convertidos para o formato FASTQ.

4.4 Plataforma Galaxy e utilização dos dados

Os dados gerados com o sequenciamento das bibliotecas foram inseridos na instância da plataforma Galaxy (GIARDINE et al., 2005) da Embrapa. Toda a análise de dados foi conduzida no Laboratório Multiusuário de Bioinformática da Embrapa, localizado na Embrapa Informática Agropecuária (Campinas-SP).

4.5 Tratamento das sequências e filtro de qualidade

Os dados foram inicialmente processados pelo programa FastQC (ANDREWS, S., 2010), versão 0.10.0, o qual gerou um relatório global da composição e qualidade de cada biblioteca. Subsequentemente, as bibliotecas foram individualmente filtradas por qualidade, utilizando-se o programa FASTQ Quality Trimmer, pertencente ao pacote FastX (http://hannonlab.cshl.edu/fastx_toolkit/) versão 0.0.13. Esse filtro elimina os nucleotídeos das extremidades considerando-se a pontuação Q (Phred score) (EWING; GREEN, 1998) de uma janela deslizante de tamanho ajustável. Os parâmetros utilizados foram: manter os IDs de sequências de tamanho 0; verificar as extremidades 5' e 3' de cada sequência; tamanho de janela de 3 nucleotídeos com deslizamento de base por base, sem limite de exclusão de bases; eliminar a última base da janela que tenha média de pontuação igual ou inferior a Q20. A pontuação Q20 na escala Phred corresponde a 1% de probabilidade da base nucleotídica estar incorreta.

4.6 Mapeamento das sequências no genoma de referência

O programa TopHat, versão 2.0.9 (KIM et al., 2013; TRAPNELL et al., 2012), mapeou os pares de sequências filtradas de cada biblioteca ao genoma de referência do *Bos taurus*, versão 3.1 Refseq UMD. Este genoma está disponível no diretório de ftp do *National Center for Biotechnology Information* – NCBI <ftp://ftp.ncbi.nih.gov/genomes/Bos_taurus/>. Foram utilizados os parâmetros padrões da ferramenta para bibliotecas de pares de sequência com distância média interna igual a 0, admitindo-se desvio padrão de 20, máximo de 2 *mismatches* e tamanho mínimo de 25 para cada segmento. O mapeamento no TopHat gera o campo CIGAR no arquivo BAM de saída. Este campo possui informações do alinhamento da sequência que são utilizados pelo HTSeq-count, para definir quais sequências serão utilizadas para a contagem.

4.7 Identificação dos transcritos

Os transcritos de cada biblioteca de sequências foram montados através do Cufflinks, versão 2.1.1 (TRAPNELL et al., 2010, 2012), utilizando as sequências previamente mapeadas pelo programa TopHat2. O Cuffmerge, versão 1.0.0 (TRAPNELL et al., 2010), utilizou os transcritos montados das bibliotecas de sequências trimadas e originais para reconstruir o transcriptoma e incluiu a anotação para os genes. Esta etapa foi necessária para criar um arquivo de referência para posterior comparação de genes entre as análises utilizando sequências trimadas e originais.

4.8 Matriz de contagens

A matriz de contagem foi produzida pelo programa HTSeq-count (ANDERS; PYL; HUBER, 2014), versão 0.6.1p1, que utilizou arquivos provenientes do mapeamento de cada biblioteca efetuado pelo programa TopHat e anotação proveniente da execução do Cufflinks. Foi utilizado o modo *intersection nonempty* para contagem das sobreposições dos mapeamentos das sequências nos genes, no qual as sequências que alinharam em mais de um gene são consideradas ambíguas e descartadas da contagem. Foram criadas matrizes distintas para cada raça. O programa HTSeq-count realiza a quantificação individual dos fragmentos de sequências em cada gene por biblioteca.

4.9 Análise de expressão diferencial com edgeR

A análise de expressão diferencial entre os extremos de maciez dentro de cada raça, foi realizada pela biblioteca EdgeR (ROBINSON; MCCARTHY; SMYTH, 2010), versão 3.4.2, executada no ambiente R versão 3.1.0 “Spring Dance” (<http://www.r-project.org/>). Esta ferramenta, também executada dentro da plataforma Galaxy, utilizou as matrizes de contagens para cada raça (Angus sequências originais, Nelore sequências originais, Angus sequências filtradas, Nelore sequências filtradas) como

entrada para análise de expressão diferencial de genes e produziu: a) uma matriz normalizada (pelo tamanho da biblioteca) utilizada para o cálculo da expressão diferencial, b) uma tabela contendo o nível de expressão relativa, log do CPM (*Counts Per Million* – Contagem Por Milhão) e, teste estatístico de probabilidade, e c) código fonte executado no R e algumas imagens comparativas.

Os parâmetros utilizados na biblioteca EdgeR foram: análise de Modelos Lineares Generalizados - GLM (*Generalized Linear Models*), cálculo da Dispersão Comum (*Common Dispersion*) e correção por Taxa de Falsas Descobertas - FDR (*False Discovery Rate*) por Benjamini and Hochberg (1995), sendo considerados diferencialmente expressos apenas aqueles genes que apresentaram FDR menor que 0,05.

4.10 Identificação de vias enriquecidas

Após a identificação dos genes diferencialmente expressos entre os grupos de animais extremos para maciez da carne, dentro de cada raça, foi realizada a análise de enriquecimento funcional, a partir das listas geradas, utilizando-se o programa QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity). Este programa possibilita a interpretação biológica de uma lista de genes, utilizando banco de informações curadas manualmente.

A partir das listas de genes diferencialmente expressos identificados pelo edgeR, foi criado um arquivo contendo o campo ID (gene) e o logFC (*log of Fold Change* – log da Razão das Contagens) para entrada de dados para o IPA. Foi utilizada a função *core analysis* com os parâmetros: todos os tipos de identificadores, todas as espécies contidas no banco de dados (humano, rato, camundongo e sem caracterização), *cutoff* de 1 baseado no logFC e análises distintas dos grupos de genes de acordo com o logFC (*upregulated* e *downregulated*). Para atribuir função biológica, os dados enviados são mapeados à base de dados do IPA, e em seguida são criadas as redes moleculares pelo algoritmo que irá calcular a significância, baseado na quantidade de genes/moléculas que mapeiam em uma função, rede ou via.

5 RESULTADOS

5.1 Seleção dos animais

Os animais utilizados para análise de expressão diferencial foram selecionados com base no teste de força de cisalhamento Warner-Bratzler, utilizando-se o aparelho Texture Analyser TA-500 aplicado sobre o *Longissimus dorsi* (contra filé) de cada um dos 50 touros Angus e 50 touros Nelore. A Tabela 1 apresenta os resultados dos animais selecionados pelo resíduo de um modelo linear envolvendo raça, data de abate e idade do animal.

Tabela 1: Valores da força de cisalhamento Warner-Bratzler (WBS) das amostras selecionadas para a análise de expressão diferencial. Valores de resíduo negativo representam os animais com valores mais macios para a característica da maciez e resíduo com valores positivos representam os animais mais resistentes. ANAN representa os dados de Angus e NENE os de Nelore.

Identificação	WBS	Nascimento	Raça	Data do abate	Data laboratório	Idade (dias)	Resíduo
animal1	3,49	10/26/2007	ANAN	1/12/2010	10/11/2010	809	-3,03
animal2	3,38	9/7/2006	ANAN	2/4/2009	18/01/2010	881	-2,23
animal3	4,5	11/10/2008	ANAN	10/28/2010	14/12/2010	717	-1,69
animal4	4,06	10/26/2006	ANAN	2/4/2009	14/01/2010	832	-1,47
animal5	3,39	9/7/2006	ANAN	12/9/2008	04/01/2010	824	-1,38
animal6	7,59	11/10/2006	ANAN	2/4/2009	18/01/2010	817	1,42
animal7	8,08	10/31/2008	ANAN	10/28/2010	14/12/2010	727	1,98
animal8	7,5	11/14/2008	ANAN	10/13/2010	01/12/2010	698	2,08
animal9	9,43	10/21/2007	ANAN	1/12/2010	10/11/2010	814	2,95
animal10	9,29	10/20/2008	ANAN	10/28/2010	14/12/2010	738	3,28
animal11	4,55	11/2/2006	NENE	12/9/2008	04/01/2010	768	-2,45
animal12	6,45	11/14/2008	NENE	10/28/2010	15/12/2010	713	-1,50
animal13	6,75	1/2/2009	NENE	10/28/2010	15/12/2010	664	-1,29
animal14	7,27	11/29/2007	NENE	1/12/2010	18/11/2010	775	-1,28
animal15	9,41	10/5/2007	NENE	1/12/2010	10/11/2010	830	1,35
animal16	9,99	11/13/2007	NENE	1/12/2010	18/11/2010	791	1,84
animal17	9,21	10/3/2008	NENE	10/28/2010	15/12/2010	755	1,94
animal18	9,46	11/16/2007	NENE	11/10/2009	09/10/2010	725	2,08

5.2 Sequenciamento das amostras

Foram gerados cerca de 313 milhões de fragmentos de 100 pares de bases (pb), distribuídos nas 18 bibliotecas do experimento, sendo 164 milhões provenientes de dados de Angus e 149 milhões de Nelore.

Na etapa final de preparação das bibliotecas, na qual as amostras são combinadas para integrar uma *lane*, pode-se observar que, pelo menos uma amostra de cada grupo estava presente em cada uma das *lanes*, conforme mostrado na Tabela 2. Esta providência possibilita minimização de viés que porventura possa ocorrer durante processo de sequenciamento, embora isto tenda a ocorrer quando são utilizadas diferentes *flowcells* (ROSS et al., 2013).

Tabela 2: Identificadores das sequências, mostrando o posicionamento das amostras no sequenciamento. ANAN Negativo: Angus classificado como carne macia, ANAN Positivo: Angus classificado como carne mais resistente, NENE Negativo: Nelore classificado como carne macia, NENE Positivo: Nelore classificado como carne resistente.

Amostra	Animal	Lane
animal2A_R1.fastq	ANAN Negativo	4
animal2A_R2.fastq	ANAN Negativo	4
animal5A_R1.fastq	ANAN Negativo	4
animal5A_R2.fastq	ANAN Negativo	4
animal6A_R1.fastq	ANAN Positivo	4
animal6A_R2.fastq	ANAN Positivo	4
animal13A_R1.fastq	NENE Negativo	4
animal13A_R2.fastq	NENE Negativo	4
animal14A_R1.fastq	NENE Negativo	4
animal14A_R2.fastq	NENE Negativo	4
animal16A_R1.fastq	NENE Positivo	4
animal16A_R2.fastq	NENE Positivo	4
animal4A_R1.fastq	ANAN Negativo	5
animal4A_R2.fastq	ANAN Negativo	5
animal7C_R1.fastq	ANAN Positivo	5
animal7C_R2.fastq	ANAN Positivo	5

animal9C_R1.fastq	ANAN Positivo	5
animal9C_R2.fastq	ANAN Positivo	5
animal11A_R1.fastq	NENE Negativo	5
animal11A_R2.fastq	NENE Negativo	5
animal15A_R1.fastq	NENE Positivo	5
animal15A_R2.fastq	NENE Positivo	5
animal18A_R1.fastq	NENE Positivo	5
animal18A_R2.fastq	NENE Positivo	5
animal1A_R1.fastq	ANAN Negativo	6
animal1A_R2.fastq	ANAN Negativo	6
animal3A_R1.fastq	ANAN Negativo	6
animal3A_R2.fastq	ANAN Negativo	6
animal8A_R1.fastq	ANAN Positivo	6
animal8A_R2.fastq	ANAN Positivo	6
animal10A_R1.fastq	ANAN Positivo	6
animal10A_R2.fastq	ANAN Positivo	6
animal12A_R1.fastq	NENE Negativo	6
animal12A_R2.fastq	NENE Negativo	6
animal17A_R1.fastq	NENE Positivo	6
animal17A_R2.fastq	NENE Positivo	6

Em Angus, obtivemos um total de 164.398.376 pares de sequências, cuja média foi 16.439.837,6 e desvio padrão 5.819.248,66. Em Nelore, obtivemos um total de 149.274.757 pares de sequências, cuja média foi 18.659.344,63 e desvio padrão 5.286.338,57. A quantidade de fragmentos gerados para cada biblioteca sequenciada está representada nas Tabelas 3 e 4.

Tabela 3: Quantidade de sequências para cada biblioteca das amostras de Angus.

Amostra	Total de sequências
1 ANAN_NEG_animal1	22.044.447
2 ANAN_NEG_animal2	23.518.565
3 ANAN_NEG_animal3	10.246.614
4 ANAN_NEG_animal4	14.611.980
5 ANAN_NEG_animal5	5.954.478
6 ANAN_POS_animal6	17.186.639
7 ANAN_POS_animal7	21.033.973
8 ANAN_POS_animal8	20.625.497
9 ANAN_POS_animal9	18.163.812
10 ANAN_POS_animal10	11.012.371

Tabela 4: Quantidade de sequências para cada biblioteca das amostras de Nelore.

Amostra	Total de sequências
1 NENE_NEG_animal11	19.217.063
2 NENE_NEG_animal12	19.987.160
3 NENE_NEG_animal13	18.011.659
4 NENE_NEG_animal14	20.162.428
5 NENE_POS_animal15	19.011.541
6 NENE_POS_animal16	26.794.188
7 NENE_POS_animal17	7.498.678
8 NENE_POS_animal18	18.592.040

5.3 Controle de qualidade

Utilizando a ferramenta FASTQC, pudemos visualizar globalmente a composição de cada uma das bibliotecas do experimento e comparar com os dados da literatura, conforme ilustrado na Figura 1. Muitos trabalhos (AMARAL et al., 2014; DEL FABBRO et al., 2013; LI; RINALDI; CAPUCO, 2011; NOOKAEW et al., 2012; RIESGO et al., 2012) relatam que o ideal é eliminar as pontas das sequências para poder retirar as bases com qualidade ruim ou mesmo retirar a sequência inteira caso apresente baixa qualidade em toda sua extensão ou tamanho muito pequeno.

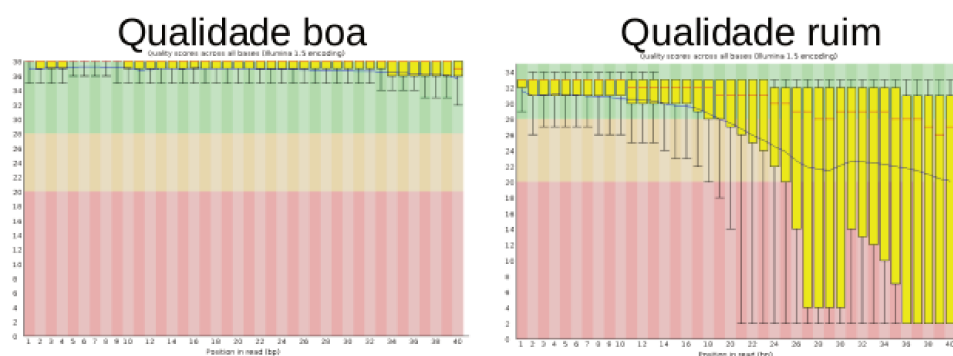


Figura 1: Imagem ilustrativa do relatório do FASTQC, exemplificando uma sequência de boa qualidade (esquerda) e uma de má qualidade (direita).

Contudo, não foram encontrados na literatura estudos que comparem os dados filtrados e não filtrados até a análise funcional dos genes. Além disso, conforme informações de cada ferramenta, elas possuem parâmetros adaptados para trabalhar com sequências que contenham baixa qualidade, tomando as ações necessárias conforme cada situação.

Dessa forma, as análises seguintes foram efetuadas utilizando-se sequências filtradas e originais, separadamente, até a análise de vias enriquecidas dos genes diferencialmente expressos.

A Figura 2 exemplifica o resultado do filtro de qualidade utilizando o FASTQ Quality Trimmer.

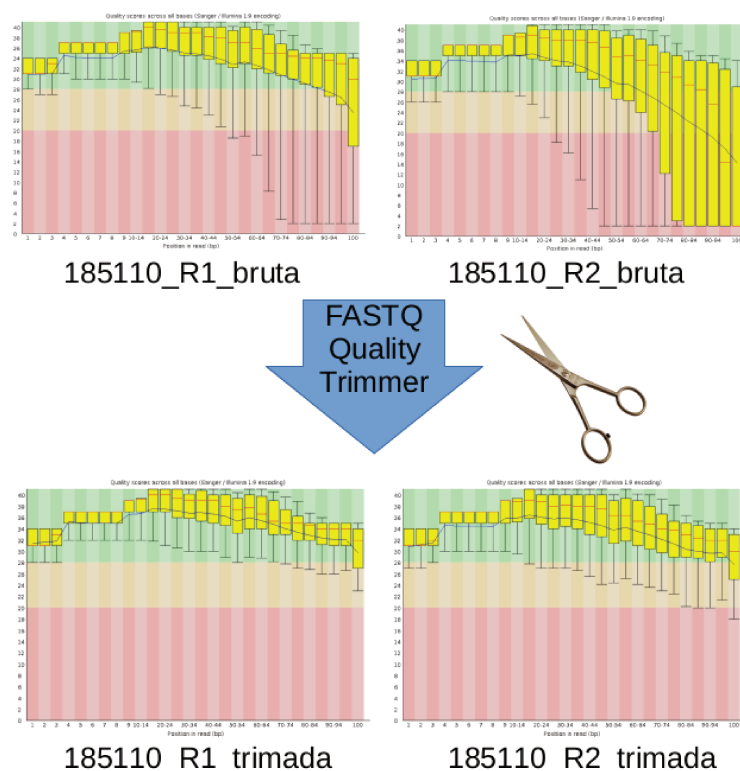


Figura 2: Imagem ilustrativa do relatório geral do FASTQC da qualidade por base, comparando as mesmas amostras antes (superior) e após (inferior) trimagem.

Assim como mostrado na Figura 2, as outras bibliotecas também tiveram suas sequências filtradas, as quais passaram a conter apenas bases com qualidade superior a Q18.

As Tabelas 5 e 6 apresentam a quantidade de sequências que foram totalmente eliminadas pelo filtro de qualidade, apresentando tamanho igual a 0.

Tabela 5: Quantidade e porcentagem das sequências trimadas de Angus de tamanho 0 após a análise de qualidade.

Biblioteca	Fragmento	Total de sequências	Total de sequências com tamanho 0	Porcentagem
ANAN_NEG_animal1	R1	22.044.447	359.082	1,63
ANAN_NEG_animal1	R2	22.044.447	213.705	0,97
ANAN_NEG_animal2	R1	23.518.565	581.778	2,47
ANAN_NEG_animal2	R2	23.518.565	388.831	1,65
ANAN_NEG_animal3	R1	10.246.614	162.783	1,59
ANAN_NEG_animal3	R2	10.246.614	113.579	1,11
ANAN_NEG_animal4	R1	14.611.980	340.661	2,33
ANAN_NEG_animal4	R2	14.611.980	229.662	1,57
ANAN_NEG_animal5	R1	5.954.478	125.749	2,11
ANAN_NEG_animal5	R2	5.954.478	76.010	1,28
ANAN_POS_animal6	R1	17.186.639	393.673	2,29
ANAN_POS_animal6	R2	17.186.639	243.973	1,42
ANAN_POS_animal7	R1	21.033.973	531.853	2,53
ANAN_POS_animal7	R2	21.033.973	384.911	1,83
ANAN_POS_animal8	R1	20.625.497	297.796	1,44
ANAN_POS_animal8	R2	20.625.497	174.870	0,85
ANAN_POS_animal9	R1	18.163.812	420.889	2,32
ANAN_POS_animal9	R2	18.163.812	292.406	1,61
ANAN_POS_animal10	R1	11.012.371	166.932	1,52
ANAN_POS_animal10	R2	11.012.371	104.669	0,95
Total		328.796.752	5.603.812	(média)1,67

Tabela 6: Quantidade e porcentagem das sequências trimadas de Nelore de tamanho 0 após a análise de qualidade.

Biblioteca	Fragmento	Total de sequências	Total de sequências com tamanho 0	Porcentagem
NENE_NEG_animal11	R1	19.217.063	448.500	2,33
NENE_NEG_animal11	R2	19.217.063	293.176	1,53
NENE_NEG_animal12	R1	19.987.160	343.748	1,72
NENE_NEG_animal12	R2	19.987.160	186.703	0,93
NENE_NEG_animal13	R1	18.011.659	425.652	2,36
NENE_NEG_animal13	R2	18.011.659	315.820	1,75
NENE_NEG_animal14	R1	20.162.428	484.849	2,4
NENE_NEG_animal14	R2	20.162.428	303.777	1,51
NENE_POS_animal15	R1	19.011.541	456.344	2,4
NENE_POS_animal15	R2	19.011.541	316.887	1,67
NENE_POS_animal16	R1	26.794.188	657.008	2,45
NENE_POS_animal16	R2	26.794.188	410.781	1,53
NENE_POS_animal17	R1	7.498.678	110.142	1,47
NENE_POS_animal17	R2	7.498.678	63.572	0,85
NENE_POS_animal18	R1	18.592.040	440.675	2,37
NENE_POS_animal18	R2	18.592.040	329.846	1,77
TOTAL		298.549.514	5.587.480	(média)1,81

Em média, cerca de 1,74% das sequências das bibliotecas filtradas apresentaram qualidade ruim e foram totalmente retiradas do conjunto de dados.

Após a trimagem das sequências, foi realizado nova verificação da qualidade das bibliotecas e principalmente a verificação do tamanho das sequências, conforme figura 3.

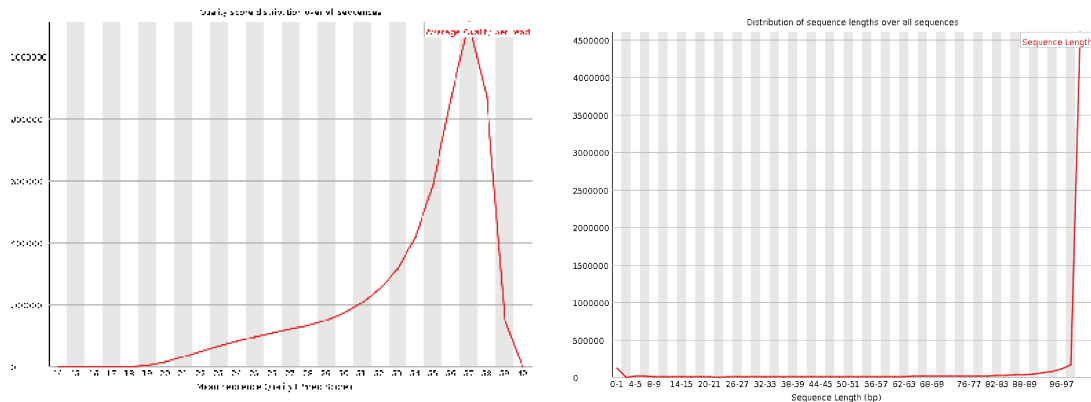


Figura 3: Imagem ilustrativa do relatório geral do FASTQC da qualidade média na biblioteca (esquerda) e do tamanho médio das sequências na biblioteca (direita) após trimagem.

5.4 Mapeamento no genoma

Foram utilizadas as bibliotecas com as sequências trimadas e originais em históricos diferentes no Galaxy. A plataforma Galaxy foi bastante útil para manter a organização e possibilitar a repetição de processos alterando-se apenas os parâmetros necessários.

Como podemos verificar na Tabela 7, as médias dos percentuais das sequências brutas R1 (*forward*) e R2 (*reverse*) de Angus mapeadas no genoma bovino de referência foram de aproximadamente 79% e 69%, respectivamente.

Tabela 7: Resultado do mapeamento de sequências brutas de Angus através do Tophat2 no genoma *Bos taurus* UMD 3.1 Refseq.

Biblioteca	Fragmento	Total de sequências	Total de sequências mapeadas	Porcentagem de mapeadas	Total de sequências únicas	Percentuais de únicas sobre mapeadas
ANAN_NEG_animal1	R1	22.044.447	18.136.259	82,27	16.242.860	89.56
ANAN_NEG_animal1	R2	22.044.447	16.414.532	74,46	15.439.018	94.06
ANAN_NEG_animal2	R1	23.518.565	18.104.386	76,98	15.798.587	87.26
ANAN_NEG_animal2	R2	23.518.565	15.515.522	65,97	14.475.089	93.29
ANAN_NEG_animal3	R1	10.246.614	8.377.028	81,75	7.467.127	89.14
ANAN_NEG_animal3	R2	10.246.614	7.541.022	73,60	7.058.261	93.60
ANAN_NEG_animal4	R1	14.611.980	11.460.591	78,43	10.048.099	87.68
ANAN_NEG_animal4	R2	14.611.980	9.726.542	66,57	9.051.352	93.06
ANAN_NEG_animal5	R1	5.954.478	4.615.303	77,51	4.058.884	87.94
ANAN_NEG_animal5	R2	5.954.478	4.000.079	67,18	3.766.727	94.17
ANAN_POS_animal6	R1	17.186.639	13.341.000	77,62	11.705.332	87.74
ANAN_POS_animal6	R2	17.186.639	11.486.696	66,84	10.768.406	93.75
ANAN_POS_animal7	R1	21.033.973	16.250.708	77,26	14.303.147	88.02
ANAN_POS_animal7	R2	21.033.973	13.748.326	65,36	12.835.818	93.36
ANAN_POS_animal8	R1	20.625.497	17.058.281	82,70	15.316.386	89.79
ANAN_POS_animal8	R2	20.625.497	15.504.604	75,17	14.601.280	94.17
ANAN_POS_animal9	R1	18.163.812	14.262.047	78,52	12.594.838	88.31
ANAN_POS_animal9	R2	18.163.812	12.086.733	66,54	11.361.970	94.00
ANAN_POS_animal10	R1	11.012.371	9.060.826	82,28	8.073.007	89.10
ANAN_POS_animal10	R2	11.012.371	8.172.556	74,21	7.645.331	93.55
TOTAL						
ANGUS_SEQ_ORIG	R1	164.398.376	130.666.429	79.48	115.608.267	88.48
ANGUS_SEQ_ORIG	R2	164.398.376	114.196.612	69.46	107.003.252	93.70

Os percentuais médios de sequências brutas R1 e R2 de Nelore mapeadas no genoma bovino de referência, conforme a Tabela 8, foram de aproximadamente 77% e 66%, respectivamente.

Tabela 8: Resultado do mapeamento de sequências brutas de Nelore através doTophat2 no genoma *Bos taurus* UMD 3.1 Refseq.

Biblioteca	Fragmento	Total de sequências	Total de sequências mapeadas	Porcentagem de mapeadas	Total de sequências únicas	Percentuais de únicas sobre mapeadas
NENE_NEG_animal11	R1	19.217.063	14.737.220	76.69	13.053.120	88.57
NENE_NEG_animal11	R2	19.217.063	12.559.051	65.35	11.795.594	93.92
NENE_NEG_animal12	R1	19.987.160	16.035.203	80.23	14.259.568	88.93
NENE_NEG_animal12	R2	19.987.160	14.512.860	72.61	13.529.798	93.23
NENE_NEG_animal13	R1	18.011.659	13.591.258	75.46	11.830.000	87.04
NENE_NEG_animal13	R2	18.011.659	11.528.204	64.00	10.733.698	93.11
NENE_NEG_animal14	R1	20.162.428	15.373.267	76.25	13.426.208	87.33
NENE_NEG_animal14	R2	20.162.428	13.225.698	65.60	12.348.456	93.37
NENE_POS_animal15	R1	19.011.541	14.331.876	75.39	12.603.030	87.94
NENE_POS_animal15	R2	19.011.541	12.030.943	63.28	11.257.006	93.57
NENE_POS_animal16	R1	26.794.188	20.095.439	75.00	17.514.622	87.16
NENE_POS_animal16	R2	26.794.188	17.219.109	64.26	16.058.302	93.26
NENE_POS_animal17	R1	7.498.678	5.967.671	79.58	5.297.814	88.78
NENE_POS_animal17	R2	7.498.678	5.363.679	71.53	5.001.604	93.25
NENE_POS_animal18	R1	18.592.040	14.235.067	76.57	12.514.852	87.92
NENE_POS_animal18	R2	18.592.040	12.048.491	64.80	11.235.586	93.25
TOTAL						
NELORE_SEQ_ORIG	R1	149.274.757	114.367.001	76.62	100.499.214	87.87
NELORE_SEQ_ORIG	R2	149.274.757	98.488.035	65.98	91.960.044	93.37

Fazendo um levantamento das sequências filtradas, primeiramente de Angus, (Tabela 9) os percentuais médios de sequências trimadas R1 e R2 mapeadas no genoma de referência foram de aproximadamente 90% e 88%, respectivamente.

Tabela 9: Resultado do mapeamento de sequências trimadas de Angus através do Tophat2 no genoma *Bos taurus* UMD 3.1 Refseq.

Biblioteca	Fragmento	Total de sequências	Total de sequências mapeadas	Porcentagem de mapeadas	Total de sequências únicas	Percentuais de únicas sobre mapeadas
ANAN_NEG_animal1	R1	22.044.447	20.108.762	91.22	18.363.989	91.32
ANAN_NEG_animal1	R2	22.044.447	19.723.411	89.47	17.986.038	91.19
ANAN_NEG_animal2	R1	23.518.565	20.921.611	88.96	18.859.443	90.14
ANAN_NEG_animal2	R2	23.518.565	20.409.672	86.78	18.354.788	89.93
ANAN_NEG_animal3	R1	10.246.614	9.291.319	90.68	8.398.604	90.39
ANAN_NEG_animal3	R2	10.246.614	9.070.991	88.53	8.183.663	90.22
ANAN_NEG_animal4	R1	14.611.980	13.083.836	89.54	11.644.060	89.00
ANAN_NEG_animal4	R2	14.611.980	12.639.750	86.50	11.212.814	88.71
ANAN_NEG_animal5	R1	5.954.478	5.332.621	89.56	4.847.017	90.89
ANAN_NEG_animal5	R2	5.954.478	5.210.771	87.51	4.726.390	90.70
ANAN_POS_animal6	R1	17.186.639	15.407.810	89.65	13.959.799	90.60
ANAN_POS_animal6	R2	17.186.639	15.043.072	87.53	13.599.614	90.40
ANAN_POS_animal7	R1	21.033.973	18.590.172	88.38	16.726.261	89.97
ANAN_POS_animal7	R2	21.033.973	17.951.278	85.34	16.100.831	89.69
ANAN_POS_animal8	R1	20.625.497	18.892.201	91.60	17.307.267	91.61
ANAN_POS_animal8	R2	20.625.497	18.543.818	89.91	16.964.601	91.48
ANAN_POS_animal9	R1	18.163.812	16.306.628	89.78	14.751.212	90.46
ANAN_POS_animal9	R2	18.163.812	15.740.291	86.66	14.197.824	90.20
ANAN_POS_animal10	R1	11.012.371	10.050.714	91.27	9.079.833	90.34
ANAN_POS_animal10	R2	11.012.371	9.847.610	89.42	8.881.244	90.19
TOTAL						
ANGUS_SEQ_TRIM	R1	164.398.376	147.985.674	90.02	133.937.485	90.51
ANGUS_SEQ_TRIM	R2	164.398.376	144.180.664	87.70	130.207.807	90.31

Em Nelore, conforme descrito na Tabela 10, os percentuais médios das sequências trimadas R1 e R2 mapeadas no genoma de referência foram de aproximadamente 88% e 86%, respectivamente.

Tabela 10: Resultado do mapeamento de sequências trimadas de Nelore através do Tophat2 no genoma *Bos taurus* UMD 3.1 Refseq.

Biblioteca	Fragmento	Total de sequências	Total de sequências mapeadas	Porcentagem de mapeadas	Total de sequências únicas	Porcentagem de únicas sobre mapeadas
NENE_NEG_animal11	R1	19.217.063	16.927.482	88.09	15.353.417	90.70
NENE_NEG_animal11	R2	19.217.063	16.374.974	85.21	14.809.387	90.44
NENE_NEG_animal13	R1	18.011.659	15.844.650	87.97	14.289.823	90.19
NENE_NEG_animal13	R2	18.011.659	15.350.175	85.22	13.804.656	89.93
NENE_NEG_animal12	R1	19.987.160	17.886.838	89.49	16.230.602	90.74
NENE_NEG_animal12	R2	19.987.160	17.565.135	87.88	15.911.982	90.59
NENE_NEG_animal14	R1	20.162.428	17.849.929	88.53	16.131.963	90.38
NENE_NEG_animal14	R2	20.162.428	17.406.607	86.33	15.693.608	90.16
NENE_POS_animal17	R1	7.498.678	6.671.608	88.97	6.049.703	90.68
NENE_POS_animal17	R2	7.498.678	6.526.083	87.03	5.906.462	90.51
NENE_POS_animal15	R1	19.011.541	16.573.488	87.18	15.054.385	90.83
NENE_POS_animal15	R2	19.011.541	15.920.619	83.74	14.412.577	90.53
NENE_POS_animal16	R1	26.794.188	23.426.929	87.43	21.215.870	90.56
NENE_POS_animal16	R2	26.794.188	22.814.606	85.15	20.610.801	90.34
NENE_POS_animal18	R1	18.592.040	16.354.527	87.97	14.690.329	89.82
NENE_POS_animal18	R2	18.592.040	15.767.807	84.81	14.117.636	89.53
TOTAL						
NELORE_SEQ_TRIM	R1	149.274.757	131.535.451	88.12	119.016.092	90.48
NELORE_SEQ_TRIM	R2	149.274.757	127.726.006	85.56	115.267.109	90.25

Pode-se observar que houve uma considerável diferença no número de sequências mapeadas quando comparamos os grupos de dados de sequências originais (Tabelas 7 e 8) e trimadas (Tabelas 9 e 10). Não apenas a quantidade alinhada no genoma, mas a quantidade de sequências que alinharam em apenas uma posição, chamadas de sequências únicas. Essas sequências (únicas) mapeadas mantiveram praticamente a mesma porcentagem em todos os mapeamentos (90% do total das sequências alinhadas no genoma) e foram subsequentemente utilizadas para se produzir a matriz de contagem.

As sequências que não foram alinhadas ao genoma não foram armazenadas no mesmo arquivo de mapeamento e por isso não passaram pela construção da matriz de

contagens.

5.5 Matriz de contagens

Após gerar um arquivo de mapeamento para cada biblioteca, foram construídas as matrizes com as contagens do número de sequências mapeadas em genes. O HTSeq-count possui parâmetros que auxiliam na filtragem de mapeamentos com qualidades baixas, excluindo aqueles que receberam pontuação menor que 10. Mapeamentos com pontuação superior a 10 garantem que as sequências não foram mapeadas em mais de uma posição do genoma (multimapeadas). Nas Tabelas 11, 12, 13 e 14 podemos verificar a diferença na quantidade de sequências multimapeadas entre as bibliotecas trimadas e originais.

**Tabela 11: Matriz de contagem produzida pelo HTSeq-count.
Dados de Angus com sequências originais.**

Biblioteca	Sequências sem sobreposição em éxons	Sequências ambíguas	Sequências Multimapeadas
ANAN_NEG_animal1	1.693.058	3.491.368	3.390.676
ANAN_NEG_animal2	1.421.806	3.429.943	3.625.800
ANAN_NEG_animal3	773.814	1.592.896	1.834.602
ANAN_NEG_animal4	1.003.778	2.075.024	3.023.009
ANAN_NEG_animal5	512.592	939.629	842.266
ANAN_POS_animal6	1.057.787	2.423.718	2.461.067
ANAN_POS_animal7	1.512.806	3.175.386	3.550.309
ANAN_POS_animal8	1.561.353	3.596.956	3.111.206
ANAN_POS_animal9	1.608.193	2.508.433	2.988.520
ANAN_POS_animal10	761.522	1.674.226	1.987.428
TOTAL			
ANGUS_SEQ_ORIG	11.906.709	24.907.579	26.814.883

**Tabela 12: Matriz de contagem produzida pelo HTSeq-count.
Dados de Nelore com sequências originais.**

Biblioteca	Sequências sem sobreposição em êxons	Sequências ambíguas	Sequências Multimapeadas
NENE_NEG_animal11	1.601.947	3.063.669	2.841.556
NENE_NEG_animal12	1.709.075	3.492.470	3.667.587
NENE_NEG_animal13	1.197.490	2.653.072	2.960.299
NENE_NEG_animal14	1.442.853	3.189.496	3.189.872
NENE_POS_animal15	1.458.453	2.900.268	2.877.701
NENE_POS_animal16	2.168.353	3.531.945	4.077.236
NENE_POS_animal17	548.173	1.101.088	1.325.068
NENE_POS_animal18	1.304.579	2.783.312	3.446.820
TOTAL			
NELORE_SEQ_ORIG	11.430.923	22.715.320	24.386.139

**Tabela 13: Matriz de contagem produzida pelo HTSeq-count.
Dados de Angus com sequências trimadas.**

Biblioteca	Sequências sem sobreposição em êxons	Sequências ambíguas	Sequências Multimapeadas
ANAN_NEG_animal1	2.810.547	6.771.045	16.300.769
ANAN_NEG_animal2	2.178.849	6.685.192	19.922.997
ANAN_NEG_animal3	1.276.543	3.074.747	8.599.148
ANAN_NEG_animal4	1.610.793	3.996.824	16.722.588
ANAN_NEG_animal5	869.666	1.831.992	5.065.951
ANAN_POS_animal6	1.614.457	4.740.333	14.285.520
ANAN_POS_animal7	2.571.771	6.980.337	15.108.107
ANAN_POS_animal8	2.485.393	6.131.986	19.401.666
ANAN_POS_animal9	2.683.454	4.839.222	17.753.381
ANAN_POS_animal10	1.179.035	3.256.464	9.586.233
TOTAL			
ANGUS_SEQ_TRIM	19.280.508	48.308.142	142.746.360

**Tabela 14: Matriz de contagem produzida pelo HTSeq-count.
Dados de Nelore com sequências trimadas.**

Biblioteca	Sequências sem sobreposição em êxons	Sequências ambíguas	Sequências Multimapeadas
NENE_NEG_animal11	2.741.627	5.918.633	16.663.469
NENE_NEG_animal12	2.402.395	6.869.711	16.122.525
NENE_NEG_animal13	1.930.492	5.139.560	15.553.228
NENE_NEG_animal14	2.300.373	6.204.086	17.369.022
NENE_POS_animal15	2.376.737	5.367.923	16.110.833
NENE_POS_animal16	3.622.731	6.682.288	21.953.051
NENE_POS_animal17	884.774	2.064.459	5.846.157
NENE_POS_animal18	2.190.291	5.340.821	18.853.510
TOTAL			
NELORE_SEQ_TRIM	18.449.420	43.587.481	128.471.795

A Figura 4 resume os resultados descritos nas Tabelas 11 a 14, permitindo uma comparação entre os valores totais de sequências de cada grupo.

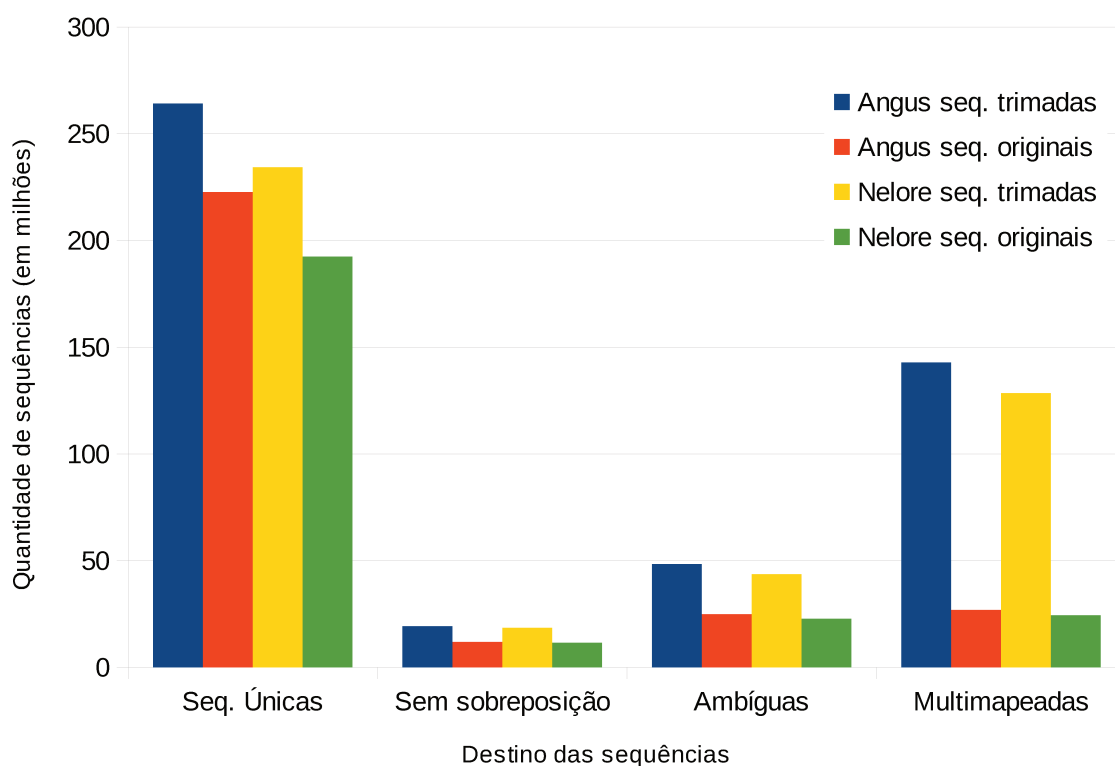


Figura 4: Gráfico comparativo entre os grupos de sequências.

Pela Figura 4 verificou-se, que embora as sequências filtradas tenham obtido maior alinhamento inicialmente (ilustrado em sequências únicas), também tiveram a maior quantidade de sequências eliminadas, conforme valores ilustrados em: sem sobreposição, ambíguas e multimapadas.

5.6 Expressão diferencial de genes

Na análise de expressão diferencial entre os extremos de maciez da carne, realizada através da pacote EdgeR, a partir do conjunto de sequências originais foram identificados 205 genes diferencialmente expressos em Angus e 327 em Nelore. Já a partir do conjunto de sequências filtradas, foram identificados 292 genes diferencialmente expressos em Angus e 435 em Nelore.

A interseção dos genes diferencialmente expressos entre os extremos de maciez de cada raça, identificados a partir de sequências originais e filtradas é exibida através do diagrama de Venn, conforme Figuras 5 e 6, respectivamente.

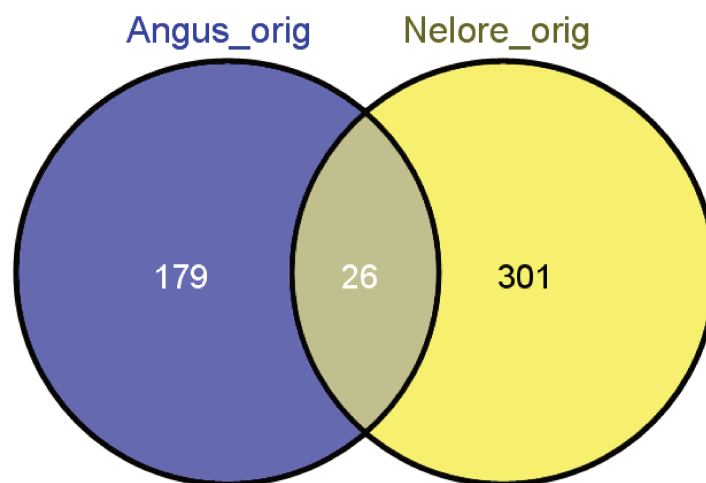


Figura 5: Diagrama de Venn mostrando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne nas raças Angus e Nelore, utilizando as sequências originais.

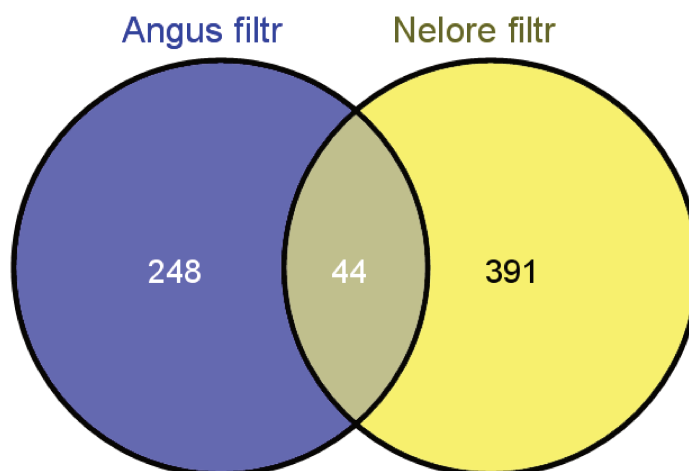


Figura 6: Diagrama de Venn mostrando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne nas raças Angus e Nelore, utilizando as sequências trimadas.

A interseção dos resultados da análise de expressão diferencial entre extremos de maciez da carne, entre os grupos de sequências originais e filtradas, para cada raça, são representados nas Figuras 7 e 8.

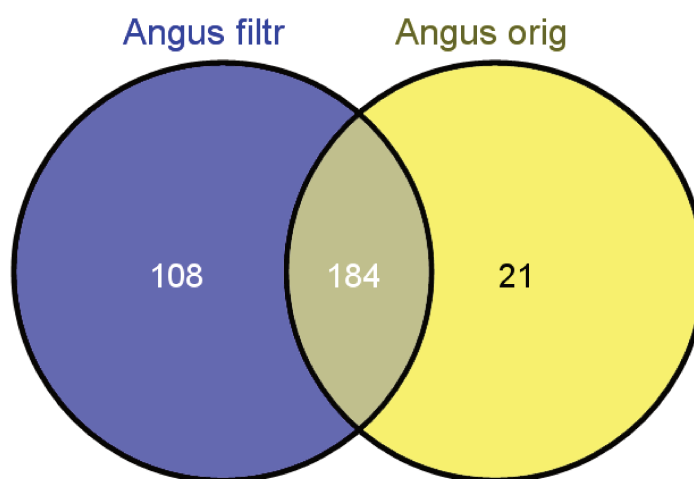


Figura 7: Diagrama de Venn apresentando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne da raça Angus, entre sequências originais e filtradas.

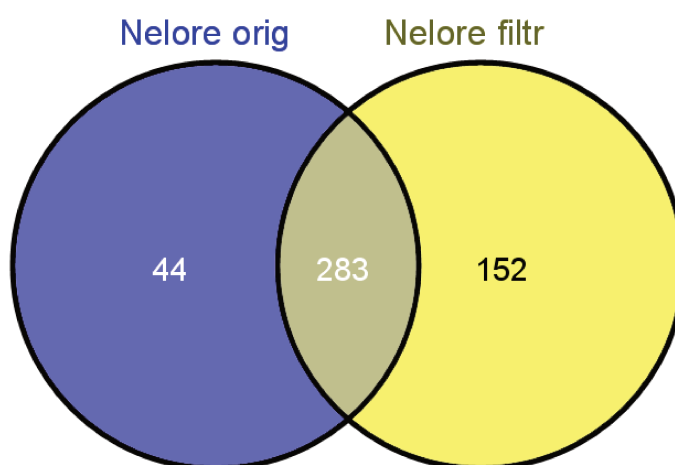


Figura 8: Diagrama de Venn apresentando a interseção dos genes diferencialmente expressos entre extremos de maciez da carne da raça Nelore, entre sequências originais e filtradas.

Do total de genes identificados para cada análise de expressão diferencial, no grupo sequências originais, 64,9% (133) genes em Angus receberam anotação e, em Nelore, 56% (183 genes). Com sequências trimadas, 58% (170) dos genes em Angus receberam anotação e 46,9% (204 genes) em Nelore. A tabela 15 apresenta os valores totais de GDEs com base no critério de anotação e também exibe os a quantidade de

genes que foram identificadas para cada grupo .

Tabela 15: Anotação de genes diferencialmente expressos. Classificação de genes expressos em grupo macio ou duro com base no *Fold change*.

Grupo	GDE	Sem anotação				Com anotação			
		Duro	Macio	Total	%	Duro	Macio	Total	%
Angus_orig	205	29	43	72	35,1	50	83	133	64,9
Angus_trim	292	51	71	122	41,8	63	107	170	58,2
Nelore_orig	327	87	57	144	44,0	118	65	183	56,0
Nelore_trim	435	137	95	231	53,1	130	74	204	46,9

5.7 Identificação de vias enriquecidas

Dos 133 genes diferencialmente expressos entre os extremos de maciez da carne, identificados e anotados, em Angus, utilizando sequências originais, 56,4% (75 genes) foram identificados no banco de dados do *Ingenuity Pathway Analysis* (IPA) e 71% (130 genes) dos 183 genes em Nelore. Nos dados com sequências trimadas, 60,6% (103 genes) e 71,6% (146 genes) dos genes em Angus e Nelore, respectivamente, tiveram identificação no IPA. A Tabela 16 sumariza os resultados de identificação dos genes diferencialmente expressos por grupo enviados à plataforma do IPA.

Tabela 16: Identificação da lista de genes enviadas ao IPA. Foram enviadas listas distintas para cada grupo.

Grupo	Com anotação			IPA			
	Duro	Macio	Total	Upregulated	Downregulated	Total	%
Angus_orig	50	83	133	23	52	75	56,4
Angus_trim	63	107	170	41	62	103	60,6
Nelore_orig	118	65	183	87	43	130	71,0
Nelore_trim	130	74	204	95	51	146	71,6

Dos 75 genes encontrados em Angus, 23 genes tiveram aumento na expressão

(up-regulated) para Angus sequências originais. A Tabela 17 contém as vias enriquecidas, ordenadas pela pontuação de acordo com o enriquecimento funcional, onde os GDEs regulados positivamente para Angus (sequências originais) aparecem em negrito.

Tabela 17: Análise de enriquecimento funcional, a partir dos genes regulados positivamente em Angus (sequências originais). Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	BCAP29, CACNG7 , CCDC186, CFDP1, CMTM6, CNOT11, DPP6, FAIM2 , FAM63B, FAS, GRIA1, KBTBD4 , KCNH7, KCNK1, NAALAD2 , NOVA1 , Pde4d, PGLYRP1 , PHF12, PMP2 , PRND , PURG, RAB23, RNY3, RORA, SLC17A6, SNAP47, SPATA5, TENM2, TNF, TROVE2 , TSPAN5, UBC, ZC3H7A, ZDHHC1	22	9	Metabolismo de lipídeo, Transporte molecular, Bioquímica de pequenas moléculas
2	Akt, API5, BPIFA1, CAMP , CCL7, chondroitin sulfate B, chondroitin sulfate E, CXCL1, CXCR2, ELAVL1, ERG, ERK1/2, ethylene glycol, FAM124B , FPR2, GSR, HAMP, KIAA2013 , LBP, LTB4R, MUC2, MYBPH , NEO1, NFkB (complex), niacinamide, NOS2 , NRG3, P38 MAPK, PRTN3, SAA1 , SDAD1, SPTBN5 , TEX14 , TGIF1 , TRIP6	22	9	Sinalização célula a célula e interação, Sistema de desenvolvimento hematológico e funcional, Tráfico de células imunes
3	MYBPC2, PCDH12	3	1	Sinalização célula a célula e interação, Desenvolvimento do sistema nervoso e função, Arritmia cardíaca
4	D-glucose, fructose-6-phosphate, glyceraldehyde-3-phosphate, TKTL1 , transketolase	2	1	Metabolismo de carboidrato, Câncer, Desenvolvimento e função do sistema cardiovascular

A Figura 9 é a representação gráfica da via de metabolismo de lipídeos. Os nove objetos em vermelho representam as moléculas que foram identificadas como sendo GDEs, a partir da lista analisada pelo IPA. Os objetos possuem representação conforme legenda e a variação da intensidade da cor varia com o nível de expressão com base no *Fold change* enviado juntamente com os lds dos genes.

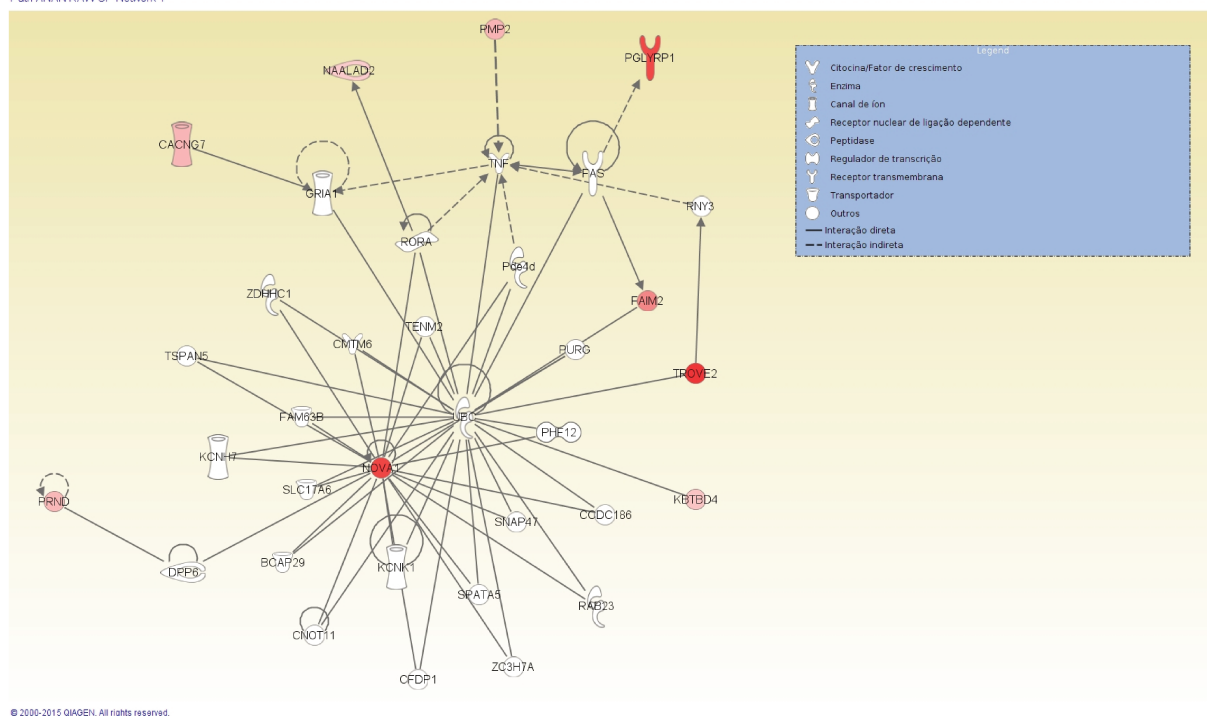


Figura 9: Via do metabolismo de lipídeos. Via enriquecida nos GDEs regulados positivamente em Angus, utilizando-se as sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais resistente.

Ainda nos animais Angus, utilizando sequências originais, os genes que tiveram expressão diminuída foram identificados nas vias enriquecidas mostradas na Tabela 18.

Tabela 18: Análise de enriquecimento funcional a partir dos genes regulados negativamente para Angus (sequências originais). Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	Akt, BCR(complex), C2, C3, CFB, COL11A2, COL22A1 , collagen, Collagen(s), CXCL9, DDIT4, DUOX2, DUOXA2 , ERK, ERK1/2, Growth hormone, IgG, IgG1, Igg3, IGHG1, IGHM, IGJ , Igm, IL1, Immunoglobulin, Interferon alpha, Notch, NTS , P38 MAPK, p70 S6k, PCBP2 , PDGF BB, POSTN, SMPD3, TTR	38	17	Resposta imune humoral, Resposta inflamatória, Transtorno do sistema endócrino
2	AASDHPPT, AKIRIN2, APEH, APIP, APLP1, CAPZB, CCDC170, CENPT , CENPW, CILP2, CPS1 , CUTC, DHX29, EIF2AK1, FBXO6, FOS, HNF4A, HUNK, LRRC40, MRPL37, NSF, OGFOD1, OTUD1 , PRUNE2, RAB6C/WTH3DI, RAVER1 , SDK1, SEC11A, SLC25A13, STT3B, TBCA, TNMD , TPP2, UBC, Wfdc21	25	12	Conjunto e organização celular, Replicação do DNA, Recombinação e reparo, Degradação de proteína
3	ABCA13 , AIF1L, beta-estradiol, BST1, CDC42EP4, CMA1, COMP, CRTAM , DUSP3, EPO, ERBB2, G6pd2, H2-K2/H2-Q9, H2-M10.1 (includes others), HGFAC, Hla-abc, HLA-B , IFNG, IL22RA2, KIR, KLHL24 , LAGE3, PDLIM4, PITX1 , Pri8a2, PTEN, RFXAP, RNA28S5 , ROBO3, SEC11A, SIM1, SPOCK2 , TACR2, TGFB1, Vnn3	22	11	Ciclo celular, Expressão do gene, Bioquímica de pequenas moléculas
4	AQP7, ARHGEF4, BPIFA1, CAMK2N2, chemokine, CLDN2, epinephrine, FBLN7, FBP1 , GIPR, GTF3A, heparin, IL3RA, Insulin, Irx5 , Jnk, L-threonine, LNx1 , mir-1, MYLK, MYO3A , NFkB (complex), NRCAM, ORMDL3, OSTN , PI3K (complex), PTGIR, Rcan1 , TIFA, TMEM14C, TNFRSF18, TPO, TYRO3, Vegf, VRK2	15	8	Desenvolvimento de tecido, Metabolismo de carboidratos, Transporte molecular

A Figura 10 representa a via mais enriquecida na lista de GDEs regulados negativamente em Angus, onde os objetos em verde representam os GDEs e variam a intensidade da cor conforme a variação do nível de expressão. Quanto mais intensa a cor verde, maior o nível de expressão com base no *Fold change*.

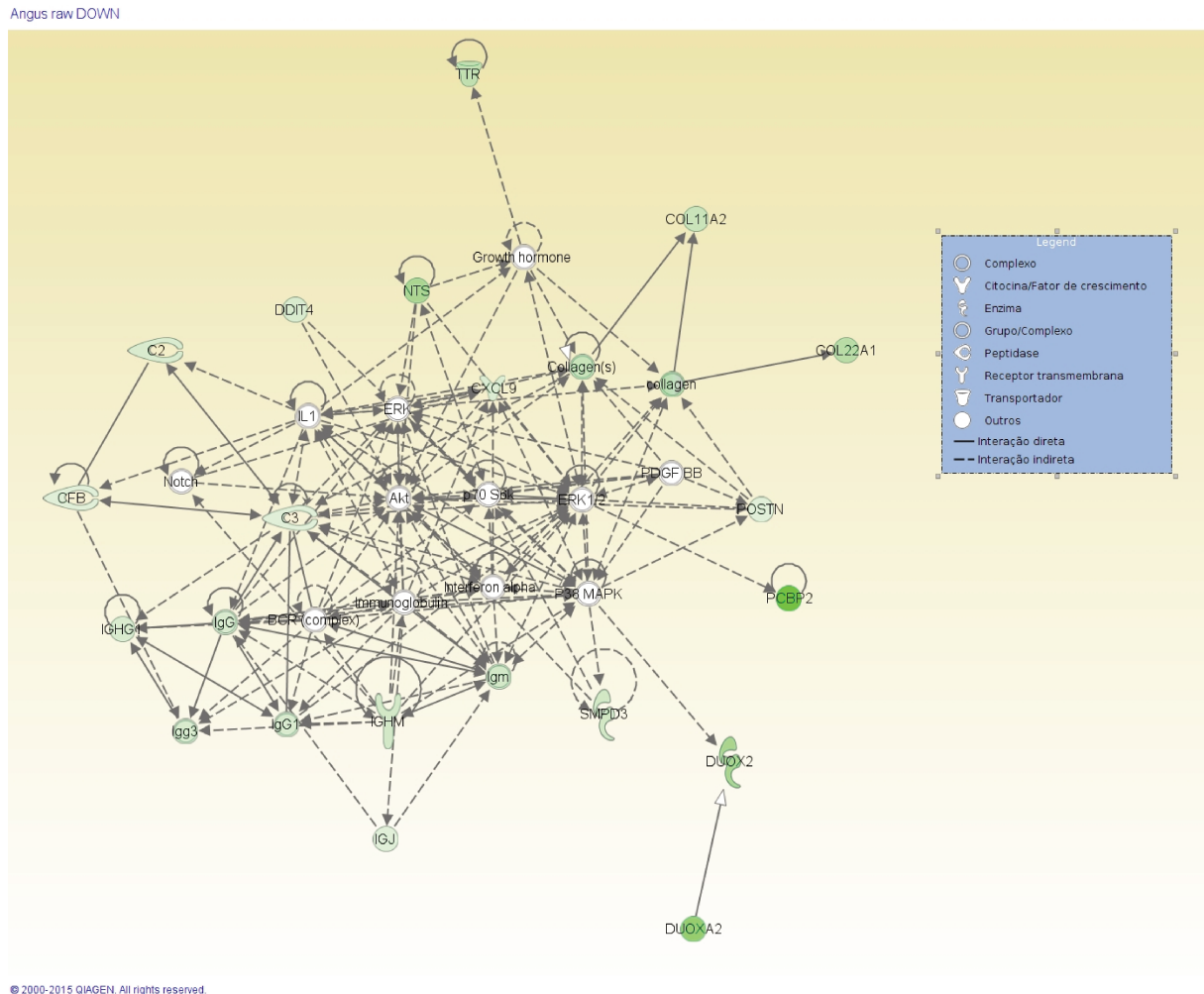


Figura 10: Via de resposta imune, inflamatória. Via enriquecida nos GDEs regulados negativamente em Angus utilizando-se sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais macia.

Em relação ao conjunto de dados de Nelore, utilizando-se as sequências

originais, foram identificados 130 dos 183 GDEs no banco de dados do IPA. Desses, 87 foram regulados positivamente e 43 negativamente. A Tabela 19 representa as principais vias enriquecidas, nas quais os genes superexpressos foram identificados pelo IPA.

Tabela 19: Análise de enriquecimento funcional dos genes regulados positivamente para Nelore (sequências originais). Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções Associadas Às Vias
1	AKR1C3, DDX58, DHX58, EIF2AK2, ERK1/2, Fcer1, GBP5, HEXB, IDO1, IFI6, IFI44, IFIH1, IFIT2, IFIT3, IFIT5, Ifn, IFN alpha/beta, IFN Beta, IFN type 1, Ifnar, Interferon-α Induced, IRF, IRF7, ISG15, ISG20, ISGF3, MX1, MX2, OAS2, PARP12, RTP4, SARAF, ULBP3, XAF1, ZBP1	57	25	Resposta antimicrobiana, Resposta inflamatória, Doenças infecciosas
2	Akt, APOL3 , calpain, CD3, CDH22 , chemokine, EGR3 , estrogen receptor, FAM3B, FGFR2, GBP1, GBP2 , Histone h3, Ifi27, IFI44L , Ifn gamma, IgG, IL12 (complex), IL12 (family), IL6R , Immunoglobulin, INPP4B , Insulin, Interferon alpha, KRT19 , LDL, NFkB (complex), P38 MAPK, PLP1, Pro-inflammatory Cytokine, PTTG1IP, RFX6, RYR2, STOM , Tgf beta	35	17	Distúrbios do sistema endócrino, Desenvolvimento e função do sistema digestivo, Desenvolvimento embrionário
3	AKAP11 , AR, ATP1A4 , CAMSAP1, CEP135, CNN3, EPCAM , EXPH5, FOXP4, IFT22, INPP4B, IQGAP2, KLHL24, MAN1A2, METAP1, MTMR10, MVB12B, MYH8, NCLN, NMO1 (includes others), ODF2L, PDDC1, PHF23, PRSS23, RHBG, RNF213, SIL1, SLFN11, SMAD9, TMEM57, TMEM222, UBC, VPS37A, ZNF83, ZNF557	32	16	Câncer, Ciclo celular, Doença cardiovascular
4	26s Proteasome, AKAP10, APLP1, B4GALNT2, CALN1 , caspase, Cg, CGB (includes others), CHST9 , CREB1, DPEP1, EPSTI1 , ERBB2, FSH, HLA-DQA1, HLA-DQA2 , HLA-DQB3, HNF4A, IFI6 , IKK (complex), Mhc2 Alpha, NOSIP , PARP, PARP14 , Pka, PKIA, RNA polymerase I, RNA28S5 , RUNX1, SLC22A4 , TCR, TNNI3, VWF, XAF1, ZNFX1	30	15	Metabolismo dos aminoácidos, Câncer, Desenvolvimento e função do Sistema cardiovascular
5	ANKS4B , APP, ATCAY, C1QTNF9B, CACNA2D3, COMP , EDN1, GPR1, HARS, HIST2H2BE , HTR6, HTT, hydrogen peroxide, LGR4, LSAMP , MEOX1, NEGR1, NTM, OGDH, OLFM1, OPCML , P2RY4, PAX3 , PCDHGC3, PCTP, PKHD1 , PKIA, PRSS1, RBM25, REEP6 , RLBP1, TBX22, TCN2, TSPY1 (includes others), YWHAZ	18	11	Morte e sobrevivência celular, Doenças neurológicas
6	FKBP4, LOXHD1	2	1	Doenças auditivas, Distúrbio hereditário, Doenças neurológicas

Tabela 20: Análise de enriquecimento funcional dos genes regulados negativamente em Nelore, utilizando as sequências originais. Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	26s Proteasome, ABCC2 , Akt, CCL5 , CD1A , chemokine, Creb, CYP11B2 , EDA2R , ERK, ERK1/2, HDL, HP , IgG, IgG1, IGHG1 , IGHM , IL1, Immunoglobulin, Insulin, Irx5 , KLK4 , LDL, MICAL2 , mir-1 , NFkB (complex), P38 MAPK, PDE10A , PITX1 , Pro-inflammatory Cytokine, PRSS2 , S100A9 , TF, trypsin, Vegf	40	17	Doenças e condições dermatológicas, Doença imunológica, Doença inflamatória
2	ACRBP , AFG3L2, ARAP3, ARFGAP2 , ASAP1 , CAMTA1 , CCNB2, COL19A1 , EMR1 , ENOPH1 , FAM184A , FOS, HIST1H2BN , IPO13, KCNC4 , LHB, MYC, MYNN, NFkB (complex), NMRAL1, NPTX1 , OCM , PHYH, Pkc(s), progesterone, RAB10, RFX2 , RNF41, RNF167 , RPS27, SLC22A18, TPMT, UBC, UCHL1 , YY1	35	15	Desenvolvimento do sistema esquelético e muscular, Desenvolvimento celular, Ciclo celular
3	ACSM3 , beta-estradiol, BMP4, BMP5, BMP7, BMPR1B, CAR ligand-CAR-Retinoic acid-RXR α , CCR1, CHRD12 , DYSF, EGR3 , EHMT2, FSHB, GDF5, hemoglobin, HS3ST2 , HS6ST1, HS6ST3 , ITGA2B, ITGB3, MED13, MYOM1 , Ncoa-Nr1i2-Rxra, Ncoa-Nr1i3-Rxra, NPPB, NR1I2, PCNA, PXR ligand-PXR-Retinoic acid-RXR α , SLC3A2, SMAD2, sulfotransferase, TTN, TWIST1, VAV3	9	5	Crescimento e proliferação celular, Desenvolvimento de tecido, Câncer
4	SPRN , STRN4	3	1	Doença neurológica, Morfologia celular, Organização celular
5	RCHY1, TMEM184A	3	1	Ciclo celular, Crescimento e proliferação celular, Modificação pós transcricional
6	NXPH2 , T3-TR-RXR	3	1	Desenvolvimento e função do sistema nervoso, Câncer, Lesão do organismo e anormalidades

A Figura 12 apresenta as dezessete moléculas identificadas como GDEs na primeira via enriquecida, cuja função está associada a problemas dermatológicos e inflamatórios.

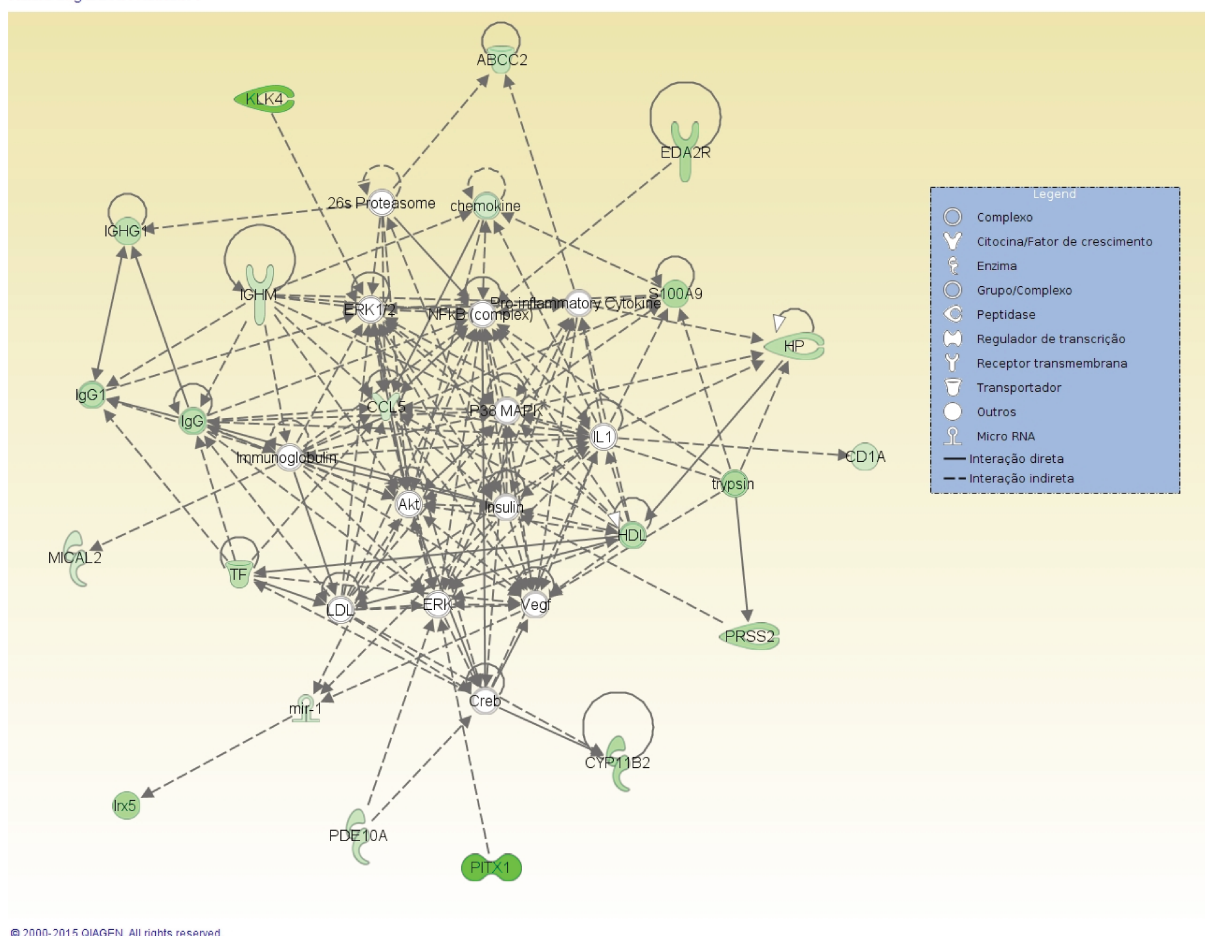


Figura 12: Via relacionada a problemas dermatológicos, imunológicos e inflamatórios. Via enriquecida nos GDEs regulados negativamente em Nelore, utilizando-se as sequências originais. A maior intensidade da cor representa o maior nível de expressão no grupo com carne mais macia.

A segunda via que mostrou-se enriquecida e representada na Figura 13, está envolvida com desenvolvimento muscular esquelético, principal tecido conjuntivo do músculo.

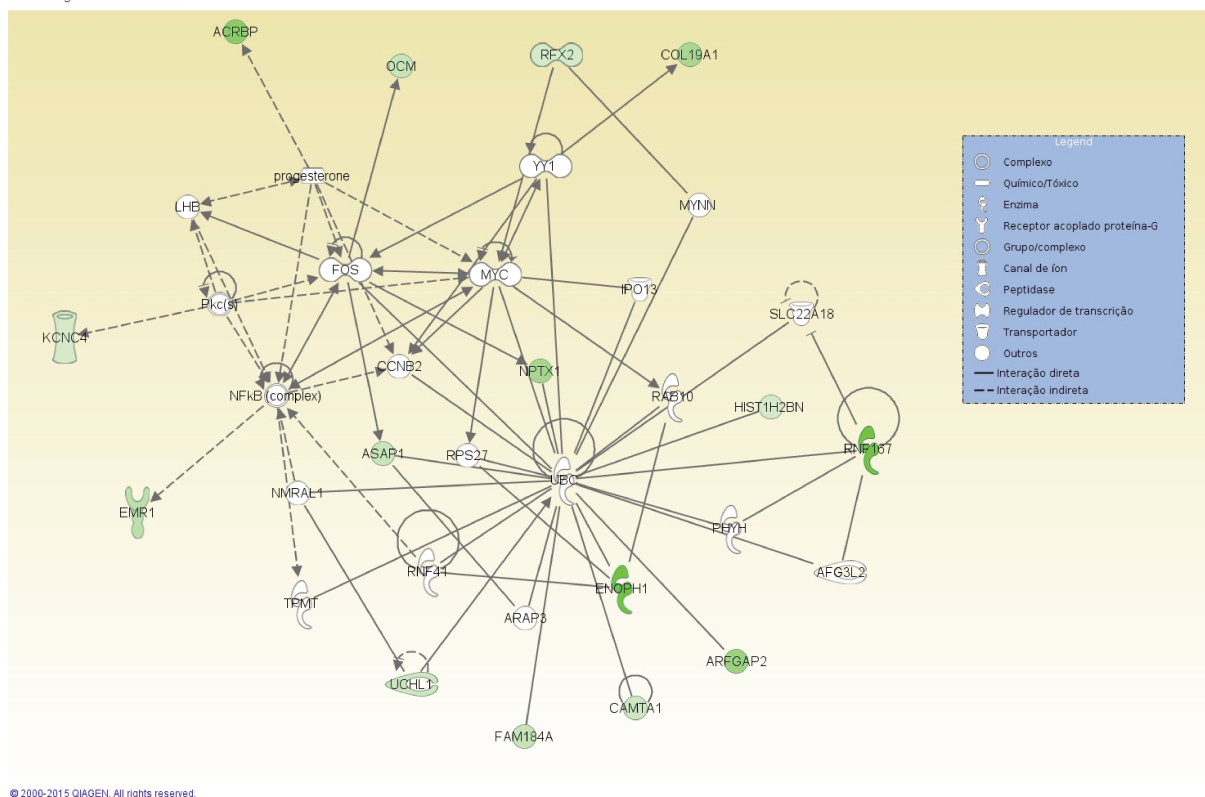


Figura 13: Via de desenvolvimento muscular e esquelético. Via enriquecida nos GDEs regulados negativamente em Nelore. utilizando-se as sequências originais. A maior intensidade da cor representa um maior nível de expressão no grupo com carne mais macia.

A terceira via que está enriquecida é representada pela Figura 14, e está envolvida com proliferação e crescimento celular e desenvolvimento de tecido.

Nos dados com sequências trimadas, dos 170 genes diferencialmente expressos identificados e anotados em Angus, 103 foram identificados no banco de dados do *Ingenuity Pathway Analysis* (IPA). Desses, 41 genes foram regulados positivamente (Tabela 21) e 62 GDEs foram regulados negativamente (Tabela 22).

Tabela 21: Análise de enriquecimento funcional dos genes regulados positivamente em Angus, utilizando as sequências trimadas. Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	Akt, APLNR, BMP8A, CAMP , Cg, CGB (includes others), chondroitin sulfate E, Cyb5r3, CYP11A1 , ERK1/2, ESM1 , ethylene glycol, FAIM2 , HBA1/HBA2 , HBD , HBQ1, Hsp70, Insulin, MYBPH , MYCL, NFkB (complex), NOS2 , NRG3, P38 MAPK, PGLYRP1 , PHEX , Pka, SAA1 , SGK1 , SLC15A2, SLC38A3, SPINK7, TGIF1 , Tnf (family), Vegf	30	13	Morte e sobrevivência celular, Transporte molecular, Desenvolvimento de tecido
2	APP, ARGLU1, BCAP29, Ca2+, CACNG7 , CCDC186, CELF4, CFDP1, CHERP , DLG4, FAM63B, GALNT5 , GSTO1 , HS6ST3 , KCNH7, L-triiodothyronine, NAALAD2 , NOVA1 , PMP2 , PPIL4, PURG, PVALB , RNY3, RORA, SNAP47, SPATA5, SULT1B1 , TENM2, TNF, TROVE2 , TSPAN5, UBC, WNT9B , ZC3H7A, ZDHHC1	26	12	Organização celular, Aumento dos níveis de LDH, Metabolismo de drogas
3	ASTN1 , CBR3, CCAR1, CDKAL1, CELF2, CELSR2, CTSL, CYB5B, D-glucose, FADS1, FAM124B , FCRL6 , HLA-DRB5, KBTBD4 , KCNQ3, KIAA2013 , KIF23, LOC102724428/SIK1, MCF2L2 , NUPR1, PARG, POM121/POM121C, PRND , RAB11FIP5, SIRT5, SRGAP2, STK16, STK17A, TEX14 , TGFB1, TGM5 , TKTL1 , TRERF1, TRIP6, UBC	22	10	Transtorno de desenvolvimento, dano glomerular, Anormalidades e lesão do organismo
4	MYBPC2, PCDH12	3	1	Sinalização celular a célula, Desenvolvimento do sistema nervoso, Arritmia cardíaca
5	NR1D1, OTOP3	3	1	Ciclo celular, Crescimento e proliferação celular, Movimentação celular

Tabela 22: Análise de enriquecimento funcional dos genes regulados negativamente em Angus, utilizando as sequências trimadas. Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	BCR (complex), C2 , C3 , CFB , chemokine, COL11A2 , COL22A1 , collagen, CXCL9 , DUOX2 , DUOX2A , ERK1/2, Growth hormone, Iga, IgG, IgG1, Igg3, Igh (family), IGHE , IGHG1 , IGHM , IGJ , Igm, IL1, Immunoglobulin, Interferon alpha, Notch, P38 MAPK, p70 S6k, PCBP2 , POSTN , Pro-inflammatory Cytokine, SLC11A1 , SMPD3 , Tnf (family)	34	16	Resposta imune humoral, resposta inflamatória, Resposta antimicrobiana
2	ALPL1 , ARHGEF19 , ASNSD1, CCDC138, CILP2 , CPS1 , DNASE2B, ESR1, FBXO6, FOS, HLA-B , LYPLAL1, NSF , NSUN5, NUDT19, OCIAD1, OGFOD1 , OTUD1 , PABPC1L, RGP1, RPL9 , RUNDC1, SDK1, SLC25A13, SNAP47, SOSTDC1 , TBCE, TIGD5, TNMD , UBC, UBD , Wfdc21 , ZNF14, ZNF543, ZNF709	27	13	Bioquímica de pequena molécula, Proliferação e crescimento celular, Desenvolvimento de tecido
3	ABCA13 , ALAS2, APP, beta-estradiol, CCDC170 , CDC42EP4, CENPM , CENPT, COL9A1, COMP , CPNE6, CRTAM , CTNND2, CYBRD1 , DUSP3, EPO, GPX3, IL17A, INPP5K, KLHL24 , MATN1, MATN3, MATN4 , MYO3A , PIK3R5, proprotein convertase, PTEN, SNCB, SP1, SPOCK2 , SYCP2L , TGFBR1, TTR , Vnn3 , WWP1	26	13	Transtorno de tecido conectivo, Transtorno de tecido, Transtorno de desenvolvimento, Transtorno hereditário
4	Akt, BPIFA1, CAMK2N2, Collagen(s), DDIT4 , DRC7 , EPHB3, ERK, FBP1 , FOXO1, GIPR, IL3RA, Insulin, Jnk, LNx1 , MYLK , NFkB (complex), NTS , ORMDL3, PBK, PDGF BB, PI3K (complex), PITX1 , Rcan1 , RFX6, RNF41, RPIA, SCLT1, TIFA, TMEM14C, TNFRSF18, Tsc1-Tsc2, TYRO3, Vegf, VRK2	14	8	Desenvolvimento do tecido conectivo e Função, Desenvolvimento embrionário, Desenvolvimento de órgãos
5	ANGPTL4, APOA5, AQP7, C4BPA , CELF2, CMA1 , CMKLR1, Collagen type ix, EIF5A, FBLN7 , FN1, GUCY2C, Ins1, Irx5 , KCNJ15, KCNQ1, KRT5, LCAT, LCP1, LIFR, MC3R, Mcpt4, mir-1, NDST2, NDUFA1, OSTN , PEA15, PTPN22, RAVER1 , SERPINA5, SIM1 , SLC25A1, SYT9, TAGLN2, TPSAB1/TPSB2	13	7	Metabolismo de carboidrato, Metabolismo de drogas, Transporte celular

Nos dados com sequências trimadas de Nelore, dos 204 genes diferencialmente expressos identificados e anotados, 146 foram identificados no banco de dados do *Ingenuity Pathway Analysis* (IPA). Desses, 95 genes foram regulados positivamente (Tabela 23) e 51 GDEs foram regulados negativamente (Tabela 24).

Tabela 23: Análise de enriquecimento funcional dos genes regulados positivamente em Nelore, utilizando as sequências trimadas. Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	AKR1C3, ASAH1 , Cg, Cyclin E, DDX58, DHX58, EPCAM, ERK1/2, FGFR2, GBP5, IDO1, IFI44, IFIH1, IFIT2, IFIT3 , Ifn, IFN alpha/beta, IFN Beta, IFN type 1, Ifnar, IL6R , Interferon- α Induced, IRF, IRF7, ISG15, ISG20, ISGF3, MX1, MX2, OAS2, PGR, PLP1, RTP4, TMPRSS2, ZBP1	53	24	Resposta antimicrobiana, Resposta inflamatória, Doença dermatológica
2	Akt, AMPK, APOL3, ART1 , calpain, caspase, CD3, chemokine, EIF2AK2 , Fcer1, GBP2, HEXB, IFI6, Ifi27, IFI44L , Ifn gamma, IgG, IL12 (complex), Immunoglobulin, INPP4B , Interferon alpha, PARP, PARP12, PARP14 , PI3K (complex), PRKAB1, PTTG1IP, RYR2, STOM, STYK1 , TCR, Tgf beta, TNNI3, XAF1, ZNFX1	40	19	Sinalização celular, Função e manutenção celular, Bioquímica de pequenas moléculas
3	ABCD4, B4GALNT2 , C16orf70, CASK , CEP72, CNN3 , CNOT3, EIF5B, FBXO11, FOXP4, HHAT, KLHL24 , LGR4, MC2R, MRPS5 , MVB12A, MVB12B, NOSIP , NUP210, PKHD1 , PRSS1, PTBP3, RAVER1 , RBM25, RNF213, SARAF, SLFN11 , SSNA1, TBR1, UBC, USP54, VPS37A, VPS37B, VWF, ZNF143	27	14	Desenvolvimento celular, Transtorno de desenvolvimento, Desenvolvimento embrionário
4	AK4, APBB3, APP, AR, ATCAY, C1QTNF9B, CALN1 , CAMK1D, CAPNS2 , CGB (includes others), CREB1, DDX42, DDX54, EPSTI1 , ERBB2, GBP6, GSTM3, hydrogen peroxide, IFI6 , IL4I1, LAMB3, LYRM7, MFRP , NNMT, norepinephrine, OGDH, P2RX3, PKIA, PNMT, RHBG, RNASE11, SYCP2L, TIMMDC1, TPD52L1, XAF1	23	13	Produção de energia, Movimento celular, Desenvolvimento do sistema esquelético e muscular
5	ANKS4B, APLP1 , APP, ATP1A4 , beta-estradiol, CACNA2D3, CHST9 , CNTN3, COMP , FAM160A2, H2-a, HIST2H2BE , HLA-DQ, HLA-DQA1, HLA-DQA2 , HLA-DQB3, HNF4A, HTT, KIAA0930, KLC4, LRRRC8C, LSAMP , NDUFA7, NEGR1, NTM, OGDH, PDXK, PPP1R3D, REEP6, RTP4 , SERPINB8, SPON1, TNF, TSPY1 (includes others), YWHAZ	22	13	Morfologia celular, Função e manutenção celular, Conjunto e organização celular

Tabela 24: Análise de enriquecimento funcional dos genes regulados negativamente em Nelore, utilizando as sequências trimadas. Os genes diferencialmente expressos (GDEs) são mostrados em negrito.

ID	Moléculas da via	Pontuação	GDEs	Funções associadas às vias
1	ABCC2 , BAI3 , CD1A , CELSR3, Creb, CYP11B2 , EDA2R , EMR1 , ERK, Gpcr, HDL, hemoglobina, HP , IgG, IgG1, Igh-8, IGHG1 , IGHM , Iglv1, IL1, KLRB1 , NFE2 , NFkB (complex), P38 MAPK, PDE10A , PIK3R5, PITX1 , Pro-inflammatory Cytokine, PRSS2 , RGR , S100A9 , TF , trypsin, TTR	44	18	Morte e sobrevivência celular, Transporte molecular, Bioquímica de pequenas moléculas, Cell Death and Survival, Molecular Transport, Small Molecule Biochemistry
2	AFG3L2, ARFGAP2 , CA5A , CAMTA1 , CCDC14, CDH15, CLIP4, COL17A1, DST , ENOPH1 , FAM184A , FRMD5, KIAA1467, KIAA1549, MLH1, NEO1 , NMRAL1, PHYH, POLDIP2, PPFIA2, RASSF3, RCHY1, RNF167 , SELENBP1 , SH3BP5, SLC22A18, TMEM171 , TMEM184A , TPRG1L , TRIM23, TUBA1C , UBC, UBXN6, UCHL1 , USP33	31	14	Transtorno do tecido conectivo, Transtorno dermatológico, Transtorno de desenvolvimento
3	ACRBP , ACSM3 , AEBP1, ASAP1 , beta-estradiol, BMP5, CAR ligand-CAR-Retinoic acid-RXR α , CGA, CHRD2 , COL19A1 , DDR1, E4F1, FOS, FSHB, GDF5, GRAP, GRB2, HS3ST2, HS6ST1 , HS6ST3, LHB, Ncoa-Nr1i2-Rxra, Ncoa-Nr1i3-Rxra, NPTX1 , OCM , progesterone, RBP4, SELENBP1 , SPRY1, SPRY2, SPRY3 , SPRY4, sulfotransferase, TESK1, YY1	23	11	Desenvolvimento do sistema renal e urológico, Renal and Urological System Development and Function, Reproductive System Development and Function, Connective Tissue Development and Function
4	SPRN , STRN4	3	1	Doença neurológica, Morfologia celular, Conjunto e organização celular
5	NXPH2 , T3-TR-RXR	3	1	Desenvolvimento do sistema nervoso, Cancer, Anormalidades e lesão do organismo

Na tabela 25 temos a sumarização dos grupos, no qual as vias com maior pontuação e representatividade de genes diferencialmente expressos encontrados foram listados. Esses dados irão contribuir para a discussão das vias encontradas.

Tabela 25: Lista de vias enriquecidas exibidas pelo IPA que foram selecionadas pelo critério de pontuação para cada grupo de acordo com a regulação dos genes.

Grupo	Regulação	Função encontrada relacionada a maciez
Angus_orig	UP	Metabolismo de lipídeos
	DOWN	Resposta inflamatória (ativação do sistema complemento), resposta imune
Angus_trim	UP	Morte e sobrevivência celular
	DOWN	Resposta inflamatória (ativação do sistema complemento), resposta imune
Nelore_orig	UP	Resposta inflamatória (resposta antiviral)
	DOWN	Desenvolvimento muscular
Nelore_trim	UP	Resposta inflamatória (resposta antiviral)
	DOWN	Desenvolvimento de tecido

6 DISCUSSÃO

Os animais foram abatidos de acordo com o critério de cobertura de gordura, peso e idade, conforme metodologia empregada. Verificou-se que existe diferença nos valores apresentados entre as raças, principalmente entre os valores para os animais considerados macios. Os animais da raça Angus produziram contrafilés com força de cisalhamento (WBS) variando de 3,38 a 9,43 kgf, enquanto os animais da raça Nelore produziram contrafilés com força de cisalhamento variando entre 4,5 e 9,99 kgf. Os valores para Angus tiveram maior variação e no geral foram esses valores foram mais altos que encontrados na literatura (ARBOITTE et al., 2011). Mesmo uma variação menor, no geral, o Nelore apresentou valores para força de cisalhamento maiores que o Angus. Os valores condizem com os resultados de trabalhos publicados (ABULARACH; ROCHA; FELÍCIO, 1998; KOOHMARAIE, 1996) onde foi indicado que o Nelore possui valores de WBS altos logo após o abate e que isso é característica dos animais *Bos taurus indicus*. Adicionalmente a essa informação, Koohmaraie (1996) descreve que é entre as primeiras 24 horas que o animal possui a carne mais rígida, influenciando na força de cisalhamento após o abate. Contudo, observou-se que mesmo os animais Nelore classificados como macios no experimento, na verdade já estão em uma classificação de mais resistente já que segundo a literatura (JOHNSON et al., 1990; KEITH et al., 1985; MILLER et al., 2001), os valores de WBS considerados macio possuem WBS entre 3,0 kgf a 5,5 kgf aproximadamente.

Neste trabalho, a tecnologia de RNA-Seq foi utilizada com o intuito de elucidar os mecanismos moleculares relacionados à maciez da carne a partir de amostras de contrafilé de duas raças bovinas (Angus e Nelore). Essa técnica tem sido amplamente utilizada em análises de transcriptômica propiciando resultados significativos e trazendo um alto rendimento por sequenciamento (LEE et al., 2013, 2013; LI; RINALDI; CAPUCO, 2011; MCCABE et al., 2012; MORTAZAVI et al., 2008; RAMAYO-CALDAS et al., 2012; SONESON; DELORENZI, 2013; ZHANG et al., 2014).

Existem diversas ferramentas de bioinformática disponíveis para cada etapa da

análise de expressão diferencial de genes e cada etapa tem potencial para modificar o resultado final. Desde a primeira etapa da análise, a avaliação de qualidade das sequências até decisões a respeito de como devem ser processadas sequências com baixa qualidade são muito relevantes. Os arquivos de sequências obtidos possuem pontuações que são atribuídas a cada base nucleica de acordo com a qualidade do sequenciamento indicando a possibilidade da base ter sido sequenciada erroneamente. Dessa forma, muitos autores optam por retirar as bases de qualidade duvidosa chegando a eliminar em torno de 7% dos dados originais (BALDWIN et al., 2012; LEE et al., 2013, 2014; LI; RINALDI; CAPUCO, 2011). Com o avanço das tecnologias de sequenciamento e a redução do custo por base sequenciada, tem-se observado um aumento na quantidade de sequências por biblioteca. Contudo, a proporção de bases com baixa qualidade continua a mesma, comumente ultrapassando milhões de sequências. Apesar da decisão de eliminar ou não as bases nucleicas de baixa qualidade ter potencial para alterar a identificação de genes diferencialmente expressos, poucos trabalhos abordam esse assunto.

Pudemos observar na literatura científica que grande parte dos pesquisadores optam por eliminar bases com qualidade inferior a 30 ou um índice menor em trabalhos de identificação de genes diferencialmente expressos (AMARAL et al., 2014; BALDWIN et al., 2012; DEL FABBRO et al., 2013; LEE et al., 2013; NOOKAEW et al., 2012; RIESGO et al., 2012). Contudo, na literatura científica recente, trabalhos como MacManes (2014) recomendam trinar sequências de forma mais estrigente apenas para estudos de genômica de populações para identificação de SNPs. Tentar aproveitar pedaços de sequência com boa qualidade e, conseqüentemente, trabalhar com sequências de tamanhos variados pode também comprometer a identificação de genes diferencialmente expressos, já que a qualidade do mapeamento é diretamente proporcional ao tamanho da sequência. Este aspecto foi verificado neste trabalho através da comparação entre os resultados da análise de expressão diferencial dos genes utilizando sequências que passaram pelo controle de qualidade e sequências originais.

Foram utilizados 5 amostras de réplicas biológicas de bovinos das raças Angus e

4 de bovinos da raça Nelore para cada condição extrema de maciez selecionadas pela força de cisalhamento. Um total de 313 milhões de sequências de 100 pb foram obtidos através do sequenciamento em 3 lanes com Illumina HiScanSQ. Além da análise de expressão diferencial entre os extremos fenotípicos de maciez, pudemos realizar todas as etapas utilizando sequências trimadas e originais e avaliar o impacto nos resultados.

Comparando-se as metodologias utilizando sequências trimadas e originais, inicialmente obtivemos menores quantidades de sequências mapeadas ao genoma de referência ao utilizar sequências originais. Foram em média 74% e 70,5% em Angus e Nelore, respectivamente, utilizando-se sequências originais e, em contrapartida, 88,5% e 86,5% utilizando-se sequências trimadas. Podemos notar que ocorreu uma redução do número total de sequências mapeadas entre médias das bibliotecas de Angus e Nelore tanto em sequências filtradas quanto nas originais. Essa diferença possivelmente está relacionada com o fato de o genoma utilizado como referência ser de um animal da raça *Bos taurus taurus* e possivelmente pelo fato de a raça Nelore aqui utilizada ter passado por várias gerações apresentando assim alguma diferença significativa da referência utilizada. Na construção das matrizes de contagens, foram retiradas muitas sequências multimapeadas, principalmente das sequências que foram trimadas, resultando em um total de 222 milhões de sequências originais de Angus, 192 milhões de sequências originais de Nelore, 264 milhões de sequências trimadas de Angus e 234 milhões de sequências trimadas de Nelore. Este resultado poderia ter correlação com o fato de ao trimar as sequências estas estariam mais suscetíveis a alinhamentos em regiões multigênicas e mais conservadas. Verificou-se que, embora a quantidade de contagens utilizadas para a avaliação de expressão diferencial fossem próximas, os genes identificados foram significativamente diferentes. Com relação às amostras da raça Angus, constatou-se que houve uma interseção de 184 genes de um total de 205 GDEs e 292 GDEs utilizando-se sequências originais e trimadas, respectivamente. Curiosamente verificou-se que 108 genes foram identificados exclusivamente utilizando-se sequências trimadas e apenas 21 ao utilizar-se sequências originais. Em Nelore, dos 327 GDEs e 435 GDEs identificados utilizando-se sequências originais e trimadas, respectivamente, a interseção foi de 283 genes. Sendo

152 específicos da análise utilizando-se sequências trimadas e 44 utilizando-se sequências originais.

A análise de enriquecimento funcional no IPA pode contribuir com a interpretação biológica dos dados. Os resultados da análise no IPA em Angus utilizando-se sequências originais indicaram GDEs constituindo via relacionada ao metabolismo de lipídeos. Essa via também foi identificada em trabalhos com RNA-Seq em bovinos (BALDWIN et al., 2012; LEE et al., 2013) e suínos (RAMAYO-CALDAS et al., 2012).

A variação de maciez da carne está relacionada, entre vários fatores, com o teor de lipídeos presentes na composição da musculatura. Esta composição, também conhecida por marmorização, tende a diluir o tecido conjuntivo na fibra muscular que está contida e difere claramente entre as raças bovinas podendo influenciar o consumidor (CHAMBAZ et al., 2003; TRAN; THU, 2006). Esta marmorização estaria relacionada com a maior quantidade de gordura entremeada na carne, e assim contribuindo para o aumento da maciez (ARBOITTE et al., 2011).

Além da marmorização, outro componente da estrutura que também foi verificado neste trabalho é o colágeno. O colágeno também influencia na maciez da carne não somente por ser um constituinte do tecido muscular, mas por determinar a forma de organização das fibras, podendo resultar em uma carne menos macia conforme a extensão das ligações entre as moléculas de colágeno (WESTON; ROGERS; ALTHEN, 2002). No presente trabalho, foram identificados genes (COL11A2 e COL22A1) participando na regulação de colágeno na via enriquecida de Angus e o gene COL19A1 em Nelore. No trabalho de Duarte *et al.* (2013) também foi verificado a importância do colágeno e gordura intramuscular em bovinos da raça Wagyu e Angus. Ambas as raças são conhecidas pela qualidade da carne, em especial ao marmoreio. Contudo, segundo os resultados em animais Wagyu, menores quantidades de fibras musculares aliado a maiores quantidades de colágeno poderiam ter contribuído para o aumento da marmorização. Weston et al. (2002) sugerem que o conteúdo ou a quantidade de colágeno são menos significativos que a qualidade do colágeno na carne. E Lee et al. (2013) também verificaram expressão de genes de colágeno em tecidos adiposos subcutâneos. Estes dados condizem com os resultados encontrados

neste trabalho para com os genes de colágeno encontrados em Angus e Nelore para o grupo macio.

Por outro lado, vias enriquecidas relacionadas à resposta imune e inflamatória foram identificadas em todos os grupos. Com relação aos resultados envolvendo resposta inflamatória, Cesar (2014) discute que podem existir dois passos no processo de deposição de gordura: o primeiro é o de síntese de lipídeos propriamente dito e o segundo é processo inflamatório. Entretanto, Zhao *et al.* (2012) identificaram vias enriquecidas relacionadas à resposta imune que foram desencadeadas por estresse agudo. Este estresse induzido estaria influenciando diretamente a maciez da carne no experimento com Angus em microarranjo.

Embora Zhao *et al.* (2012) tenham indicado uma relação de resposta imune induzida por fatores externos, Titto *et al.* (2010) indicam que o estresse está relacionado a fatores comportamentais ao afirmar que há grande correlação entre temperamento e produtividade em bovinos em geral e, que o Nelore particularmente possui um comportamento mais agressivo.

Dessa forma, uma possível hipótese no presente trabalho é que a presença de vias enriquecidas relacionadas à resposta imune e inflamatória estejam influenciando o mecanismo de maciez em bovinos das raças Angus e Nelore, porém com

Estudos sugerem que 30% da variação na maciez da carne entre as raças se explica pelo fator genético e 70% da variação pelo fator ambiental. WARNER *et al.* (2010) relatam que tanto o fator genético, como o fator ambiental, interagem para contribuir fenotipicamente na variação da maciez da carne e que certamente outros fatores devem influenciar, como as condições de abate.

Devido à grande importância econômica da criação de bovinos de corte, muitos estudos são realizados para encontrar marcadores genéticos, associados à *loci* de caracteres quantitativos (QTL), que possibilitem a seleção precoce de animais a fim de melhorar tanto em qualidade quanto em quantidade produzida por animal. Dessa forma, este estudo identificou candidatos a genes que possivelmente podem estar relacionados com a maciez da carne e espera contribuir com pesquisas futuras para o desenvolvimento de novos marcadores para essa característica.

7 CONCLUSÃO

Os resultados descritos no presente trabalho confirmam a viabilidade de se utilizar a técnica de RNA-Seq na comparação de populações selecionadas por extremos fenotípicos para força de cisalhamento da carne.

As ferramentas de bioinformática utilizada se mostraram bastante eficazes na análise de sequências originais e trimadas, no qual o resultado com sequências originais em Angus teve maior correlação com o que se é observado na literatura.

Quanto ao objetivo principal do trabalho, pudemos encontrar candidatos a genes diferencialmente expressos nas raças Angus e Nelore, entre as condições de variação fenotípicas, que podem estar correlacionados no processo de maciez da carne.

8 REFERÊNCIAS

ABULARACH, M. L. S.; ROCHA, C. E.; FELÍCIO, P. E. DE. CARACTERÍSTICAS DE QUALIDADE DO CONTRAFILÉ (m. L. dorsi) DE TOUROS JOVENS DA RAÇA NELORE. **Ciência e Tecnologia de Alimentos**, v. 18, n. 2, p. 205–210, maio 1998.

AMARAL, A. J. et al. Quality assessment and control of tissue specific RNA-seq libraries of *Drosophila* transgenic RNAi models. **Frontiers in Genetics**, v. 5, 5 mar. 2014.

ANDERS, S. et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. **Nature protocols**, v. 8, n. 9, p. 1765–86, set. 2013.

ANDERS, S.; PYL, P. T.; HUBER, W. HTSeq A Python framework to work with high-throughput sequencing data. p. 0–5, 20 fev. 2014.

ANDREWS, S., ET AL. **Andrews, S. “FastQC: A quality control tool for high throughput sequence data.” Reference Source (2010).** Disponível em: <http://scholar.google.com.br/scholar?q=related:7Au96aB8tVoJ:scholar.google.com/&hl=pt-BR&as_sdt=0,5>. Acesso em: 13 jan. 2015.

ARBOITTE, M. Z. et al. Qualidade da carne do músculo longissimus dorsi de novilhos superjovens Aberdeen Angus de biótipo pequeno e médio abatidos com o mesmo estágio de acabamento na carcaça. **Acta Scientiarum. Animal Sciences**, v. 33, n. 2, 29 abr. 2011.

BAI, Y.; SARTOR, M.; CAVALCOLI, J. Current status and future perspectives for sequencing livestock genomes. **Journal of animal science and biotechnology**, v. 3, n. 1, p. 8–8, jan. 2012.

BALDWIN, R. L. et al. Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. **Physiological Genomics**, v. 44, n. 10, p. 543–550, 17 maio 2012.

BERNARD, C. et al. New indicators of beef sensory quality revealed by expression of specific genes. **Journal of agricultural and food chemistry**, v. 55, n. 13, p. 5229–37, jun. 2007.

CESAR, A. S. M. **Identification of genes associated with intramuscular fat deposition and composition in Nellore breed.** 2014. 100 f.. Tese Doutorado em Ciência Animal e Pastagens – Escola Superior de Agricultura Luiz de Queiroz – USP, Piracicaba.

CHAMBAZ, A. et al. Meat quality of Angus, Simmental, Charolais and Limousin steers

compared at the same intramuscular fat content. **Meat Science**, v. 63, n. 4, p. 491–500, abr. 2003.

CHITWOOD, J. L. et al. RNA-seq analysis of single bovine blastocysts. **BMC genomics**, v. 14, n. 1, p. 350–350, jan. 2013.

CORVA, P. et al. Association of CAPN1 and CAST gene polymorphisms with meat tenderness in Bos taurus beef cattle from Argentina. v. 1069, p. 1064–1069, 2007.

CROUSE, J. D. et al. Comparisons of Bos Indicus and Bos Taurus Inheritance for Carcass Beef Characteristics and Meat Palatability. n. 67, p. 2661–2668, 1989.

DEL FABBRO, C. et al. An extensive evaluation of read trimming effects on Illumina NGS data analysis. **PloS one**, v. 8, n. 12, p. e85024–e85024, jan. 2013.

DILLIES, M.-A. et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. **Briefings in bioinformatics**, v. 14, n. 6, p. 671–83, nov. 2013.

DUARTE, M. S. et al. Enhancement of adipogenesis and fibrogenesis in skeletal muscle of Wagyu compared with Angus cattle. **Journal of animal science**, v. 91, n. 6, p. 2938–46, jun. 2013.

EWING, B.; GREEN, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1 mar. 1998.

FAO. **Meat Consumption**. Disponível em: <<http://www.fao.org/ag/againfo/themes/en/meat/background.html>>. Acesso em: 01 Fev. 2015.

FERRAZ, J. B. S.; FELÍCIO, P. E. DE. Production systems – An example from Brazil - 1-s2.0-S0309174009001648-main.pdf. **Meat science**, v. 84, n. 2, p. 238–243, 2010.

GIARDINE, B. et al. Galaxy: a platform for interactive large-scale genome analysis. **Genome research**, v. 15, n. 10, p. 1451–5, out. 2005.

IBGE (Instituto Brasileiro de Geografia e Estatística). Produção da Pecuária Municipal. Disponível em: ftp://ftp.ibge.gov.br/Producao_Pecuaria/Producao_da_Pecuaria_Municipal/2013/ppm2013.pdf>. Acesso em: 05 Nov 2013.

IGARASI, M. S. et al. Características de carcaça e parâmetros de qualidade de carne de bovinos jovens alimentados com grãos úmidos de milho ou sorgo. **Revista Brasileira de Zootecnia**, v. v.37, n. n.3, p. p.520–528, 2008.

IPA. Ingenuity Systems. Disponível em: <www.ingenuity.com> Acesso em 15 Out 2014.

JIMÉNEZ-MARÍN, Á. et al. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. **BMC Proceedings**, v. 3, n. Suppl 4, p. S6, 2009.

JOHNSON, D. D. et al. Effects of percentage Brahman and Angus breeding age-season of feeding and slaughter end point on meat palatability and muscle characteristics. **Journal of Animal Science**, v. v.68, p. p.1980–1986, 1990.

KEITH, F. K. M. et al. Chemical and Sensory Properties of Thirteen Major Beef Muscles. **Journal of Food Science**, v. 50, n. 4, p. 869–872, jul. 1985.

KIM, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome biology**, v. 14, n. 4, p. R36–R36, abr. 2013.

KOOHMARAIE, M. Muscle Proteinases and Meat Aging Mohammad Koohmaraie. **Meat Science**, v. 36, p. 93–104, 1994.

KOOHMARAIE, M. Biochemical factors regulating the toughening and tenderization processes of meat. **Meat science**, v. 43S1, n. 96, p. 193–201, jan. 1996.

KOOHMARAIE, M. et al. Meat tenderness and muscle growth: is there any relationship? **Meat Science**, v. 62, p. 345–352, 2002.

LEE, H.-J. et al. Comparative Transcriptome Analysis of Adipose Tissues Reveals that ECM-Receptor Interaction Is Involved in the Depot-Specific Adipogenesis in Cattle. **PloS one**, v. 8, n. 6, p. e66267–e66267, jan. 2013.

LEE, H.-J. et al. Comparison of Metabolic Network between Muscle and Intramuscular Adipose Tissues in Hanwoo Beef Cattle Using a Systems Biology Approach. **International Journal of Genomics**, v. 2014, p. 1–6, 2014.

LEHMANN, K. B. Studies of the causes for the toughness of meats. **Arch. Hyg.**, v. 3, p. 134, 1907.

LI, R. W.; RINALDI, M.; CAPUCO, A. V. Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. **Veterinary research**, v. 42, n. 1, p. 114–114, 2011.

LIU, G. E. et al. Analysis of copy number variations among diverse cattle breeds. **Genome Research**, v. 20, n. 5, p. 693–703, 1 maio 2010.

MACMANES, M. D. On the optimal trimming of high-throughput mRNA sequence data. **Frontiers in Genetics**, v. 5, 2014.

MCCABE, M. et al. RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. **BMC genomics**, v. 13, n.

Vldl, p. 193–193, jan. 2012.

MILLER, M. F. et al. Consumer thresholds for establishing the value of beef tenderness. **Journal of Animal Science**, v. 79, n. 12, p. 3062–8, 2001.

MORGAN, J. B. et al. National beef tenderness survey. **Journal of Animal Science**, v. 69, n. 8, p. 3274–83, 1991.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. n. May, p. 1–8, 2008.

NOOKAEW, I. et al. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. **Nucleic acids research**, v. 40, n. 20, p. 10084–97, nov. 2012.

OSHLACK, A.; ROBINSON, M. D.; YOUNG, M. D. From RNA-seq reads to differential expression results. **Genome biology**, v. 11, n. 12, p. 220–220, jan. 2010.

PAGE, B. T. et al. Association of markers in the bovine CAPN1 gene with meat tenderness in large crossbred populations that sample influential industry sires The online version of this article , along with updated information and services , is located on the World Wide Web. p. 3474–3481, 2004.

PRADO, I. N. D. et al. Meat quality of crossbred bulls fed with sorghum silage or sugar cane and slaughtered at two levels of fat thickness. **Acta Scientiarum. Technology**, v. 34, n. 3, p. 337–344, maio 2012.

RAMAYO-CALDAS, Y. et al. Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. **BMC genomics**, v. 13, p. 547–547, jan. 2012.

RIESGO, A. et al. Optimization of preservation and storage time of sponge tissues to obtain quality mRNA for next-generation sequencing: PRESERVATION AND STORAGE TIME OF SPONGE mRNA. **Molecular Ecology Resources**, v. 12, n. 2, p. 312–322, mar. 2012.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010.

ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biol**, v. 11, n. 3, p. R25, 2010.

ROLF, M. M. et al. Genomics in the United States beef industry. **Livestock Science**, v. 166, p. 84–93, ago. 2014.

- ROSS, M. G. et al. Characterizing and measuring bias in sequence data. **Genome Biol**, v. 14, n. 5, p. R51, 2013.
- SAVELL, J. W. et al. National beef market basket survey. **Journal of Animal Science**, v. 69, n. 6, p. 2883–93, 1991.
- SCHENKEL, F. S. et al. Association of single nucleotide polymorphisms in the leptin gene with carcass and meat quality traits of beef cattle. **Journal of Animal Science**, v. 83, n. 9, p. 2009–20, set. 2005.
- SCHOLEY, R. A et al. Identifying host pathogenic pathways in bovine digital dermatitis by RNA-Seq analysis. **Veterinary journal (London, England : 1997)**, v. 197, n. 3, p. 699–706, set. 2013.
- SHACKELFORD, S. D. et al. AN EVALUATION OF TENDERNESS OF THE LONGISSIMUS MUSCLE OF ANGUS BY HEREFORD VERSUS BRAHMAN CROSSBRED HEIFERS. **Journal of animal science**, v. 69, n. 1, p. 171–177, 1991.
- SMULDERS, F. J. M. et al. Beef Tenderness and Sarcomere Length. **Meat Science**, v. 28, p. 349–363, 1990.
- SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of RNA-seq data. **BMC bioinformatics**, v. 14, n. 1, p. 91–91, jan. 2013.
- THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM et al. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. **Science**, v. 324, n. 5926, p. 522–528, 24 abr. 2009.
- TITTO, E. A. L. et al. Reactivity of Nellore steers in two feedlot housing systems and its relationship with plasmatic cortisol. **Livestock Science**, v. 129, n. 1-3, p. 146–150, abr. 2010.
- TIZIOTO, P. C. et al. Calcium and potassium content in beef: influences on tenderness and associations with molecular markers in Nellore cattle. **Meat science**, v. 96, n. 1, p. 436–40, jan. 2014.
- TRAN, D.; THU, N. MEAT QUALITY : UNDERSTANDING OF MEAT TENDERNESS AND INFLUENCE OF FAT CONTENT ON MEAT FLAVOR. p. 65–70, 2006.
- TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, n. 5, p. 511–515, 2010.
- TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562–578, 2012.

TRAPNELL, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. **Nature biotechnology**, v. 31, n. 1, p. 46–53, jan. 2013.

USDA (United States Departmente of Agriculture). **Livestock and Poultry: World Markets and Trade**. Disponível em:

<http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf>. Acesso em: 28 Nov 2014.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews. Genetics**, v. 10, n. 1, p. 57–63, jan. 2009.

WARNER, R. D. et al. Genetic and environmental effects on meat quality. **Meat Science**, v. 86, n. 1, p. 171–183, set. 2010.

WESTON, A. R.; ROGERS, R. W.; ALTHEN, T. G. The Role of Collagen in R Meat : Tenderness. n. 6, 2002.

WHEELER, T. L. et al. The effects of Piedmontese inheritance and myostatin genotype on the palatability of longissimus thoracis, gluteus medius, semimembranosus, and biceps femoris. **Journal of Animal Science**, v. 79, n. 12, p. 3069–74, dez. 2001.

ZHANG, Z. H. et al. A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. **PLoS ONE**, v. 9, n. 8, p. e103207–e103207, ago. 2014.

ZHAO, C. et al. Functional genomic analysis of variation on beef tenderness induced by acute stress in angus cattle. **Comparative and functional genomics**, v. 2012, p. 756284–756284, jan. 2012.