

**UNIVERSIDADE ESTADUAL PAULISTA**  
**FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS**  
**CAMPUS DE JABOTICABAL**

**TRANSCRIPTOMA DO MÚSCULO *LONGISSIMUS DORSI* DE  
BOVINOS MACHOS ADULTOS DA RAÇA NELORE**

**Eduardo Paulino Castan**

Zootecnista

JABOTICABAL - SÃO PAULO - BRASIL

FEVEREIRO - 2014

**UNIVERSIDADE ESTADUAL PAULISTA  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CAMPUS DE JABOTICABAL**

**TRANSCRIPTOMA DO MÚSCULO *LONGISSIMUS DORSI* DE  
BOVINOS MACHOS ADULTOS DA RAÇA NELORE.**

**Eduardo Paulino Castan**

**Orientador: Prof. Dr. Henrique Nunes De Oliveira**

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Genética e Melhoramento Animal.

JABOTICABAL - SÃO PAULO - BRASIL

FEVEREIRO - 2014

C346t Castan, Eduardo Paulino  
Transcriptoma do músculo longissimus dorsi de bovinos machos adultos da raça nelore. / Eduardo Paulino Castan. -- Jaboticabal, 2014  
iii, 80 p. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2014

Orientador: Henrique Nunes de Oliveira

Banca examinadora: Luciana de Almeida Regitano, André Luiz Julien Ferraz, Rogério Abdallah Curi, Luis Artur Loyola Chardulo  
Bibliografia

1. Transcriptoma. 2. Longissimus dorsi. 3. Expressão gênica. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.08:636.2

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

## **DADOS CURRICULARES DO AUTOR**

**EDUARDO PAULINO CASTAN** - nasceu no dia 03 de janeiro de 1985, na cidade de Recife - PE, filho de José Milton Castan Júnior e Celene Carvalho Paulino Fontenla. Em dezembro de 2007 obteve o título de zootecnista pela Universidade Estadual Paulista (UNESP) e em março de 2008 ingressou no Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal - SP como bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, obtendo o grau de mestre em 23 de fevereiro de 2010, sob orientação do Prof. Dr. Henrique Nunes de Oliveira e co-orientação da Profa. Dra. Maeli Dal Pai Silva. Em março de 2010 ingressou no curso de Doutorado do Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal - SP como bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, sob orientação do Prof. Dr. Henrique Nunes de Oliveira.

*Eu não vim até aqui pra desistir  
agora! Se depender de mim, eu vou até  
o fim!*

*Até o fim (Engenheiros do Hawaií)*

*Talvez o sucesso esteja escondido em  
algum lugar entre o “não deixe pra  
amanhã o que você pode fazer hoje” e o  
“por que fazer hoje se você pode fazer  
amanhã”*

*Dedico este trabalho...*

*Aos meus pais Castan e Celene por  
terem me guiado até aqui.*

*À minha esposa Renata por ter me  
escolhido para compartilhar a vida.*

## AGRADECIMENTOS

À Life Technologies que me escolheu como membro de sua equipe e que possibilitou o desenvolvimento deste trabalho.

Aos meus gestores Gustavo e Munera por permitirem que eu dividisse meu tempo entre o trabalho na Life e o trabalho na pós graduação.

À Carla Abdo pela **enorme** ajuda na compra dos kits e reagentes, o que possibilitou as seis corridas de sequenciamento e a conclusão deste trabalho.

Ao Léo pelas análises de bioinformática e que pacientemente me ajudou de várias formas, inclusive atendendo minhas ligações e respondendo meus emails durante suas férias.

À Dra. Dirce e ao Dr. Emmanuel que permitiram que eu utilizasse o Ion Proton do laboratório deles mesmo durante a intensa rotina de corridas e análises.

À Bruna que sempre bem humorada e muito pacientemente me ajudou com cada uma das seis corridas de sequenciamento. E, as vezes, ficando até tarde no laboratório.

Ao Rafael Oliveira que fez de tudo para que eu conseguisse receber todos os kits e reagentes a tempo.

Ao Professor Henrique por acreditar que eu conseguiria levar uma vida dupla e concluir o doutorado.

À minha família Felipe, Marcelo, Fernando, Carol, Orlando, Cinira, Ha, Gu, Ibis, Ana e Gaby.

À toda equipe envolvida na produção dos animais utilizados neste trabalho, desde criação até a coleta do tecido.

Ao Prof. Towia Libermann por me aceitar em seu laboratório e me dar a oportunidade única de estudar em Harvard.

Ao Prof. Manoj Bhasin, um dos melhores bioinformatas do mundo, por me passar um pouquinho do seu enorme conhecimento.

À Fundação de Amparo a Pesquisa do Estado de São Paulo pelo Auxílio Financeiro, processo 2011/12520-3, e pelo fornecimento da bolsa de doutorado processo 2010/01012-4.

## SUMÁRIO

<b>TRANSCRIPTOMA DO MÚSCULO LONGISSIMUS DORSI DE BOVINOS MACHOS ADULTOS DA RAÇA NELORE .....</b>	<b>II</b>
<b>TRANSCRIPTOME OF LONGISSIMUS DORSI MUSCLE OF NELLORE BREED CATTLE MALE BULLS .....</b>	<b>III</b>
<b>1. Introdução.....</b>	<b>1</b>
Transcritos .....	1
Análise de transcriptoma .....	4
Sequenciamento de nova geração .....	5
Tecnologia Ion Torrent .....	6
Produção animal, genética e qualidade da carne.....	7
Crescimento muscular.....	11
Análise de transcriptoma por sequenciamento de RNA.....	12
Transcriptoma e a qualidade da carne.....	17
<b>2. Objetivos .....</b>	<b>28</b>
<b>3. Material e Métodos.....</b>	<b>28</b>
Extração de RNA mensageiro com Dynabeads® mRNA DIRECT™ Kit (Life Technologies, Carlsbad, CA, EUA) .....	29
Construção da biblioteca de cDNA com Ion Total RNA-Seq Kit v2 (Life Technologies, Carlsbad, CA, EUA).....	29
Amplificação da biblioteca por PCR em emulsão (ePCR) com o Ion PI™ Template OT2 200 Kit v2 (Life Technologies, Carlsbad, CA, EUA).....	30
Preparo da reação de sequenciamento com Ion PI™ Sequencing 200 Kit v2 (Life Technologies, Carlsbad, CA, EUA) .....	31
Análise dos resultados.....	31
<b>4. Resultados .....</b>	<b>36</b>
Rendimento da extração dos mRNAs e da construção das bibliotecas de cDNA .....	36
Sequenciamento do transcriptoma do músculo LD de bovinos da raça Nelore.....	37
Análise dos dados .....	38
<b>5. Discussão .....</b>	<b>57</b>
<b>6. Conclusão.....</b>	<b>62</b>
<b>7. Considerações finais .....</b>	<b>63</b>
<b>8. Referências .....</b>	<b>65</b>

## **TRANSCRIPTOMA DO MÚSCULO LONGISSIMUS DORSI DE BOVINOS MACHOS ADULTOS DA RAÇA NELORE**

**RESUMO** - O transcriptoma é o conjunto e a quantidade de transcritos de uma célula em um estágio de desenvolvimento específico ou condição fisiológica. Estudar o transcriptoma é essencial para interpretar os elementos funcionais do genoma e revelar os constituintes moleculares de células e tecidos. Levando em consideração a grande importância do mercado da carne de bovinos e a necessidade e dificuldade de melhoria da qualidade da carne, o presente projeto objetivou traçar um panorama completo do transcriptoma do músculo *longissimus dorsi* (LD) de bovinos da raça Nelore, bem como identificar genes com diferentes perfis de expressão associados ao crescimento muscular por meio da técnica de RNA-seq. Foram utilizadas amostras de 10 animais da raça Nelore com dois diferentes perfis de crescimento muscular provenientes de um rebanho participante do programa de melhoramento genético. Amostras de mRNA do músculo LD foram coletadas para extração, processamento do RNA, sequenciamento e posterior análise do transcriptoma. No total, foram sequenciados aproximadamente 453 milhões de fragmentos e identificados 14.876 genes conhecidos, 45.938 potenciais novos genes e 10.960 potenciais novas isoformas de transcritos. Foi calculado o nível de expressão de 12.082 genes e 110 tiveram valor de expressão diferencial significativo entre os grupos ( $p \leq 0,01$ ). Por meio da análise de relevância de termos ontológicos foram identificados 52 termos relevantes entre os genes diferentemente expressos, dos quais 12 estavam associados ao desenvolvimento muscular. Neste estudo, foi caracterizado o transcriptoma do músculo *longissimus dorsi* de bovinos da raça Nelore por meio de sequenciamento de nova geração, fornecendo assim, uma valiosa fonte de entendimento do genoma do boi.

**Palavras-chave:** Transcriptoma, *Longissimus dorsi*, Expressão Gênica, Sequenciamento de RNA.

## TRANSCRIPTOME OF LONGISSIMUS DORSI MUSCLE OF NELLORE BREED CATTLE MALE BULLS

**ABSTRACT** - The transcriptome is the set and the amount of transcripts in a cell at a specific developmental stage or physiological condition. Understanding the transcriptome is essential for interpreting the functional elements of the genome and to reveal molecular constituents of cells and tissues. Considering the great importance of the beef market, the need and difficulty of improving meat quality, this research aimed to provide an overview of the complete longissimus dorsi (LD) transcriptome of Nellore cattle and to identify genes with differential expression profiles associated with muscle growth using RNA-seq. Ten mRNA samples of Nellore breed steers with two different muscle growth profiles from a herd participating of a breeding program were used. LD muscle samples were collected for RNA extraction and processing, sequencing and subsequent transcriptome analysis. In total, approximately 453 million fragments were sequenced and 14,876 known genes, 45,938 potential new genes and 10,960 potential new transcripts isoforms were identified. The expression levels of 12,082 genes were calculated and 110 have significant differential expression values between the groups ( $p \leq 0.01$ ). The ontology terms relevance analysis identified 52 relevant terms among the differentially expressed genes, which 12 were associated with muscle development. In this study, we obtained the transcriptome of Nellore cattle *longissimus dorsi* muscle through next-generation sequencing, providing a valuable source of information about the bovine genome.

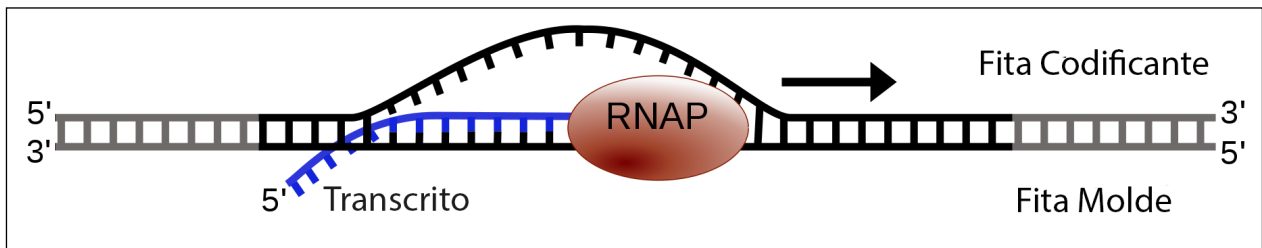
**Keywords:** Transcriptome, *Longissimus dorsi*, Gene expression, RNA sequencing.

## 1. Introdução

Rápidos progressos têm acontecido no melhoramento animal nas últimas décadas. Tecnologias avançadas em genética molecular para a análise genômica, têm permitido a identificação de muitos genes e marcadores associados a genes que afetam características de interesse econômico. Ao mesmo tempo, a integração entre a pesquisa operacional ou aplicada e a pesquisa básica possibilitaram grande avanço biotecnológico, sendo seus resultados implementados rapidamente em diversas áreas da indústria e da saúde. Uma abordagem que contribui muito para o desenvolvimento do conhecimento e suas aplicações nas áreas biológicas é o estudo do transcriptoma (REECY *et al.*, 2006). O transcriptoma pode ser definido como o conjunto e a quantidade de transcritos de uma célula em um estágio de desenvolvimento específico ou condição fisiológica. A população de transcritos especifica a identidade da célula e participa na modulação de suas atividades em diferentes momentos de sua vida.

## Transcritos

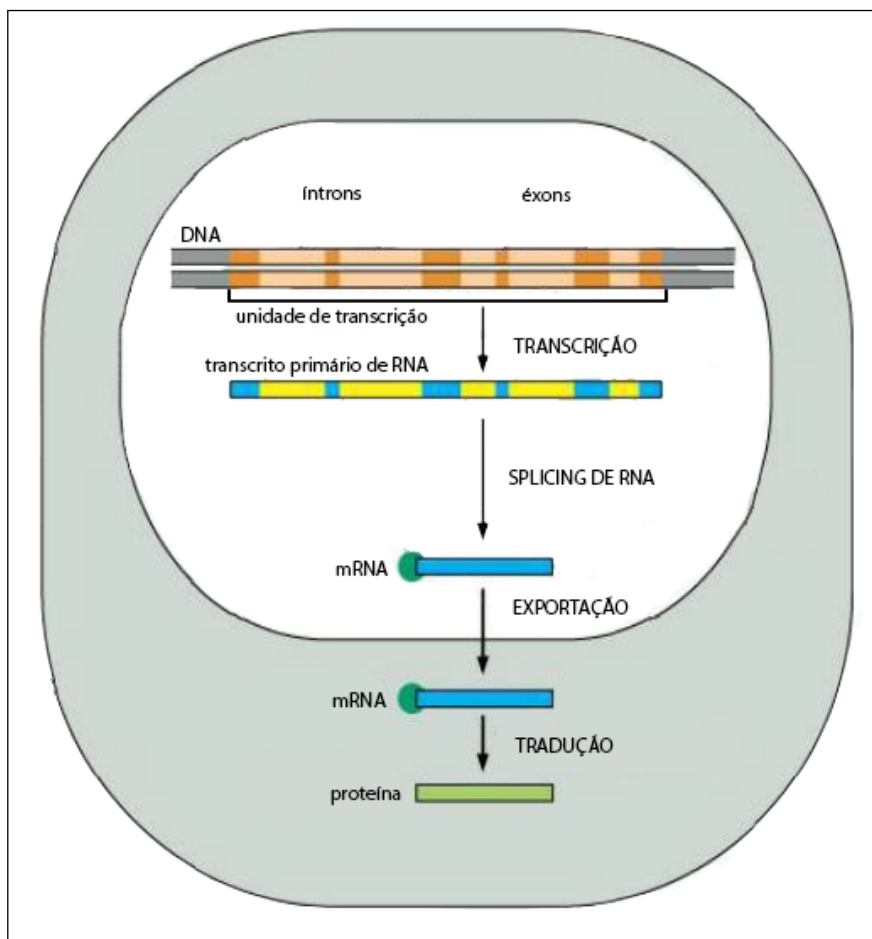
A síntese de proteínas envolve copiar regiões específicas do DNA (os genes) em polinucleotídeos química e funcionalmente diferentes conhecidos como ácido ribonucleico ou RNA. O RNA contém toda informação do DNA que foi utilizado como molde para sintetizá-lo, por meio de um processo chamado transcrição. Neste processo uma enzima chamada de RNA polimerase se complexa a uma das fitas do DNA e catalisa a polimerização da molécula de RNA, incorporando um nucleotídeo por vez e utilizando a fita do DNA como molde. O RNA proveniente deste processo é chamado de transcrito (Figura 1). Transcritos de RNA que direcionam a síntese de proteínas são chamados de RNAs mensageiros (mRNAs), enquanto outros transcritos servem como transportador (RNA transportador ou tRNA) ou como componentes dos ribossomos (RNA ribossomal ou rRNA) ou são pequenos RNAs, como os micro RNAs (miRNAs).



**Figura 1.** A síntese de um transcrito. A RNA polimerase (RNAP) move-se paulatinamente ao longo do DNA. Conforme avança, a polimerase adiciona nucleotídeos um a um, à cadeia de RNA, no sítio de polimerização, utilizando uma fita de DNA exposta como molde. O RNA transcrito é, conseqüentemente, uma cópia complementar fita simples de uma das fitas do DNA.

A quantidade de RNA gerada a partir de uma região particular do DNA é controlada por proteínas regulatórias de genes que se ligam em sítios específicos no DNA perto das sequências codificantes de um gene. Em qualquer célula, em um dado momento, alguns genes são utilizados para gerar RNA em uma escala muito grande enquanto outros genes não são transcritos. Para um gene ativo, milhares de transcritos de RNA podem ser gerados a partir do mesmo segmento de DNA e este processo pode ocorrer em várias células ao mesmo tempo. Desta maneira, cada mRNA pode direcionar a tradução de milhares de cópias de cadeias polipeptídicas e portanto, a informação contida em uma pequena região do DNA pode direcionar a síntese de milhões de cópias de uma proteína específica. Diferentemente dos organismos procariontes, a maioria dos genes dos eucariontes têm as suas sequências codificantes (chamadas de éxons) interrompidas por sequências não codificantes (chamada de íntrons). Para produzir uma proteína, um fragmento inteiro do gene, incluindo éxons e íntrons, é primeiramente transcrito em uma grande molécula de RNA (o transcrito primário). Antes dessa molécula de RNA deixar o núcleo da célula, um complexo de RNA e enzimas de processamento removem as sequências intrônicas, produzindo, portanto, uma molécula consideravelmente menor de RNA. Após esse processamento do RNA (chamado de *splice*) ter sido concluído, a molécula de RNA migra para o citoplasma e direciona a síntese de uma proteína específica (Figura 2). Esta, aparentemente não econômica maneira de transferir informação em eucariontes, é responsável por fazer a síntese de proteína muito mais versátil. Os transcritos primários

de RNA de alguns genes podem ser processados de várias maneiras diferentes para produzir diferentes mRNAs. Isso permite a produção de diferentes proteínas a partir de um mesmo gene. Além disso, a presença dos vários íntrons facilita eventos de recombinação entre éxons e esse tipo de arranjo gênico foi extremamente importante no processo evolutivo dos genes, acelerando o processo pelo qual os organismos desenvolveram novas proteínas a partir de “peças” pré-existentes, ao invés de desenvolver sequências totalmente novas de aminoácidos.



**Figura 2.** Nas células eucarióticas a molécula de RNA produzida por transcrição podem estar contidas tanto sequências codificantes (éxons), como sequências não codificantes (íntrons). Antes do RNA ser traduzido em proteína, os íntrons são removidos no processo chamado de splicing e o mRNA resultante é transportado do núcleo para o citoplasma.

## Análise de transcriptoma

A análise do transcriptoma engloba a organização de todas as espécies de transcritos, incluindo os RNAs mensageiros (mRNAs), os RNAs não codificantes e os micro RNAs (miRNAs); a determinação da estrutura transcricional dos genes, isto é, sítios de início, terminações 5' e 3', padrões de splicing e outras modificações pós-transcripcionais; e a quantificação na mudança do nível de expressão de cada transcrito durante o desenvolvimento e sob diferentes condições (WANG et al., 2008).

Apesar das extensivas análises com o intuito de esclarecer o transcriptoma dos mamíferos, as tentativas de colocar a complexidade dos mRNAs dentro do contexto biológico vinham sendo sem sucesso, principalmente devido às limitações das tecnologias utilizadas até o momento. Metodologias baseadas em arranjos (*arrays*) eram as mais populares, porém essa abordagem é fisicamente limitada por problemas como densidade das sondas (quantidade de sondas em uma lâmina), tamanho do éxon, conteúdo adequado das sequências das sondas, hibridização cruzada, baixa sensibilidade para transcritos raros e dificuldade da identificação da combinação de éxons. Metodologias baseadas em *tags*, como *massively parallel signature sequencing* (MPSS) (REINARTZ et al., 2002), *serial analysis of gene expression* (SAGE) (VELCULESCU et al., 1995), *cap analysis of gene expression* (CAGE) (SHIRAKI et al., 2003) e *polony multiplex analysis of gene expression* (PMAGE) (KIM et al., 2007), são melhores em relação à sensibilidade e discriminação de alguns sinais em relação ao microarray, mas podem ter ambiguidade no mapeamento devido ao pequeno tamanho (17-20 nucleotídeos) das *tags*.

Muito do progresso nos estudos com transcriptomas foi obtido, até agora, por estudos de genes únicos, proteínas, ou vias metabólicas, ou seja, são resultados provenientes da participação de vários estudos individuais nessa área, sem uma abordagem ampla e sem um resultado global em cada estudo. No entanto, o surgimento das tecnologias dos sequenciadores de nova geração, proporciona oportunidades de gerar perfis globais de transcriptomas completos (MARDIS, 2008; SCHUSTER, 2008; TANG et al., 2009). Por meio da análise do transcriptoma, milhares de variantes e isoformas transcricionais expressas em tecidos e órgãos de mamíferos

têm sido detectadas (MORTAZAVI *et al.*, 2008; SULTAN *et al.*, 2008; CLOONAN *et al.* 2009). Este fato pode permitir e acelerar o entendimento sobre a complexidade da expressão e regulação gênica, e as relações celulares, bem como ajudar na elucidação de descobertas mais recentes, como a possibilidade de determinados *loci* expressarem transcritos não codificantes (CAWLEY *et al.*, 2004), e que a existência de genes *sense* e *antisense* é comum (ENGSTROM *et al.*, 2006).

### **Sequenciamento de nova geração**

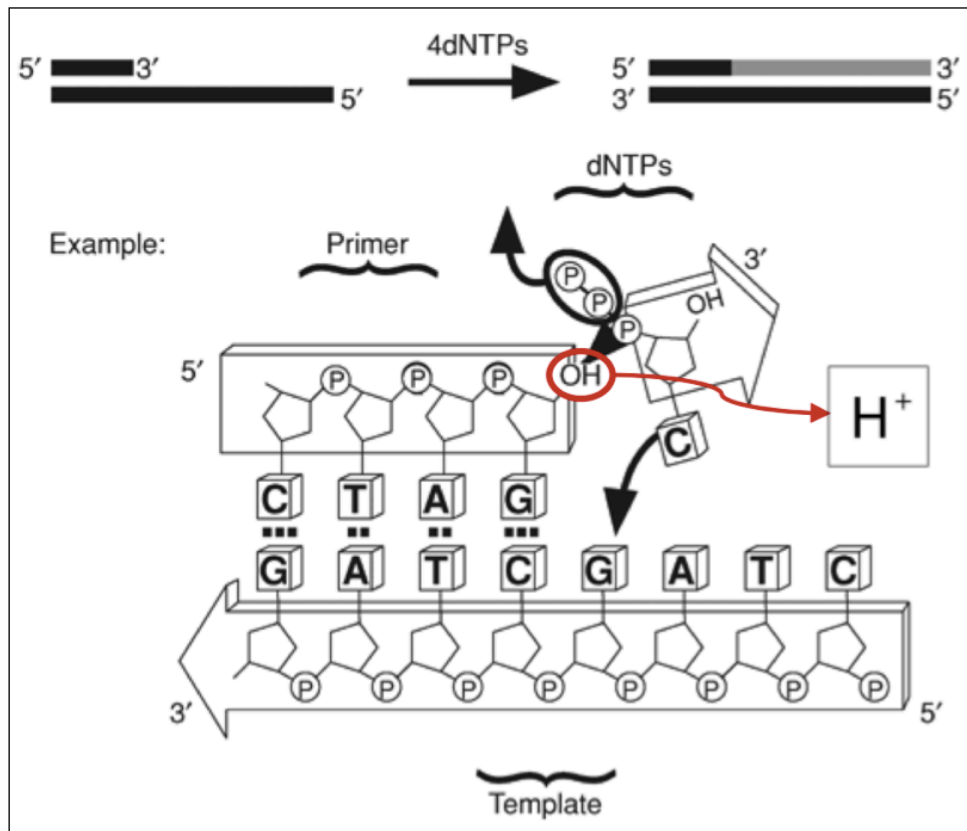
No campo da genômica, os sequenciadores de nova geração representam o atual estado da arte. Eles tem altíssima produtividade em termos de dados produzidos com um custo muito mais baixo, quando comparados com os mais modernos sequenciadores capilares baseados no método de Sanger (SANGER *et al.*, 1977; KATO, 2009). Sendo assim, essa tecnologia é uma ferramenta poderosa e de baixo custo para análise de transcriptoma em larga escala, que vem causando grande impacto na genômica e mudando a natureza da experimentação em genética (SCHUSTER, 2008; WOLD & MYERS, 2008). Com essa abordagem é possível processar milhões de leituras de sequências ao mesmo tempo, sendo viável terminar um experimento com apenas uma ou duas reações de sequenciamento (MARDIS, 2008).

Uma grande vantagem desses sequenciadores é a possibilidade de utilizar fragmentos provenientes de bibliotecas geradas sem a utilização de vetores de clonagem e amplificação em bactérias, evitando dessa maneira, problemas provenientes dessa estratégia na representação do transcriptoma. O procedimento para produzir as bibliotecas é simples: fragmentos de DNA ou cDNA que podem ser originadas de diferentes processos, são preparadas para o sequenciamento, por meio da ligação de adaptadores específicos em ambas extremidades dos fragmentos, o que permite a manipulação de fragmentos de DNA/cDNA de sequências desconhecidas. Outra grande vantagem, é a pouca quantidade de DNA/RNA necessária (no máximo poucos microgramas) para produzir as bibliotecas (TANG *et al.*, 2009; WANG *et al.*, 2008).

Entre os atuais sequenciadores de nova geração, existe a Ion Proton™ Sequencer (Life Technologies), que comparado com outras plataformas, possui baixo custo e gera quantidade suficiente de dados por corrida, sendo indicada para análise de transcriptoma de mamíferos. Além disso, possui um sistema completo que inclui um potente computador e software robusto para processar grandes quantidades de informação.

### **Tecnologia Ion Torrent**

A tecnologia Ion Torrent, que é a tecnologia utilizada pela plataforma Ion Proton™, utiliza um processo de sequenciamento único catalisado pela DNA polimerase. Cada corrida dura por volta de quatro horas e podem ser produzidos atualmente 15 giga-base (Gb) de dados, ou seja, podem ser sequenciadas em uma única corrida até 15 bilhões de bases nitrogenadas com média de leitura de fragmentos de 200 pares de bases (pb). Alguns dos sequenciadores de nova geração utilizam desoxirribonucleotídeos fosfatados (dNTPs) marcados por fluoróforos para a detecção das bases. Ion Torrent é diferente nesse ponto pois a detecção é feita diretamente da reação de incorporação dos dNTPs pela polimerase. Isso é possível pois cada dNTP é disponibilizado para polimerização individualmente e sua incorporação gera naturalmente um próton de hidrogênio que altera o pH do meio. Essa alteração do pH é detectada por um transistor ISFET (BERGVELD, 2003) e, então convertida em sinal elétrico (figura 3).



**Figura 3.** Tecnologia Ion Torrent de sequenciamento. A reação de incorporação de um nucleotídeo pela DNA polimerase libera íons H<sup>+</sup> e, conseqüentemente, altera o pH do meio. Essa variação de pH é detectada por um transistor ISFET.

### Produção animal, genética e qualidade da carne

A tecnologia de sequenciamento da nova geração está perto de revolucionar a genômica na área da produção animal. Pode-se esperar em breve, o sequenciamento da maioria, se não de todas as espécies com importância zootécnica, o que fornecerá ferramentas fundamentais para análises comparativas, genotipagem de SNPs de genomas completos, estudo da expressão de mRNAs e miRNAs e análise da metilação de DNA de cada espécie. Devido à capacidade de análise em larga escala, tem havido grande redução de custo de reagentes, e grande aumento de quantidade de dados gerados por amostra.

A comunidade de pesquisa de genomas de animais de produção está preparada para as tecnologias comumente utilizadas, porém ainda não está pronta para essas novas tecnologias. O fator limitante das descobertas por meio da pesquisa partirá das

tecnologias disponíveis e chegará à habilidade de integrar, analisar e interpretar grandes volumes de dados díspares. A próxima década será caracterizada por grandes mudanças e grandes desafios na melhoria dos sistemas da produção animal.

Uma das áreas da produção animal com enorme importância econômica e igual potencial para melhoria é a da qualidade da carne de bovinos. Na última década foi o tema de muitas pesquisas e a melhoria desta característica é, no momento, essencial para se manter o consumo e satisfazer a demanda do consumidor por carne macia e com sabor. Porém, as características que determinam a qualidade da carne são difíceis de serem melhoradas por meio de seleção convencional devido, principalmente, aos problemas relacionados às mensurações que, além do alto custo, somente são possíveis após o abate. Além disso, a falta de conhecimento da complexidade da arquitetura gênica relacionada à carne é um grande limitante na exploração completa das oportunidades de seleção. Ainda, a produção animal vem se tornando cada vez mais competitiva e com baixas margens de lucro. Desta maneira, muitas pesquisas tem dado enfoque na melhoria da eficiência da produção animal. Para a indústria da produção de carne, a ênfase vem sendo dada no aumento do entendimento de processos biológicos e vias moleculares associadas com o crescimento e desenvolvimento muscular.

Na última década a genética da qualidade da carne de bovinos se tornou alvo de inúmeras pesquisas e a melhoria desta característica é um elemento presente e essencial para dar suporte ao consumo de carne e satisfazer a demanda do consumidor por uma alimentação melhor, mais saudável e de maior qualidade nutricional. O aumento da conscientização do consumidor em relação a alimentos mais saudáveis e de melhor qualidade vem encorajando empresas, como frigoríficos, grupos de melhoramento e mesmo o produtor, a considerar com mais atenção a qualidade da carne, e incluir características da qualidade como parte integrante dos objetivos de seleção de programas de melhoramento, que podem incorporar informações genéticas e moleculares (DRANSFIELD, 2001; TARRANT, 1998).

Sendo o músculo esquelético o tecido de maior importância para produção de carne, a seleção tem sido direcionada em favor do desenvolvimento muscular com o objetivo de se produzir carcaças maiores, e muitas vezes aumentando a proporção de

músculo em relação à gordura. O melhoramento genético em bovinos de corte tem aumentado as taxas de crescimento muscular e isso pode ter modificado algumas características musculares (KOOHMARAIE *et al.*, 2002). Músculos são compostos de uma mistura heterogênea de fibras que diferem entre si de acordo com suas propriedades metabólicas e contráteis. A diversidade dos tipos de fibras comanda, não somente a demanda energética do músculo como um todo, mas também influencia as características finais de qualidade da carne em todas as espécies (KLONT *et al.*, 1998). De fato, a variabilidade dos tipos de fibras musculares está associada com diferenças no teor de gordura e no perfil de ácidos graxos (que influencia o sabor e propriedades dietéticas da carne), teor de glicogênio (que determina em grande parte o pH final da carne) e a atividade proteolítica (que contribui com a maturação e a maciez). Outros aspectos musculares também são responsáveis por influenciar a qualidade da carne. Entre eles estão as características do tecido conectivo intramuscular, como teor e solubilidade de colágeno (MCCORNICK, 1999). Uma parte significativa na variabilidade e no sabor da carne é dada pela área e tipo de fibra muscular, bem como o colágeno e as características de lipídios. De um quarto a um terço da variabilidade da maciez aos dois e aos quinze dias após o abate e o *score* de sabor, foram relacionados com a variabilidade nestas características dos músculos (RENAND *et al.*, 2001). Além disso, atividade de calpastatina a 24h *postmortem* ou a razão da atividade da calpastatina com a atividade da calpaína são indicadores de maturação da fibra, o que pode também ajudar a prever a maciez final da carne (KOOHMARAIE *et al.*, 1995). Da mesma maneira, gordura muscular tem papel importante no sabor (DENOYELLE, 1995). Até o momento, somente as características mais importantes do músculo foram estudadas, mas a maior parte da variabilidade na qualidade da carne, como maciez e sabor, permanecem independente destes fatores (RENAND *et al.*, 2001).

Como apontado por CAMERON (1990), a qualidade da carne é um conceito composto e portanto difícil de ser mensurado de maneira simples e única, e é influenciada, não somente pelas características do músculo (tamanho e tipo de fibra, gordura e tecido conectivo), mas também por um grande número de outros fatores. Entre estes fatores estão, condições ambientais e de produção (taxa de crescimento, nutrição, idade, condições de (pré-) abate, maturação *postmortem*) e a genética do

animal (raça, genótipo). Características da qualidade da carne são controladas por um número desconhecido de genes pertencentes a *loci* de características quantitativas, sendo que alguns desses genes podem ter grande efeito em uma característica específica (ANDERSSON, 2001; ANDERSSON & GEORGES, 2004). Muitas das pesquisas vêm sendo direcionadas para essa área com o objetivo de encontrar genes potenciais ou regiões cromossômicas responsáveis pela qualidade da carne e seus derivados (TIZIOTO *et al.*, 2013; UIMARI *et al.*, 2013; ROTHSCCHILD *et al.*, 2007). Apesar da baixa/moderada herdabilidade, os efeitos genéticos desempenham um papel crucial na “modelação” da composição e qualidade da carcaça (SELLIER, 1998). Geralmente, entre 10% e 30% da variação de características da qualidade da carne e de seus derivados, como pH final, cor, capacidade de retenção de água, perda por gotejamento, maciez e marmoreio, é determinada pela genética do animal (SELLIER, 1998).

Assim como outras características zootécnicas, há certa dificuldade para um melhor desempenho do melhoramento genético da qualidade da carne (baixa herdabilidade, alto custo de mensuração, etc) e com outro limitante que é a sua mensuração somente pós abate. Além disso, a falta de conhecimento sobre o número de genes de características qualitativas da carne, sobre os efeitos quantitativos de genes únicos e sobre suas interações, representa um limite para a completa exploração das oportunidades de um plano de seleção. Parte da falta de compreensão da qualidade da carne em uma base genética é devido a grande parte do progresso na área da biologia do músculo vir acontecendo em análises individuais de genes, proteínas e vias metabólicas e não em análises globais.

Por outro lado, a rápida evolução dos sequenciadores de nova geração (WANG *et al.*, 2009) e a enorme quantidade de dados que eles podem gerar em um curto período, fornece aos pesquisadores novas oportunidades de estudar, de maneira global, toda complexidade da estrutura gênica de organismos e/ou tecidos de importância econômica. Com essa abordagem em larga escala é possível estudar não somente genes e/ou vias metabólicas individualmente, e sim mensurar diretamente um transcriptoma inteiro englobando componentes conhecidos e desconhecidos. O transcriptoma é dinâmico, está constantemente respondendo às condições fisiológicas

e ambientais em uma célula, tecido ou organismo, e sua análise fornece a primeira leitura funcional entre o genoma e o fenótipo expresso. Este tipo de análise tem expandido nosso entendimento da complexidade do transcriptoma de mamíferos e as inúmeras maneiras dele ser alterado. A análise do transcriptoma pode superar os limites relacionados à falta de conhecimento em relação à complexidade transcricional do músculo esquelético de bovinos, oferecendo novas oportunidades de melhoria da qualidade da carne, já que ela fornece ferramentas para analisar variabilidade genética quantitativa diretamente, e ao mesmo tempo permite detectar genes individuais que influenciam características quantitativas.

### **Crescimento muscular**

Os determinantes da qualidade da carne ainda não estão completamente elucidados, mas sabe-se que as características intrínsecas do músculo tem um papel importante (GEAY *et al.* 2001). Portanto, para melhor compreensão dos fatores que influenciam a qualidade da carne é muito importante um conhecimento mais profundo sobre os aspectos genéticos e os constituintes moleculares do músculo esquelético. Estes aspectos, no entanto, podem estar associados ou ser alterados por diversos fatores, entre eles, o crescimento muscular.

Como mencionado acima, a seleção de animais de produção, especialmente de gado de corte, tem sido direcionada para maior crescimento muscular, o que pode ter influenciado características do músculo e qualidade da carne. De fato, quanto maior a taxa de crescimento de um animal, maior a taxa de *turnover* proteico e maior, portanto, a taxa de degradação proteica. Estes fatos estariam diretamente relacionados com a ação do complexo enzimático responsável pelo processo de maturação e com a maciez da carne. RENAND e colaboradores (1994) mostraram que o aumento da proporção de músculo “magro” em relação à gordura em bois Charolês selecionados divergentemente para potencial de crescimento muscular, estava associado com maior velocidade da glicólise muscular e menor teor de gordura intramuscular.

Outros fatores que estão relacionados com a qualidade da carne e que podem ser alterados de acordo com o crescimento muscular são a proporção dos tipos de

fibras musculares e o teor de colágeno. PICARD *et al.* (1999) e PICARD *et al.*, (2006) mostraram, respectivamente, que o processo de amaciamento da carne no *postmortem* envolve mecanismos enzimáticos e físico-químicos que são dependentes da proporção de distintos tipos de fibras e que animais com maior potencial de crescimento tinham maior número de fibras, maior proporção de fibras glicolíticas rápidas e menor proporção de fibras lentas do que animais com menor potencial de crescimento. Por outro lado, RENAND *et al.*, (1994), mostraram que seleção para capacidade de crescimento não aumentou o teor de colágeno total, mas aumentou ligeiramente o teor de colágeno termo-lábil. Esses estudos mostraram, portanto, que é fundamental aumentar o conhecimento dos efeitos da seleção nas características do músculo com o objetivo final de compreender a variabilidade na qualidade da carne. Ferramentas de análise de transcriptoma nos fornecem novas perspectivas para o estudo da influência da taxa de crescimento dos animais nas características biológicas do músculo relacionadas com a qualidade da carne.

### **Análise de transcriptoma por sequenciamento de RNA**

Tecnologias de sequenciamento de RNA em larga escala (RNA-seq) (MORTAZAVI *et al.*, 2008) permitem a exploração do processo global de transcrição de uma célula ou tecido com detalhes requintados. RNA-seq é uma abordagem atrativa pois com ela o perfil de um transcriptoma pode ser caracterizado diretamente por meio de sequenciamento, sem a necessidade de conhecimento prévio do transcriptoma. Estas tecnologias vem sendo o foco de vários estudos recentes que demonstraram alta resolução e acurácia na quantificação e descoberta de transcritos, bem como na análise de modulação pós transcricional da expressão gênica e variações moleculares. Devido ao seu grande potencial, RNA-seq é considerado padrão ouro para estudos de transcriptomas e vem substituindo tecnologias baseadas em *arrays*, como microarray. De modo geral, as etapas de um experimento com RNA-seq envolvem isolar todos os RNAs de uma célula, convertê-los em cDNA, sequenciar os fragmentos de cDNA em um sequenciador de nova geração, mapear os fragmentos ao genoma de referência,

identificar variações de *splice*, descobrir novos transcritos e quantificar a expressão de transcritos.

O primeiro passo no processo de análise de RNA-seq é mapear os fragmentos sequenciados de RNA ao genoma de referência, o qual fornece a localização genômica de onde os fragmentos foram originados. Diferentemente do alinhamento de fragmentos provenientes de sequenciamento de DNA, o algoritmo para mapeamento de RNA-seq apresenta desafios adicionais. Por exemplo, devido aos genes de genomas eucariotos conterem íntrons, e devido aos fragmentos sequenciados a partir de transcritos maduros de mRNAs não incluírem esses íntrons, quaisquer programas de alinhamento de RNA-seq devem ser capazes de lidar com alinhamento espaçado (produtos de *splice*) e espaçamentos grandes. Em genoma de mamíferos, íntrons abrangem uma gama muito grande de comprimentos, tipicamente de 50 a 100.000 bases, os quais o algoritmo de alinhamento deve acomodar. Fragmentos alinhados dizem muito sobre a amostra sequenciada. Fragmentos que não foram alinhados completamente e inserções e deleções no alinhamento podem identificar polimorfismos entre a amostra sequenciada e o genoma de referência, ou mesmo eventos de fusão gênica. Fragmentos que foram alinhados fora de anotações gênicas são, geralmente fortes evidências de novos genes codificadores de proteínas e RNAs não codificantes. O alinhamento de fragmentos de RNA-seq podem revelar novos eventos de *splicing* alternativos e isoformas e também pode ser utilizado para quantificação acurada da expressão de genes e de transcritos, já que o número de fragmentos produzidos por um transcrito é proporcional à sua abundância.

O número de fragmentos de RNA-seq gerados a partir de um transcrito é diretamente proporcional a quantidade relativa daquele transcrito na amostra. No entanto, devido aos fragmentos de cDNA serem geralmente selecionados por tamanho durante o processo de construção de biblioteca, transcritos longos produzem mais fragmentos passíveis de serem sequenciados do que transcritos curtos. Por exemplo, supondo que uma amostra tenha dois transcritos, A e B, os quais estão presentes na mesma quantidade. Se B é duas vezes mais longo que A, uma biblioteca de RNA-seq terá (em média) duas vezes mais fragmentos B em relação à A. Para calcular o nível correto de expressão de cada transcrito, os softwares para quantificação de expressão

gênica para RNA-seq fazem uma normalização que considera a quantidade de fragmentos e o tamanho dos transcritos (esta normalização está explicada no tópico Material e Métodos).

Para quantificação acurada do nível de expressão de um transcrito é necessário identificação precisa de qual isoforma de um dado gene foi produzido cada fragmento. Isso somente é possível quando todas as variações de *splice* (isoformas) de um gene são conhecidas. Tentar quantificar expressão de genes e/ou de transcritos utilizando uma anotação incompleta ou incorreta de transcriptoma pode levar a valores imprecisos de expressão (TRAPNELL *et al.*, 2010). Isso, porque quando um gene sofre *splicing* alternativo e produz várias isoformas na mesma amostra, muitos fragmentos que se alinham a esse gene, podem ser referentes a éxons que compõe apenas uma única isoforma ou éxons que compõe mais de uma isoforma, o que complica o processo de contagem de cada transcrito. Para computar precisamente o nível de expressão de cada transcrito, uma simples contagem não é suficiente e uma inferência estatística mais sofisticada é necessária.

Mesmo para organismo bem estudados, a maioria de experimentos de RNA-seq pode revelar novos genes e novos transcritos. Uma análise recente e profunda de amostras de RNA-seq de 24 tecidos humanos e linhagens celulares revelou mais de 8.000 novos longos RNAs não codificantes juntamente com vários possíveis genes codificadores de proteínas (CABILI *et al.*, 2011). Um dos objetivos principais ao se fazer análise de transcriptoma por RNA-seq é a descoberta de novos genes e transcritos e fazer a análise de expressão diferencial deles entre diferentes condições. No entanto, pode ser difícil de distinguir sequências completas de novos transcritos de sequências parciais, utilizando apenas RNA-seq.

Outro tipo de abordagem que também pode ser feita por meio de RNA-seq é a análise diferencial de expressão gênica entre duas ou mais condições biológicas por meio da contagem de fragmentos. Vários programas foram desenvolvidos nos últimos anos com modelos estatísticos para verificar, para um dado gene, se a diferença observada na contagem de fragmentos era significativa, ou seja, se a diferença era maior do que seria esperado somente devido às variações aleatórias naturais. Entre eles estão os que utilizam métodos paramétricos (os mais utilizados) com variantes de

distribuição binomial negativa como o edgeR (ROBINSON *et al.*, 2010), HTseq (ANDERS & HUBER, 2010), bayseq (HARDCASTLE & KELLY, 2010) e NBPseq (SCHAFER *et al.*, 2011). E métodos não paramétricos como NOISeq (TARAZONA *et al.*, 2011) e Samseq (LI & TIBSHIRANI, 2013). No entanto, mesmo os algoritmos mais utilizados possuem falhas que podem comprometer a análise correta dos dados, como aqueles que utilizam modelo de Poisson. Considerando que os fragmentos são amostrados independentemente a partir de uma população, a contagem deles segue uma distribuição multinomial, a qual pode ser aproximada pela distribuição de Poisson. Consequentemente, a distribuição de Poisson foi utilizada para análise de expressão gênica diferencial (MARIONI *et al.*, 2008; WANG *et al.*, 2010). No entanto, foi verificado que a suposição da distribuição de Poisson é muito restrita: ela prediz menores variações do que foi observado (NAGALAKSHMI *et al.*, 2008; ROBINSON *et al.*, 2007). Portanto, o teste estatístico resultante desta suposição não controla erro tipo 1 (probabilidade de falsos positivos). Para resolver este problema foi proposto modelar dados de contagem com distribuições binomiais negativas (WHITAKER, 1914; ROBINSON *et al.*, 2010). A distribuição binomial negativa possui parâmetros que são unicamente determinados pela média  $\mu$  e variância  $\sigma^2$ . Porém, o número de replicatas em banco de dados de interesse é, geralmente, muito pequeno para se estimar ambos os parâmetros (média e variância) com segurança para cada gene. Para o *edgeR*, ROBINSON & SMYTH (2008) assumiram que a média e a variância são relacionadas por  $\sigma^2 = \mu + \alpha\mu^2$ , com uma única constante de proporcionalidade  $\alpha$  que é a mesma durante todo o experimento e que pode ser estimada a partir dos dados. Consequentemente, um parâmetro somente, precisa ser estimado para cada gene, permitindo aplicação em experimentos com pequeno número de replicatas. O pacote para ambiente R chamado DESeq2, que foi utilizado neste trabalho, estendeu este modelo permitindo um algoritmo eficiente para o ajuste do modelo aos dados. Desta maneira, pôde ser obtida uma seleção mais balanceada de genes diferencialmente expressos em todo o intervalo dinâmico dos dados.

Outra técnica simples, mas extremamente utilizada, para destacar processos biológicos, é a análise de super-representação de categoria gênica. Para realizar este tipo de análise, genes são agrupados em categorias por alguma característica biológica

comum, e em seguida são testados para encontrar categorias que são super-representadas entre genes diferencialmente expressos. Categorias de ontologia gênica (OG) são comumente utilizada nesta técnica e existem várias ferramentas disponíveis para realização de análise de OG, por exemplo: EasyGO (ZHOU & SU, 2007), GOMiner (ZEEBERG *et al.*, 2003), Gostat (BEISSBARTH & SPEED, 2004) e GOseq (YOUNG *et al.*, 2010), o qual foi utilizado neste trabalho. Embora estas ferramentas tenham algumas diferenças metodológicas (HUANG *et al.*, 2009), todas funcionam sob suposições semelhantes sobre a distribuição de genes diferencialmente expressos. Especificamente, assume-se que todos os genes são independentes e igualmente prováveis de serem selecionados como genes diferencialmente expressos, sob a hipótese nula. É esta suposição que faz as metodologias de análises padrões de OG inadequadas para dados de RNA-seq, devido ao viés na detecção de expressão gênica diferencial para genes com grande número de leituras. Consequentemente, quando análises padrões são utilizadas, quaisquer categorias contendo uma preponderância de genes longos serão mais prováveis de serem super-representadas do que categorias com genes de tamanhos médios. De maneira semelhante, categorias com genes altamente expressos são mais prováveis de serem detectadas como super-representadas do que categoria com genes de baixo valor de expressão. Tendo estes problemas identificados, é possível compensar o efeito de viés de seleção no teste estatístico do enriquecimento de uma categoria entre genes diferencialmente expressos.

A profundidade do RNA-seq, que está diretamente relacionada com a quantidade e qualidade de dados gerados durante o processo de sequenciamento, é a média de cobertura de todos os *loci* referentes às sequências alvo, e é fundamental para análises precisas e acuradas do transcriptoma (MARIONI *et al.*, 2008; LIU *et al.*, 2012). Espaçamentos na cobertura do sequenciamento causarão quebras na reconstrução dos transcritos, da mesma maneira que acontece durante a montagem de um genoma. Reconstrução de transcriptoma de eucariotos com dados de alta qualidade conterà milhares de sequências completas de transcritos. Por outro lado, reconstrução com dados de baixa qualidade, principalmente aquelas com poucos dados (menos que 10 milhões de fragmentos), pode conter dezenas ou mesmo centenas de milhares de

seqüências parciais de transcritos. Quantidades insuficientes de dados podem acarretar no aumento de detecção de falsos novos transcritos, na quantificação imprecisa de expressão gênica, na identificação errônea de eventos de *splice* e na perda de detecção e quantificação de transcritos raros (IOANNIDIS, 2005; ROBLES *et al.*, 2012). Apesar do RNA-seq ser padrão ouro para análise de transcriptoma, alguns aspectos importantes devem ser levados em consideração durante as análises, como vieses, artefatos do processo de sequenciamento, falsos positivos, quantificação imprecisa de genes pouco expressos, entre outros. Estes aspectos estão discutidos no tópico Discussão.

### **Transcriptoma e a qualidade da carne**

A análise global do transcriptoma é um novo nicho científico entre a genética e a fisiologia, e é capaz de gerar novas hipóteses biológicas que, por sua vez, podem ser estudadas mais profundamente e com abordagens mais focadas (ANDERSSON & GEORGES, 2004). Isto impactará principalmente a descrição de características complexas, as quais são governadas por meio de interações entre muitos genes de efeitos pequenos como a qualidade da carne. Esta, por sua vez, depende das características do tecido conjuntivo, conteúdo de lipídeos, composição das fibras, nível de proteólise muscular e todos os aspectos biológicos complexos por si mesmos (OUALI, 1990). Embora vários estudos de transcriptomas de diferentes tecidos tenham sido conduzidos em humanos e camundongos (MORTAZAVI *et al.*, 2008; TOUNG *et al.*, 2011; WANG *et al.*, 2008; BOTTOMLY *et al.*, 2011), nosso entendimento sobre animais de produção ainda é muito limitado.

Por muito tempo as pesquisas relacionadas com a carne foram direcionadas para o descobrimento de marcadores moleculares e biológicos de qualidade da carne (principalmente maciez) que: tornassem possível a classificação, logo após o abate, dos cortes de carne, baseada em seu potencial final de maciez; e a otimização de seleção genética de animais de corte, baseada em características da qualidade da carne. No entanto, tais marcadores ainda não foram identificados provavelmente devido ao entendimento extremamente limitado dos mecanismos biológicos responsáveis pela

remodelação *postmortem* da textura da carne (OUALI *et al.*, 2013). Em resumo, a análise global do transcriptoma do músculo esquelético de bovinos tem o potencial de esclarecer a complexidade da arquitetura gênica e servir como resultado intermediário dos fatores genéticos que contribuem com a qualidade da carne. Estratégias de seleção indireta baseadas nesses fenótipos moleculares intermediários poderiam ser muito úteis para o melhoramento genético de características relacionadas à qualidade da carne (CHEREL *et al.*, 2012). Abaixo, segue uma revisão sobre alguns genes potenciais marcadores de qualidade da carne, no qual a Tabela 4, no tópico Discussão, resume características destes genes identificadas neste trabalho.

#### *Leptina (LEP) e receptor da leptina (LEPR)*

O gene da leptina (NM\_173928) , também conhecido como o “gene da obesidade”, localizado no cromossomo 4 de bovinos, codifica a leptina, uma citocina que é sintetizada e secretada pelo tecido adiposo (ZHANG *et al.*, 1994; STONE *et al.*, 1996), e tem papel importante na regulação do balanço energético. A concentração de leptina pode ser vista como um indicador de marmoreio, profundidade de toucinho, deposição de gordura, eficiência de reprodução, ingestão de alimentos e grau de rendimento e qualidade em gado confinado (GEARY *et al.*, 2003; MCFADIN *et al.*, 2003, HOUSEKNECHT *et al.*, 1998; LAGONIGRO *et al.*, 2003; PANNIER *et al.*, 2009; SCHENKEL *et al.*, 2005). Ainda, foi mostrado que marcadores em regiões promotoras do gene bovino da leptina estavam associados com o marmoreio e a qualidade da carcaça (NKRUMAH *et al.*, 2005). Apesar da sua grande importância, a LEP ainda não atingiu as expectativas que os cientistas da área da produção animal almejavam para um marcador robusto e confiável de gordura corporal e de carcaça. Mas levando em consideração a sua potencial importância em características relacionadas à qualidade da carne, estudos buscando associações mais robustas entre o gene da LEP com estas características são extremamente promissores. Polimorfismos no gene da LEP e em genes de seus receptores podem fornecer ferramentas poderosas para auxiliar a seleção e, conseqüentemente, o melhoramento de características que afetam a qualidade da carne e com impacto ambiental minimizado (WYLIE, 2011). O receptor da

leptina, codificado pelo gene LEPR (NM\_001012285), é um receptor transmembranar de um único domínio que inibe a via orexígena e estimula a via anorexígena no hipotálamo. O receptor da leptina desempenha um papel importante no controle de consumo de alimento, secreção do hormônio de crescimento, homeostase energética e mobilização de gordura (BARB *et al.* 2006; BELL *et al.* 2005).

#### *Diacilglicerol O-aciltransferase 1 (DGAT1)*

O gene DGAT1 (NM\_174693) localizado no cromossomo 14 de bovinos, codifica a enzima microssomal diacilglicerol O-aciltransferase, que por sua vez catalisa a etapa final da síntese de triglicerídeos. Este gene foi associado a teor de gordura no leite (GRISART *et al.*, 2002; WINTER *et al.*, 2002), teor de gordura intramuscular e adiposidade da carcaça (THALLER *et al.*, 2003, ZHENGRONG *et al.*, 2003). Desta maneira, o DGAT1 poderia ser utilizado como marcador no melhoramento de característica como deposição de gordura.

#### *Calpaína (CAPN) e Calpastatina (CAST)*

O sistema da CAPN foi considerado o principal mecanismo envolvido na maciez da carne (WHEELER, *et al.*, 1996). Este sistema consiste de duas proteases cálcio dependentes calpaína 1 e calpaína 2, codificadas pelos genes CAPN1 (NM\_174259) e CAPN2 (NM\_001103086), respectivamente, e um polipeptídeo chamado calpastatina, codificado pelo gene CAST (NM\_001030318, NM\_001030319, NM\_001030320, NM\_174003), cuja a função é inibir ambas as CAPNs. Calpaínas são provavelmente as proteases mais estudadas na área da maciez da carne e é amplamente aceito que a atividade proteolítica das CAPNs contribui para o processo de amaciamento da carne (KOOHMARAIE & GEESINK, 2006; SENTANDREU *et al.*, 2002). Um estudo que fortaleceu a importância das CAPNs com relação à maciez da carne, foi a identificação de dois SNPs não sinônimos associados com maciez (PAGE *et al.*, 2002) Diferentemente, foi verificado que o alto nível de CAST está relacionado com a

diminuição da maciez da carne (SHACKELFORD *et al.* 1991), e que polimorfismos no gene CAST estão relacionados com capacidade de retenção de água e coloração em bovinos e suínos (CIOBANU *et al.*, 2004; REARDON *et al.*, 2010).

#### *Estearoil-CoA desaturase (SCD)*

SCD é a enzima responsável pela conversão de ácidos graxos saturados em ácidos graxos monoinsaturado em adipócito de mamíferos. Os ácidos graxos monoinsaturados estão relacionados com a maciez da gordura, sabor da carne e diminuição da concentração de colesterol (LDL) circulante (MELTON *et al.*, 1982; SMITH 1994; RUDEL *et al.*, 1995). Os fatores genéticos que determinam o perfil de ácidos graxos ainda estão sendo definidos. No entanto, dado o papel determinante na oxidação de ácidos graxos, o gene da SCD (NM\_173959) é um bom candidato para ser associado à variação da composição de ácidos graxos. YANG e colaboradores (1999) apresentaram correlações interessantes entre a atividade da enzima SCD e a composição de ácidos graxos em uma pesquisa com tecido adiposo de bovinos.

#### *Fosfoglicomutase (PGM)*

Fosfoglicomutase é uma enzima membro da via glicolítica com atuação na primeira fase do processo de glicólise. Nesta fase, ocorre a conversão de glicose-1-fosfato, gerado a partir de glicose livre ou a partir do glicogênio por fosforilase, para dois fosfatos de triose. HAMILL e colaboradores (2012) associaram valores de expressão gênica diferencial do gene PGM (NM\_001076903) com maciez e teor de gordura intramuscular de suínos.

#### *Tiroglobulina (TG)*

TG (NM\_173883) é o gene que codifica a proteína tiroglobulina, uma proteína sintetizada nas células do folículo tireoidiano que é uma molécula precursora dos hormônios da tireoide, tiroxina e triiodotironina. Os hormônios da tireoide desempenham

um papel importante na regulação do metabolismo e pode afetar o crescimento de adipócitos, diferenciação e homeostase da gordura (CASAS *et al.*, 2005). Alguns estudos mostraram associações deste gene com espessura de gordura (CASAS *et al.*, 2005) e score de marmoreio (GAN *et al.*, 2008), e um teste para um marcador do gene da TG foi patenteado (BARENDSE, 1997) e disponível comercialmente para melhoristas de bovinos na Austrália.

#### *Piruvato quinase (músculo) (PKM)*

Piruvato quinase (PK) é uma enzima que na via glicolítica catalisa a conversão de fosfoenolpiruvato em piruvato, o qual mais tarde é reduzido a lactato em condições anaeróbicas (PÖSÖ & PUOLANNE, 2005). A isoforma de PK que é expressa em músculo é codificada pelo gene PKM (NM\_001205727). FONTANESI e colaboradores (2003) mostraram uma associação entre polimorfismo do gene de uma das isoformas de PKM de suínos com o conteúdo de glicogênio. E SIECZKOWSKA e colaboradores (2010) verificaram expressão gênica diferencial de PKM em suínos, relacionada com o aumento do teor de gordura intramuscular, bem como um progresso mais favorável do metabolismo energético e glicolítico no estágio inicial da transformação do músculo em carne.

#### *Homogentisato oxidase (HGD)*

Homogentisato oxidase é uma das seis enzimas necessárias para o catabolismo dos aminoácidos aromáticos fenilalanina e tirosina. Estes, desempenham um papel importante na síntese de algumas proteínas, hormônio, pigmentos e neurotransmissores. Homogentisato oxidase é codificada pelo gene HGD (NM\_001144852) que foi considerado um gene candidato associado com características de carne e de carcaça. ZHOU e colaboradores (2011) verificaram associação de SNPs no HGD com características de qualidade da carne de bovinos.

#### *Proteína de ligação de ácido graxo (FABP)*

Proteína de ligação de ácido graxo são pequenos polipeptídeos intracelulares encontrados em muitos tecidos, estão envolvidos na transferência e metabolismo de ácidos graxos e são codificados por vários genes, a família multigênica FABP (GOMEZ *et al.*, 2007). JURIE e colaboradores (2007), mostraram que o nível de expressão do mRNA de uma isoforma de FABP (A-FABP) pode ser um dos indicadores metabólicos da habilidade de bovinos em depositar gordura intramuscular de acordo com o tipo de músculo e raça. Também foram associados SNPs no transcrito da isoforma FABP4 com gordura intramuscular (AVILÉS *et al.*, 2013).

### *Miostatina (MSTN)*

Miostatina é um fator de diferenciação de crescimento secretado, membro da família da proteína TGF beta que inibe a diferenciação muscular e o crescimento no processo da miogênese e é codificado pelo gene MSTN (NM\_001001525) (CARNAC *et al.*, 2006). Em ovinos, suínos e bovinos, várias mutações na MSTN ou em sua região promotora foram identificadas na alteração do tamanho do músculo (KAMBADUR *et al.*, 2007; GUIMARAES *et al.*, 2007; HADJIPAVLOU *et al.*, 2008) . Além disso, MSTN também influencia no metabolismo de glicose e no acúmulo de gordura (MCPHERRON LEE, 2002). Decréscimo do conteúdo de gordura intramuscular e proporção de gordura na carcaça também foram descritas para o fenótipo de dupla musculatura em gado, o qual é causado por uma mutação na MSTN (HOCQUETTE *et al.*, 1999). No entanto diferentes mutações na MSTN podem causar diferentes graus de hipermusculatura. Tendo em vista a importância da MSTN, esforços tem sido direcionados para a manipulação de sua expressão com o intuito de se produzir maior quantidade de massa magra (D-B *et al.*, 2009).

### *Proteína quinase, AMP-ativada, gama 3 subunidade não catalítica (PRKAG3) e Proteína quinase, AMP-ativada, beta 2 (PRKAB2)*

O gene bovino PRKAG3 (NM\_001162421) codifica a subunidade AMPK gama 3, uma isoforma da subunidade regulatória gama da *Proteína quinase AMP-ativada*

(AMPK). A AMPK tem um papel na regulação do metabolismo energético e foi mostrado que mutações que afetam os genes das subunidades gama influenciaram a atividade da AMPK (CARLING, 2004). A subunidade gama 3 está envolvida na regulação da atividade da AMPK em músculo esquelético e influencia fortemente o metabolismo de glicogênio (MILAN *et al.* 2000). O conteúdo de glicogênio no músculo está correlacionado com a qualidade da carne devido a sua influência no processo de maturação *postmortem* e pH final. Portanto mutação no gene PRKAG3 pode afetar a qualidade da carne através do conteúdo de glicogênio antes do abate. De maneira semelhante, foi verificada correlação do gene PRKAB2 (NM\_001192328) com características determinantes na qualidade da carne em suínos (YERLE *et al.*, 1996).

#### *Desmina (DES) e Titina (TTN)*

Parte do amaciamento da carne foi atribuído ao rompimento da estrutura altamente organizada do músculo por meio da degradação de proteínas miofibrilares chave como a desmina e a titina. Desmina é o membro músculo específico da família da proteína IF (filamento intermediário) e é codificada unicamente pelo gene DES (NM\_001081575), que é expresso em todos os tecidos musculares (CAPETANAKI *et al.* 2007). Foi documentado que a desmina é degradada mais rapidamente em miofibrilas de amostras com baixa força de cisalhamento e maior capacidade de retenção de água (HUFF-LONERGAN & LONERGAN, 1999; HUFF-LONERGAN, *et al.*, 1996). O maior produto da degradação que geralmente é encontrado no músculo é um polipeptídeo de aproximadamente 38kDA. Este produto de degradação também foi mostrado estar presente em miofibrilas digeridas pela  $\mu$ -calpaína (CARLIN *et al.*, 2006; HUFF-LONERGAN & LONERGAN, 1999; HUFF-LONERGAN *et al.*, 1996). Deste modo, a  $\mu$ -calpaína pode ser, pelo menos em parte, responsável pela degradação da desmina sob condições normais de maturação *postmortem*. A titina ou conectina é uma proteína gigante de aproximadamente 3 mega-Daltons, codificada pelo gene TTN (código de acesso no banco de dados ENSEMBL - ENSBTAT00000061449). Ela é a maior proteína encontrada em tecido de mamíferos e é a terceira mais abundante. Devido a essas e outras características é concebível que a sua degradação *postmortem* no

músculo levaria ao enfraquecimento da estrutura longitudinal do sarcômero miofibrilar e integridade do músculo. Este enfraquecimento, em conjunto com outras mudanças postmortem, poderia elevar o grau de amaciamento. A degradação da timina foi observada em vários estudos (ASTIER *et al.*, 1993; FRITZ & GREASER, 1991; FRITZ *et al.*, 1993; HUFF-LONERGAN *et al.*, 1995) e é importante notar que a sua degradação e não a quantidade absoluta de timina no músculo *postmortem* que contribui com o amaciamento *postmortem* (FRITZ *et al.*, 1993).

#### *Canal de cálcio, dependente de voltagem, subunidade 1 alfa2/delta (CACNA2D1)*

O gene CACNA2D1 (NM\_001205987) codifica um membro da família da subunidade alfa2/delta, uma proteína do complexo do canal de cálcio dependente de voltagem. Em bovinos este gene foi mapeado em uma região genômica de sete QTLs, para peso da carcaça, ganho médio diário (CASAS *et al.*, 2001), estatura (ASWELL *et al.*, 2005), porcentagem de osso e razão carne/osso (GUTIÉRREZ-GIL *et al.* 2009). Outros quatro QTLs que afetam características de qualidade da carne também estavam perto desta região. LI e colaboradores (2007 e 2008) verificaram que um SNP na região 3'UTR do CACNA2D1 poderia ser um marcador para características relacionadas com reprodução, carcaça e qualidade da carne em suínos.

#### *Hormônio do crescimento (GH), receptor do hormônio de crescimento (GHR) e fator de crescimento 1 e 2 tipo insulina (IGF1 e 2)*

Genes envolvidos no controle do eixo somatotrópico, como GH (NM\_180996), GHR (NM\_176608) e IGF1 (NM\_00107782) são bons candidatos a terem influência em características relacionadas com a taxa de crescimento e composição corporal (HALE *et al.*, 2000; THOMAS *et al.*, 2002; FARBER *et al.*, 2006). GH é um gene importante envolvido no crescimento pós natal, desenvolvimento do músculo e da glândula mamária e no metabolismo de lipídeos e carboidratos (AKERS, 2006). GH exerce sua ação por meio de seu receptor (GHR), o qual ativa GH ao ligar-se a ele. Vários estudos associaram polimorfismos nos genes GH e GHR com características da carcaça em

diversas raças (HALE *et al.*, 2000; DI STASIO *et al.*, 2005; GARRET *et al.*, 2008). IGF1 é um hormônio similar a estrutura molecular da insulina e foi reportado que a sua concentração estava negativamente correlacionada com a porcentagem de gordura na carcaça, taxa de deposição de gordura, gordura total da carcaça e espessura de gordura, e positivamente relacionada com a porcentagem de proteína na carcaça (ANDERSON *et al.*, 1998).

#### *Fator miogênico 5 e 6 (MyF5 e MyF6), MYOD1 e Miogenina (MYOG)*

Os fatores miogênicos 5 (NM\_174116) e 6 (NM\_181811) pertencem a uma família de fatores de transcrição que incluem a MYOD1 (NM\_001040478) e a miogenina (NM\_001111325). Eles são essenciais para o início e o desenvolvimento do músculo esquelético. MyF5, em combinação com a MYOD1, determinam a linhagem muscular e foi mapeado QTL para peso ao nascimento, ganho médio de peso diário ao desmame e ganho médio de peso diário (LI *et al.*, 2002a,b). Considerando a importância destes quatro fatores transcricionais na formação do tecido muscular, acredita-se que eles sejam bons genes candidatos para crescimento e características da qualidade da carne (MAAK *et al.*, 2006; VERNER *et al.*, 2007)

#### *Peroxisomo proliferador-ativado receptor gama (PPARG) e peroxisomo proliferador-ativado receptor gama, coativador 1 alfa (PPARGC1A)*

PPARG (NM\_181024) é regulador transcricional crítico de genes que controlam o metabolismo energético, adipogênese e manutenção do estado diferenciado (MEMISOGLU *et al.*, 2003; ROSEN & MACDOUGALD, 2006; XU *et al.*, 1999). Assim, PPARG tem um efeito lipogênico e adipogênico em adipócitos e hepatócitos (KERSTEN, 2001) e também foi mostrado que ele tem efeito considerável no omega-3 (SEVANE *et al.*, 2013). PPARGC1A (NM\_177945) é um coativador do PPARG e tem função chave na ativação de uma variedade de receptores de hormônios nucleares e em fatores de transcrição que regulam homeostase de energia (PUIGSERVER & SPIEGELMAN, 2003). Foi mostrado que este gene regula a expressão de genes

envolvidos no metabolismo oxidativo, adipogênese e gliconeogênese e que ele estaria associado a características organolépticas da carne (SEVANE *et al.*, 2013).

#### *Troponina C tipos 1 e 2 (TNNC1 e 2)*

As proteínas contráteis troponina, que formam um complexo localizado nos filamentos do músculo estriado, estão envolvidas na inibição da atividade da ATPase induzida pelo cálcio da miosina (portanto, na contração muscular), e são expressas de maneira fibra-específica. Os genes TNNC1 (NM\_001034351) e TNNC2 (NM\_001076373) são expressas exclusivamente em fibras de contração lenta e fibras de contração rápida, respectivamente (CORIN *et al.* 1994; MULLEN & BARTON, 2000). Considerando a importância destas proteínas na contração muscular, pode-se inferir que os genes que as codificam são candidatos potenciais para se associar à qualidade da carne.

#### *Fator potenciador de miócitos 2A (MEF2A)*

Os genes MEF2 (fator potenciador específico de ligação de miócitos 2) pertencem a família MADS (*MCM1 Agamous, Deficiens, Serum Response Factor*) de fatores de transcrição, e foi mostrado que eles desempenham um papel importantíssimo na morfogênese e miogênese das células dos músculos esquelético, cardíaco e liso (OLSON *et al.*, 1995). Em vertebrados existem quatro genes MEF2, entre eles está o MEF2A. MEF2 controla a formação do músculo esquelético depois da diferenciação terminal em novas fibras e é responsável pela expressão de genes que codificam proteínas de filamento grosso (HINITS & HUGHES, 2007). Cada vez mais estudos são realizados mostrando a importância dos genes MEF2 no desenvolvimento e diferenciação do músculo esquelético (POTTHOFF *et al.*, 2007; L'HONORE *et al.*, 2007). Existem poucos estudos que relacionam o nível expressão e características do MEF2A com a qualidade da carne. No entanto ZHAO e colaboradores (2011) propuseram um modelo no qual a MyOD e o MEF2A controlam o balanço entre a

miogênese e a adiposidade intramuscular. Outros três trabalhos mostraram em suínos que a atividade de MEF2 poderia aumentar a miogênese e conseqüentemente afetar a qualidade da carne (CREEMERS *et al.*, 2006; HAN *et al.*, 2012). Em bovinos, estudos mostraram que polimorfismos no gene MEF2A estavam associados com o peso corporal (CHEN *et al.*, 2010). Este gene está localizado no cromossomo 21 que possui 2 QTLs associados com peso ao nascimento. Isso implica que MEF2A pode ser considerado um potencial marcador molecular para características da carcaça. Os resultados encontrados neste trabalho contribuem bastante com esta implicação e com a possibilidade do MEF2A ter papel significativo no crescimento muscular. Neste trabalho o gene MEF2A apresentou menor valor significativo de expressão no grupo de menor perfil de crescimento muscular em relação ao grupo com maior perfil de crescimento.

#### *Proteína de ligação do hormônio de crescimento tipo insulina 3 (IGFBP3)*

Como mencionado acima, fatores de crescimento tipo insulina (IGF) são peptídeos promotores de crescimento que tem estrutura homóloga com a insulina. No entanto, diferente da insulina, IGFs circulam no plasma complexados com uma família de proteínas de ligação chamadas de proteínas de ligação do IGF (IGFBP) (RAJARAM *et al.* 1997). Estas proteínas circulam por uma variedade de tecidos e são encontradas em fluídos de tecidos biológicos. IGFBP3 é a maior forma de proteínas de ligação e possui vários papéis endócrinos fundamentais e ainda, podem afetar funções celulares independentemente dos IGFs. Vários estudos investigaram o efeito da concentração do IGFBP3 no desempenho de suínos na produção de carne. WALTON e colaboradores (1989) verificaram que o IGFBP3 poderia bloquear a degradação de IGF1. OWENS e colaboradores (1999) acharam que IGF1 e IGFBP3 estavam positivamente associados com a taxa de crescimento, com a ingestão voluntária de alimento e com a relação ganho de peso/conversão alimentar. LEE e colaboradores (2002) sugeriram que a concentração de IGFBP3 pode ser utilizada como índice de crescimento em suínos. Em bovinos foi verificado que o nível de expressão do IGFBP3 era significativamente diferente entre bovinos com diferentes perfis de crescimento muscular (LISOWSKI *et al.*, 2013) e também foram associadas características relacionadas com a composição

de lipídeos e propriedades organolépticas da carne (DUNNER *et al.*, 2013). SUN e colaboradores (2003) encontraram diversas características da carcaça associadas a polimorfismos no gene IGFBP3. Neste estudo o gene IGFBP3 apresentou menor valor significativo de expressão para o grupo de menor perfil de crescimento muscular.

Levando em consideração a grande importância do mercado da carne de bovinos, a necessidade e dificuldade de melhoria da qualidade da carne, a relevância do crescimento muscular na qualidade da carne e o enorme potencial dos sequenciadores de nova geração para análise de transcriptoma, o presente trabalho objetivou caracterizar o panorama completo do transcriptoma do músculo *longissimus dorsi* de bovinos da raça Nelore.

## 2. Objetivos

Caracterizar o panorama completo do transcriptoma do músculo *longissimus dorsi* de bovinos da raça Nelore, avaliar toda a sua complexidade, dinâmica e conteúdo e identificar genes com perfis de expressão associados ao crescimento muscular.

## 3. Material e Métodos

Foram utilizadas amostras de 10 animais da raça Nelore provenientes de um rebanho participante do programa de melhoramento genético coordenado pelo grupo Qualitas, para a raça Nelore. Nos rebanhos participantes deste programa, os animais são criados a pasto até a idade de seleção (por volta de 550 dias), que é baseada em características de crescimento e avaliação visual. Após a seleção, cerca de 20% dos machos de cada rebanho são destinados à reposição ou vendidos como reprodutores. O restante dos animais é vendido como boi magro ou terminado na própria fazenda e enviados para o abate. Os animais utilizados neste estudo foram provenientes de uma propriedade localizada no oeste do estado de São Paulo, e faziam parte de um grupo de 118 animais não selecionados para reprodutores e que, após mais um período de 100 dias a pasto e cerca de 90 a 100 dias em confinamento, foram abatidos em um frigorífico comercial, com idades entre 720 e 760 dias. Todos os 118 animais eram

machos nascidos nos meses de agosto e setembro de 2010 e estavam divididos em dois grupos de manejo na propriedade. A escolha dos 10 animais que participaram do estudo foi baseada no peso aos 550 dias e na área de olho de lombo (corrigidos para grupo de contemporâneo e padronizados para média zero e variância um). O índice para classificação dos animais considerou ponderação de 70% para peso aos 550 dias e 30% para área de olho de lombo. Foram escolhidos os cinco animais com maior (Grupo 1) e os cinco com menor (Grupo2) valor para o índice. Amostras do músculo LD para extração de RNA foram coletadas de cada animal ainda na linha de abate.

### **Extração de RNA mensageiro com Dynabeads® mRNA DIRECT™ Kit (Life Technologies, Carlsbad, CA, EUA)**

Foi realizada a extração dos RNAs mensageiros (mRNAs) provenientes das amostras coletadas dos músculos LD por meio do sistema de purificação Dynabeads® mRNA DIRECT™ kit seguindo as recomendações do fabricante. Este sistema de extração é baseado no pareamento da cauda de poli(A) da extremidade 3' da maioria dos mRNAs com os oligonucleotídeos (dT)<sup>25</sup> covalentemente associados às esferas metálicas chamadas comercialmente de Dynabeads® (JAKOBSEN, 1994). Durante o processo de purificação as moléculas de mRNA ficam presas às esferas metálicas, que por sua vez ficam presas a um ímã, separando assim outras moléculas que não são de interesse, como DNA, proteínas, debris celulares e outros tipo de RNA. A quantificação dos mRNAs provenientes dessa purificação foi feita através do Qubit® RNA Assay Kit utilizando o fluorímetro Qubit® 2.0 Fluorometer (ambos da Life Technologies, Carlsbad, CA, EUA). O manual do Dynabeads® mRNA DIRECT™ Kit pode ser encontrado no endereço:

[http://tools.lifetechnologies.com/content/sfs/manuals/DynabeadmRNADIRECTMicro\\_UG\\_Rev004\\_20120514.pdf](http://tools.lifetechnologies.com/content/sfs/manuals/DynabeadmRNADIRECTMicro_UG_Rev004_20120514.pdf).

### **Construção da biblioteca de cDNA com Ion Total RNA-Seq Kit v2 (Life Technologies, Carlsbad, CA, EUA)**

Para o sequenciamento de nova geração chamado de sequenciamento paralelo em massa, no qual são sequenciados milhares de pequenos fragmentos de cDNA ao mesmo tempo, é necessária a construção da biblioteca de cDNA. Foi utilizado o Ion Total RNA-Seq Kit v2 seguindo as recomendações do fabricante. Foi feita conversão do mRNA purificado no passo anterior em biblioteca de cDNA para sequenciamento fita específica, ou seja, mantendo a informação de orientação da transcrição na molécula de DNA. Durante a construção da biblioteca foram feitos os seguintes procedimentos: fragmentação do RNA utilizando RNase III, ligação de oligonucleotídeos chamados de adaptadores nas extremidades dos RNAs, reação de transcrição reversa, amplificação do produto da transcrição reversa por PCR. Todos os passos de verificação de quantidade, rendimento e distribuição dos fragmentos da biblioteca foram feitos pelo Agilent® DNA 1000 Kit utilizando o 2100 Bioanalyzer (ambos da Agilent, Santa Clara, CA, EUA).

Durante a ligação dos adaptadores foram utilizados, para cada amostra, marcadores moleculares individuais contendo aproximadamente 8 pb, chamados comercialmente de código de barras. Desta maneira, cada amostra recebeu um código de barra específico, o que permitiu que amostras diferentes fossem misturadas e sequenciadas na mesma corrida. Posteriormente, durante as análises preliminares de bioinformática, os resultados do sequenciamento de cada amostra foram individualizados e analisados separadamente. O manual do Ion Total RNA-Seq Kit v2 pode ser encontrado no endereço: <http://ioncommunity.lifetechnologies.com/docs/DOC-2999>.

### **Amplificação da biblioteca por PCR em emulsão (ePCR) com o Ion PI™ Template OT2 200 Kit v2 (Life Technologies, Carlsbad, CA, EUA)**

Amplificação de bibliotecas tem como objetivo gerar em um pequeno espaço físico milhares de cópias de cada fragmento de DNA produzidos na etapa anterior. O objetivo dessa amplificação é aumentar a fonte de sinal iônico que será detectado na etapa de sequenciamento. Para tanto foi utilizada a técnica de PCR de emulsão no qual são criados milhares de micro reatores em uma emulsão de óleo. Esses reatores

contém todos os reagentes necessários para uma reação de PCR e pequenas esferas cobertas com sequências de oligonucleotídeos complementares ao adaptador ligado ao fragmento. Esses fragmentos servem para fixar as moléculas alvos na esfera e servem também como primer para a reação de PCR. No fim dessa etapa foram obtidas milhares de esferas contendo, em cada uma delas, milhares de cópias de cada um dos fragmentos da biblioteca associadas à sua superfície. O manual do Ion PI™ Template OT2 200 Kit v2 pode ser encontrado no endereço: <http://ioncommunity.lifetechnologies.com/docs/DOC-7457>.

### **Preparo da reação de sequenciamento com Ion PI™ Sequencing 200 Kit v2 (Life Technologies, Carlsbad, CA, EUA)**

Nessa etapa do protocolo os fragmentos associados às esferas gerados no passo anterior foram preparados para a reação de sequenciamento e foram depositados nos chips semicondutores. Os chips, por sua vez, foram carregados na plataforma Ion Proton. Essa etapa também engloba a lavagem do equipamento e o preparo das soluções de lavagem necessárias na reação. O manual do Ion PI™ Sequencing 200 Kit v2 pode ser encontrado no endereço: <http://ioncommunity.lifetechnologies.com/docs/DOC-7459>.

### **Análise dos resultados**

As sequências geradas pelo Ion Proton foram previamente analisadas no Torrent Server (servidor associado ao sequenciador) por meio do software Torrent Suit™. Essa análise inicial tem como objetivo verificar a quantidade e a qualidade dos dados gerados durante a corrida de sequenciamento. Cada um dos fragmentos sequenciados foi submetido a alinhamento ao genoma de referência do *bosTau7* (<http://genome.ucsc.edu/cgi-bin/hgGateway?clade=mammal&org=Cow&db=bosTau7>) por meio dos programas Bowtie (LANGMEAD *et al.*, 2009) e TopHat2 (KIM *et al.*, 2003).

Bowtie é um programa de alinhamento extremamente rápido e eficiente desenvolvido para alinhar grande quantidade de pequenos fragmentos de DNA/cDNA a grandes genomas. Como parte do processo de alinhamento, o genoma de referência foi

indexado com um índice Burrows-Wheeler (BURROWS & WHEELER, 1994), o que permitiu grande diminuição no tempo de análise. No entanto, Bowtie não permite alinhamento entre fragmentos que contenham espaçamentos grandes em relação ao genoma, e deste modo, ele não pode alinhar os fragmentos que são separados por íntrons. Para resolver esta limitação e, portanto, identificar eventos de *splice*, foi utilizado o programa TopHat2. Este programa utiliza mecanismo de alinhamento do Bowtie e quebra os fragmentos que o Bowtie não conseguiu alinhar em pequenas porções chamadas de segmentos. Geralmente, essas porções, quando processadas independentemente irão se alinhar ao genoma, mas quando vários segmentos de um único fragmento são alinhados em regiões distantes (entre 100 pb e várias centenas de quilobases) do genoma, TopHat2 infere que o fragmento está em uma junção de *splice* e estima onde essa junção de *splice* está. Ao processar cada um dos fragmentos inicialmente não alinhados, o TopHat2 pode construir um índice de regiões de *splice* do transcriptoma sem a necessidade de anotação de genes ou regiões de *splice*. Esta capacidade é crucial, já que vários estudos de RNA-seq mostraram que o nosso atual catálogo de eventos de *splicing* alternativo ainda está muito incompleto. Mesmo em transcriptomas de organismos modelos muito estudados, novos eventos de *splicing* são descobertos a cada novo estudo de RNA-seq. A montagem dos transcritos a partir das sequências dos fragmentos alinhados, bem como estimativas da quantidade de transcritos, a regulação gênica e descoberta de novos transcritos foram feitas por meio do programa Cufflinks (TRAPNELL *et al.*, 2010).

A quantidade de transcritos foi medida pelo número de fragmentos por quilobase por milhão de fragmentos mapeados (FPKM) (MORTAZAVI *et al.*, 2008). Para calcular o nível correto de expressão de cada transcrito, foram feitas duas normalizações pelo Cufflinks: normalização pelo tamanho do transcrito e normalização pela quantidade total de fragmentos sequenciados. Na primeira, foram contados os fragmentos que se alinharam a cada transcrito, e então esta contagem foi normalizada pelo tamanho do transcrito. Na segunda, para comparar o nível de expressão de cada transcrito entre amostras, a contagem foi normalizada pelo rendimento total do sequenciamento ao longo do experimento para cada amostra. FPKM incorpora esses dois passos de

normalização para garantir que os níveis de expressão de diferentes genes e transcritos possam ser comparados entre diferentes corridas.

FPKM = fragmentos por quilobase por milhão  
= [número de fragmentos mapeados]/[tamanho do transcrito em quilobase]/[milhão de fragmentos mapeados]  
= [número de fragmentos mapeados]/([tamanho do transcrito]/1000)/([número total de fragmentos]/10<sup>6</sup>)

Para a quantificação de transcritos provenientes de *splicing* alternativo por meio do Cufflinks, foi implementado um modelo estatístico linear para estimar a abundância de cada transcrito. Depois de calculado o valor de expressão de cada isoforma de um gene, o valor de expressão deste gene foi calculado a partir da soma dos valores de expressão de cada um de seus transcritos. Isso é possível pois FPKM é diretamente proporcional a abundância do transcrito. De fato, o valor de expressão de qualquer grupo de transcritos (como um grupo de transcritos que compartilham um mesmo promotor) pode ser seguramente computado por meio da soma dos níveis de expressão dos membros daquele grupo.

Para auxiliar a descoberta de novos genes, foi utilizado o algoritmo de alinhamento Basic Local Alignment Search Tool para busca de similaridade de sequências de nucleotídeos (blastn). Todas as sequências dos genes que foram considerados potenciais novos genes (PNG) foram submetidas a uma busca de similaridade a sequências de nucleotídeos depositadas no banco de dados *database Nucleotide collection (nr/nt)*. Dos resultados de alinhamento, foi selecionado somente o alinhamento com maior valor de significância de similaridade para cada PNG. Nenhum parâmetro do blastn foi alterado para as buscas de similaridade.

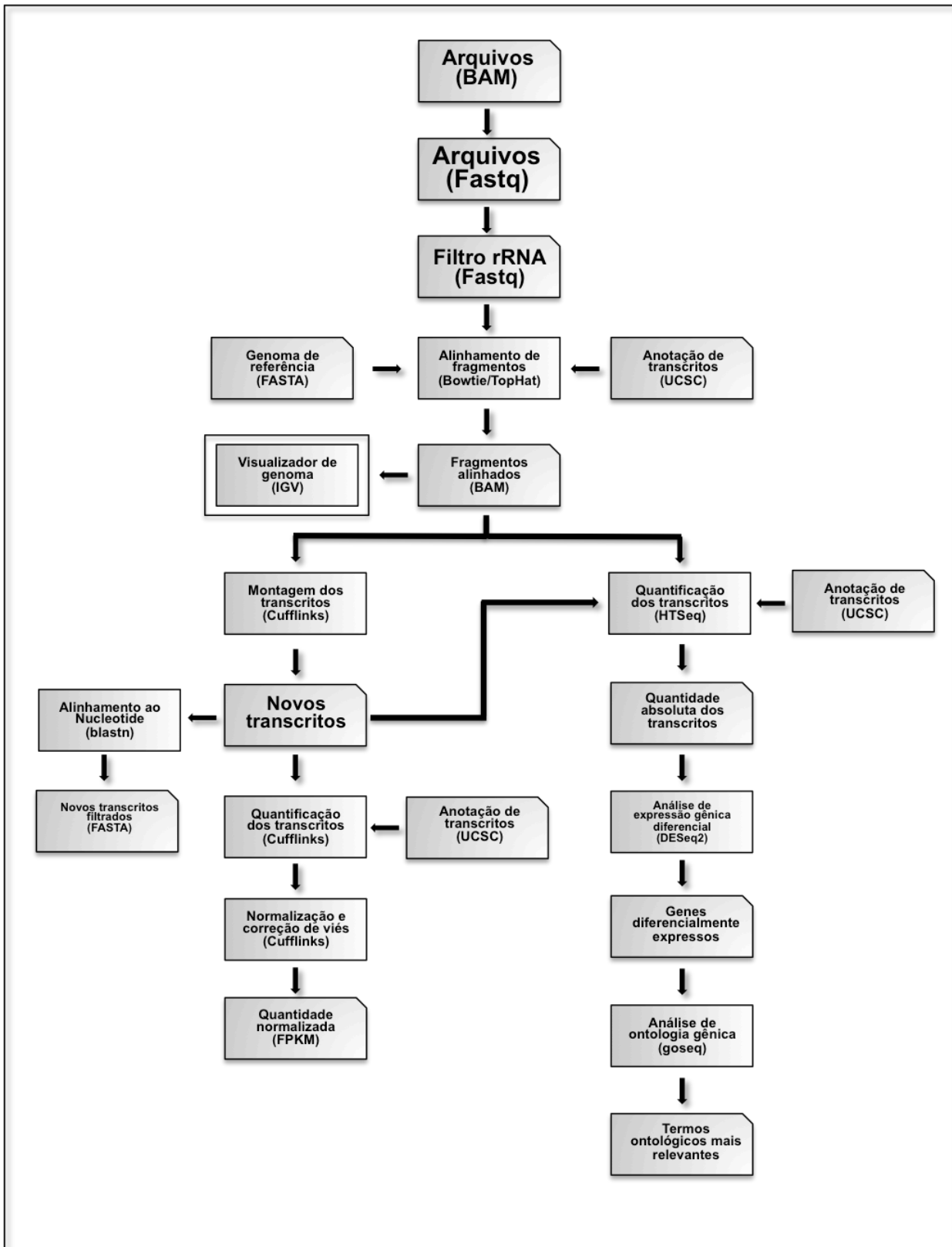
Para análise de expressão gênica diferencial entre os grupos foi utilizado o pacote DESeq2 para ambiente R (ANDERS & HUBER, 2010) . O DESeq2 utiliza como

---

entrada valores brutos da contagem dos fragmentos, os quais são importantes para o seu modelo estatístico. Para fazer a contagem bruta dos fragmentos foi utilizado o programa HTSeq. Diferente do Cufflinks que faz a contagem normalizada dos fragmentos (FPKM).

A partir dos resultados de expressão gênica diferencial foi utilizado o pacote para ambiente R, *GOseq* para identificação dos termos ontológicos mais relevantes entre os genes diferencialmente expressos. Este pacote utiliza o método "Wallenius" que aproxima a distribuição real dos números dos membros de uma categoria entre os genes diferencialmente expressos pela distribuição hipergeométrica não central Wallenius. Esta distribuição assume que dentro de uma categoria todos os genes tem a mesma probabilidade de ser escolhido.

Foi utilizado o "visualizador de genoma" (*genome browser*) *Integrative Genomics Viewer* (IGV) (ROBINSON et al., 2011) para auxiliar as análises dos dados, interpretação dos resultados e identificação visual de potenciais artefatos. A Figura 4 representa de maneira simplificada o fluxograma utilizado nas análises dos dados.



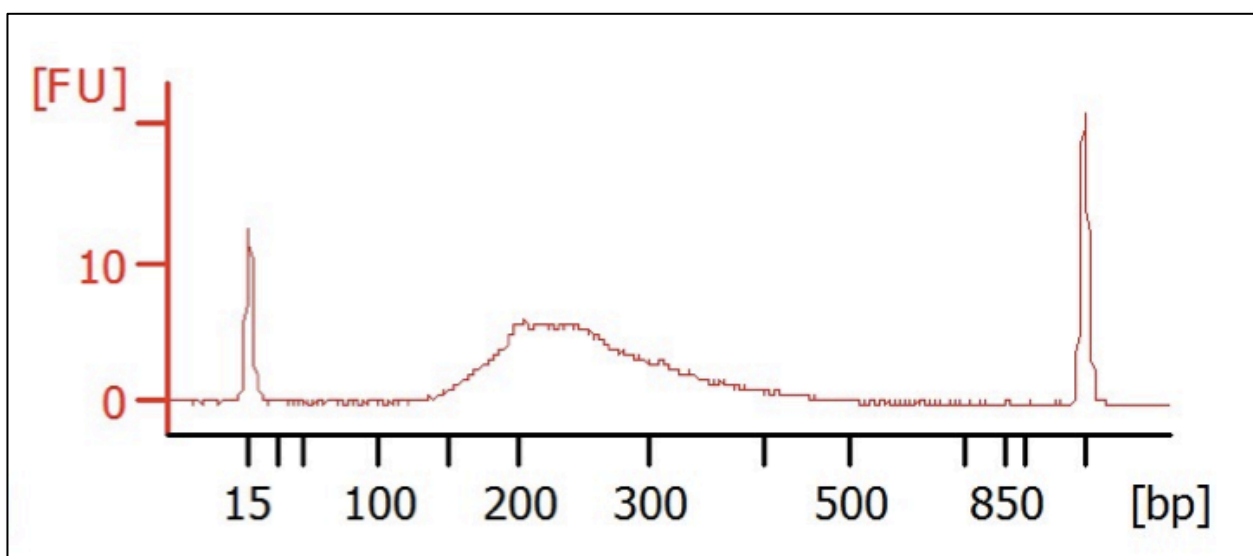
**Figura 4.** Fluxograma das análises dos dados do sequenciamento. Retângulos inteiros representam programas computacionais e os retângulos com um dos vértices cortados representam arquivos. As setas indicam a direção do fluxo das análises. O arquivo de formato *.bam* ou arquivo BAM é a versão binária do arquivo SAM (*Sequence Alignment/Map*), que por sua vez é o formato genérico para se armazenar grandes sequências alinhadas da nucleotídeos. Arquivo FASTA (*.fasta*) tem formato baseado em texto para representar sequência de nucleotídeos utilizando letras como código. Bowtie e TopHat são softwares de alinhamento de fragmentos sequenciados ao genoma de referência e de detecção de eventos de *splice*. Cufflinks é o programa responsável pela montagem dos transcritos, estimativas de quantidade de transcritos e descoberta de novos transcritos. O arquivo FPKM (fragmentos por quilobase por milhão) contém as quantidades dos transcritos normalizadas pelo tamanho deles e pela quantidade total de fragmentos sequenciados. O banco de dados do UCSC foi utilizado para fazer a identificação de transcritos anotados. HTSeq também é um programa de quantificação de transcritos e os dados provenientes de seu resultado são utilizados como entrada para o pacote DESeq2. DESeq2 é um pacote de ambiente R responsável pela análise de expressão gênica diferencial. IGV é um programa auxiliar de visualização de genomas. Goseq é um pacote de ambiente R utilizado para identificação dos termos ontológicos mais relevantes entre os genes diferencialmente expressos. Blastn é uma ferramenta de alinhamento utilizada para verificar similaridade entre sequências alvos e diversos bancos de dados de nucleotídeos.

#### 4. Resultados

##### Rendimento da extração dos mRNAs e da construção das bibliotecas de cDNA

Por meio do sequenciamento de nova geração, este estudo gerou e comparou o panorama do transcriptoma do músculo LD de bovinos da raça Nelore com dois perfis de crescimento muscular. Para isso, mRNAs das dez amostras do músculo LD foram extraídos e purificados com concentração média de  $1,7 \text{ ng}/\mu\text{L} \pm 0,607$ , sendo ressuspensos em  $20 \mu\text{L}$  de água. Em seguida, bibliotecas de mRNA de cada amostra foram construídas a partir de  $8 \mu\text{L}$  de cada produto da etapa de purificação. A quantidade média resultante da construção das bibliotecas foi de  $17,78 \text{ ng}/\mu\text{L} \pm 2,01$ . Durante o processo de construção das bibliotecas, são gerados milhões de fragmentos de diferentes tamanhos, no geral de 50 a 1000 pb. No entanto, apenas uma porção desses fragmentos (até 200 pb) podem ser utilizados durante a amplificação por PCR em emulsão. Portanto, para a obtenção de um bom rendimento durante a amplificação são necessárias bibliotecas com quantidade suficiente de fragmentos com até 200 pb. Uma maneira para verificar quantidade e a distribuição de tamanho de fragmentos, é

por meio de eletroforese. A Figura 5 mostra o perfil de migração dos fragmentos de uma das bibliotecas durante a eletroforese no Bioanalyzer com o Agilent® DNA 1000 Kit. Segundo as recomendações do fabricante do Ion Total RNA-Seq Kit v2, os perfis de distribuição dos fragmentos obtidos estavam de acordo com o necessário para bom rendimento da PCR em emulsão.

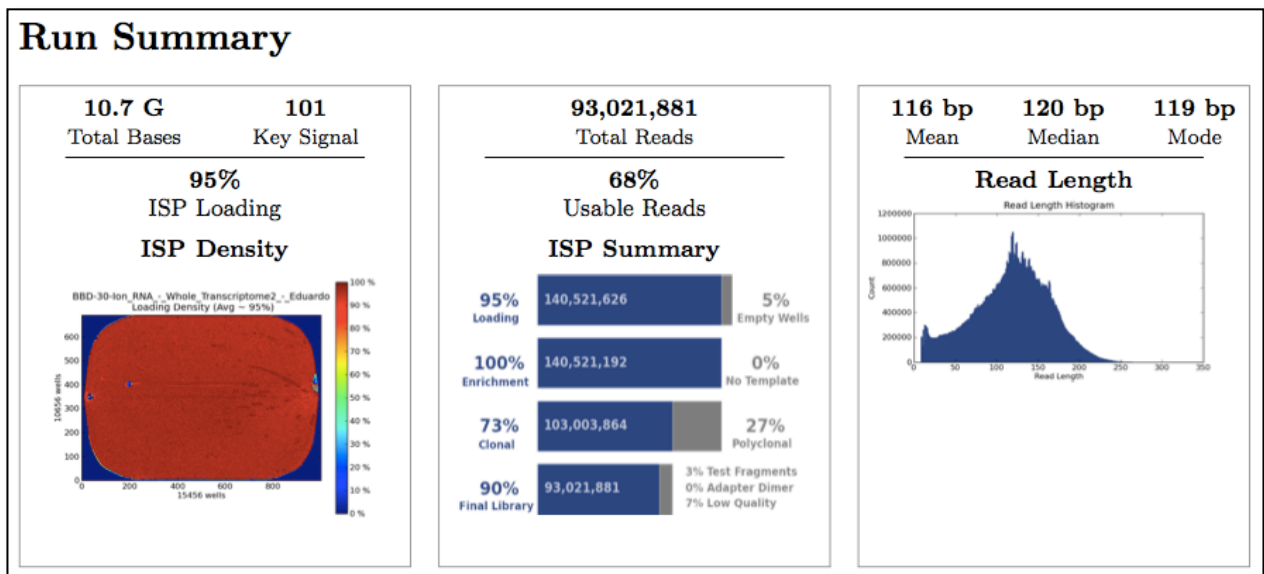


**Figura 5.** Perfil de migração dos fragmentos de uma das bibliotecas em uma eletroforese capilar. No eixo y estão as unidades de fluorescência (FU) e no eixo x os pares de bases dos fragmentos (bp). Os picos das extremidades são referentes a fragmentos de DNA marcadores (componente do kit) com tamanho e quantidade conhecidos. E o “arraste” no centro é referente á migração dos fragmentos da biblioteca que variaram de ~120 a ~600 pb.

### **Sequenciamento do transcriptoma do músculo LD de bovinos da raça Nelore**

Foram realizadas cinco corridas de sequenciamento com duas amostras em cada corrida e mais uma corrida extra repetindo o sequenciamento de duas das dez amostras. Ou seja, oito amostras foram sequenciadas uma vez cada e duas amostras foram sequenciadas duas vezes cada, no total de doze corridas. A corrida extra foi feita com o objetivo de aumentar a quantidade de dados e, conseqüentemente, obter informações mais acuradas.

O software Torrent Suit™, que funciona em conjunto com o Ion Proton™, gera um relatório com alguns dados preliminares após as corridas de sequenciamento. Com este relatório é possível verificar aspectos relacionados a todo o fluxograma da parte laboratorial do experimento, desde a construção da biblioteca até ao sequenciamento propriamente dito. A Figura 6 é referente ao resumo do resultado gerado pelo software Torrent Suit™ de uma corrida de sequenciamento com duas amostras.



**Figura 6.** Resumo do resultado de uma corrida de sequenciamento. O quadro da esquerda mostra a quantidade de bases sequenciadas (10,7 bilhões) e a percentagem de poços do chip que receberam fragmentos da biblioteca (95%). A quantidade de poços preenchidos também é mostrada pelo *heatmap*. O quadro central mostra a quantidade total de fragmentos que foi sequenciada (mais de 93 milhões) e a percentagem desses fragmentos que gerou informação útil para a análise do transcriptoma (mais de 63 milhões). O quadro da direita mostra a distribuição de tamanho de fragmentos de biblioteca que foram sequenciados (média de 116 pb).

## Análise dos dados

### Alinhamento

A Tabela 1 apresenta os resultados gerais do sequenciamento e do alinhamento dos fragmentos ao genoma de referência do *bosTau7*. No total, foram sequenciados mais de 40 bilhões de pares de bases e mais de 450 milhões de fragmentos com média

---

de tamanho de 105 pb ( $\pm 7,53$ ). Do total dos fragmentos sequenciados, 91,21% foram alinhados ao genoma de referência. Destes, 27,6% eram referentes aos RNAs ribossomais e 50,92% foram alinhados unicamente a uma região específica do genoma. Ainda como resultado do alinhamento, em média, foram identificadas 577.499,5 ( $\pm 272.433,14$ ) deleções, 585.113,6 ( $\pm 302.244,63$ ) inserções e 82.992,2 ( $\pm 15.214,91$ ) junções de éxons para as dez amostras.

**Tabela 1. Resumo do sequenciamento e do alinhamento dos fragmentos ao genoma de referência bosTau7.**

	Quantidade	Porcentagem
Quantidade total de pares de bases	40.696.389.940	100
Quantidade total de pares de bases (≥Q20)	29.992.028.105	73,70
Quantidade total de fragmentos	452.929.550	100
rRNA	123.464.188	27,26
Fragmentos alinhados + rRNA	413.124.897	91,21
Fragmentos alinhados - rRNA	289.660.709	63,95
Fragmentos alinhados a uma única posição	230.609.371	50,92
Fragmentos alinhados a múltiplas posições	59.051.338	13,04
Fragmentos não alinhados	21.063.181	8,76

Quantidade total de pares de bases: quantidade de pares de bases sequenciados nas dez corridas. Quantidade total de pares de bases (≥Q20): quantidade de pares de bases sequenciados com *Phred Score* igual ou maior que 20 nas dez corridas. Quantidade total de fragmentos: número total de fragmentos sequenciados nas dez corridas. rRNA: fragmentos alinhados aos RNAs ribossomais. Fragmentos alinhados + rRNA: quantidade total de fragmentos que foram alinhados ao genoma de referencia, considerando os RNAs ribossomais. Fragmentos alinhados - rRNA: quantidade total de fragmentos que foram alinhados ao genoma de referencia, sem considerar os RNAs ribossomais. Fragmentos alinhados a uma única posição: fragmentos que foram alinhados a apenas uma região do genoma de referência sem considerar os RNAs ribossomais. Fragmentos alinhados a múltiplas posições: fragmentos que foram alinhados a mais de uma posição do genoma de referência sem considerar os RNAs ribossomais. Fragmentos não alinhados: fragmentos que não foram alinhados ao genoma de referência.

### *Anotação, montagem e contagem*

Mais de 60.000 montagens realizadas pelo Cufflinks foram classificadas como genes e mais de 79.000 transcritos foram identificados a partir desses genes. Da quantidade total de transcritos, 15.977 (20,12%) eram transcritos já anotados, 10.960

---

(13,80%) foram classificados como potenciais novas isoformas e 51.140 (64,39%) foram classificados como desconhecidos. Na Tabela 2 estão sumarizados os resultados referentes à anotação, montagem e contagem do transcriptoma. Entre os 34.703 genes anotados do genoma de referência de bovinos (BosTau7), 14.876 (43%) genes foram expressos no músculo LD. Destes, 12.082 tiveram valores de FPKM calculado para pelo menos uma amostra, os demais não passaram em um ou mais parâmetros de controle de qualidade do Cufflinks e não tiveram seus valores calculados. Na Tabela 3 estão os 50 genes anotados mais abundantes encontrados neste trabalho.

**Tabela 2. Resumo dos resultados da montagem e anotação gênica.**

	Quantidade	Porcentagem
Quantidade total de genes	60.814	100
Quantidade de genes anotados	14,876	24,46
Quantidade de potenciais novos genes	45.938	75,54
Quantidade total de transcritos	79.427	100
Quantidade de potenciais novas isoformas	10.960	13,80
Quantidade de transcritos completos (conhecidos)	15.977	20,12
Quantidade de transcritos desconhecido, intergênicos	51.140	64,39
Quantidade de transcritos parcialmente alinhados a um éxon	358	0,45
Quantidade de transcritos parcialmente alinhados a um éxon e referência na fita oposta	992	1,25

Quantidade total de genes: quantidade total de montagens classificadas como genes pelo Cufflinks para as dez amostras. Quantidade de genes anotados: quantidade genes sequenciados com anotação no banco de dados UCSC. Quantidade de potenciais novos genes: quantidade de montagens classificadas como genes pelo Cufflinks e sem anotação no banco de dados UCSC. Quantidade total de transcritos: quantidade total de transcritos que foram sequenciados para as dez amostras. Quantidade de potenciais novas isoformas: Quantidade de transcritos não anotados em que, pelo menos uma junção de *splice* é compartilhada com o transcrito de referência. Quantidade de transcritos completos: quantidade de transcritos que tiveram todos os seus éxons alinhados ao transcrito de referência. Quantidade de transcritos desconhecidos, intergênicos: quantidade de transcritos que não foram alinhados a nenhum transcrito de referência. Quantidade de transcritos parcialmente alinhados a um éxon: quantidade de transcritos que foram alinhados a um éxon e suas regiões adjacentes. Quantidade de transcritos parcialmente alinhados a um éxon e referência na fita oposta: quantidade de transcritos que foram alinhados a um éxon e suas regiões adjacentes, porém com referência na fita oposta.

**Tabela 3. Cinquenta genes mais abundantes no músculo LD de bovinos da raça Nelore**

<b>Símbolo</b>	<b>Nome do gene</b>	<b>FPKM</b>
LBH	limb bud and heart development homolog (mouse)	908387,30
TMEM63A	transmembrane protein 63A	578651,00
THBS1	thrombospondin 1	505974,00
TMEM141	transmembrane protein 141	411825,23
DHRS13	dehydrogenase/reductase (SDR family) member 13	205269,40
UBE2C	ubiquitin-conjugating enzyme E2C	161456,28
TMEM63A	transmembrane protein 63A	140460,00
MIR206	microRNA mir-206	102955,15
MYLPF	myosin light chain, phosphorylatable, fast skeletal muscle	87497,19
MIR30C	microRNA mir-30c	78852,91
ACTA1	actin, alpha 1, skeletal muscle	76806,90
CKM	creatine kinase, muscle	64862,56
ALDOA	aldolase A, fructose-bisphosphate	58043,50
TPM2	tropomyosin 2 (beta)	53111,34
TPT1	tumor protein, translationally-controlled 1	52891,68
TNNT3	troponin T type 3 (skeletal, fast)	40228,58
ATP2A1	ATPase, Ca <sup>++</sup> transporting, cardiac muscle, fast twitch 1	32998,42
MYL2	myosin, light chain 2, regulatory, cardiac, slow	30737,00
FHL1	four and a half LIM domains 1	28854,42
MB	myoglobin	28041,66
ACTA1	actin, alpha 1, skeletal muscle	25258,70
TNNT3	troponin T type 3 (skeletal, fast)	24698,43
TPM3	tropomyosin 3	24656,27
MIR380	microRNA mir-380	24150,74
TNNC1	troponin C type 1 (slow)	23382,30
TNNT1	troponin T type 1 (skeletal, slow)	22549,42

---

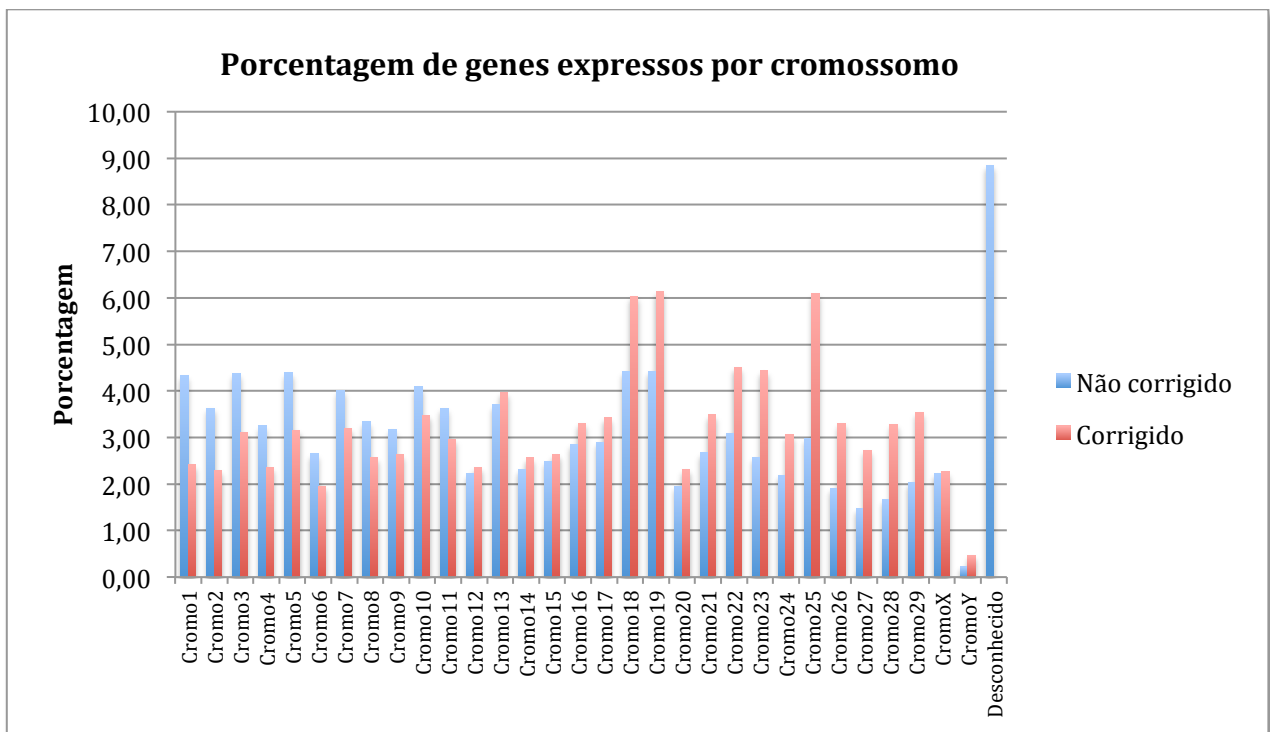
MB	myoglobin	22351,67
ACTA1	actin, alpha 1, skeletal muscle	21975,70
TPM1	tropomyosin 1 (alpha)	20250,15
CSRFP3	cysteine and glycine-rich protein 3 (cardiac LIM protein)	19990,50
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	19922,42
MYL2	myosin, light chain 2, regulatory, cardiac, slow	19745,51
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	18835,60
TVP23B	trans-golgi network vesicle protein 23 homolog B (S. cerevisiae)	18632,40
ACTA1	actin, alpha 1, skeletal muscle	17572,00
NDUFA3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa	17269,99
MYH7	myosin, heavy chain 7, cardiac muscle, beta	17097,44
MIRLET7B	microRNA let-7b	16139,63
DES	desmin	15999,22
MIRLET7A-3	microRNA let-7a-3	15350,47
RPLP1	ribosomal protein, large, P1	15347,94
PGAM2	phosphoglycerate mutase 2 (muscle)	14646,99
MYL2	myosin, light chain 2, regulatory, cardiac, slow	14628,03
ALDOA	aldolase A, fructose-bisphosphate	14610,40
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	14533,48
CSDA	cold shock domain protein A	14291,27
ALDOA	aldolase A, fructose-bisphosphate	14280,10
FTH1	ferritin, heavy polypeptide 1	14235,55
ACTA1	actin, alpha 1, skeletal muscle	13281,80
MYL1	myosin, light chain 1, alkali; skeletal, fast	13096,31

---

FPKM: Soma dos FPKMs das 10 amostras nas 12 corridas.

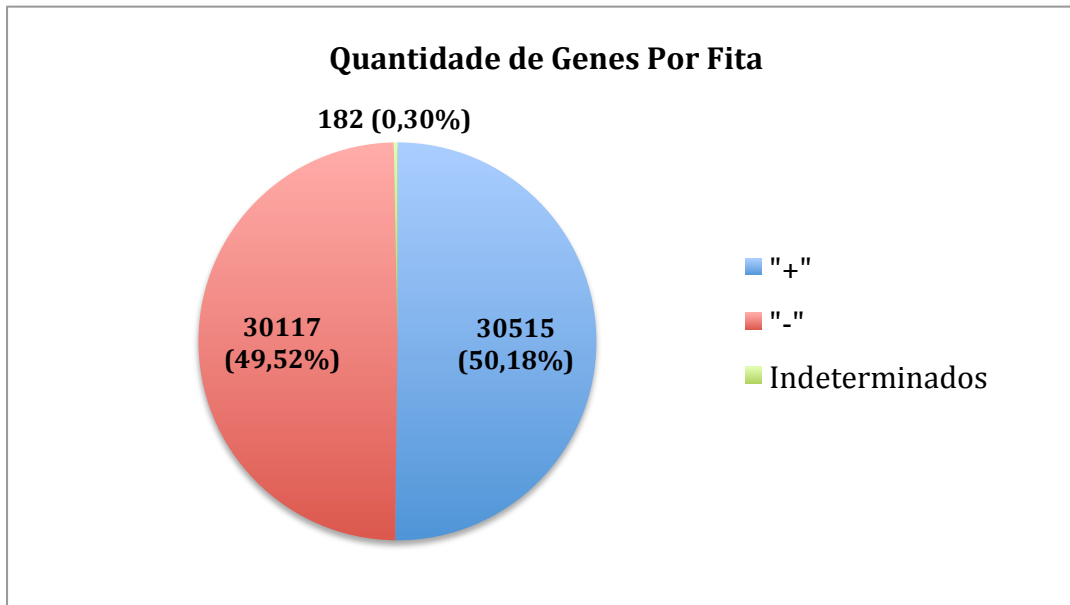
Foi detectada expressão gênica em todos os 29 autossomos e nos 2 cromossomos sexuais. O cromossomo 19 teve maior número de genes expressos

(2.687) e o cromossomo y teve o menor número de genes expressos (137). Dos genes detectados 5.377 (8,84%) foram atribuídos a regiões genômicas de localização ainda não determinada. A Figura 7 mostra a porcentagem de genes expressos por cromossomo. Considerando apenas os genes anotados, a expressão também foi distribuída por todos os cromossomos. O cromossomo 19 também teve a maior quantidade de genes expressos (868), enquanto o cromossomo y teve a menor quantidade (63).



**Figura 7.** Porcentagem de genes expressos por cromossomo. As barras vermelhas são referentes a porcentagem de genes expressos por cromossomo corrigida pelo tamanho do cromossomo. As barras azuis são referentes a porcentagem de genes expressos por cromossomo sem correção.

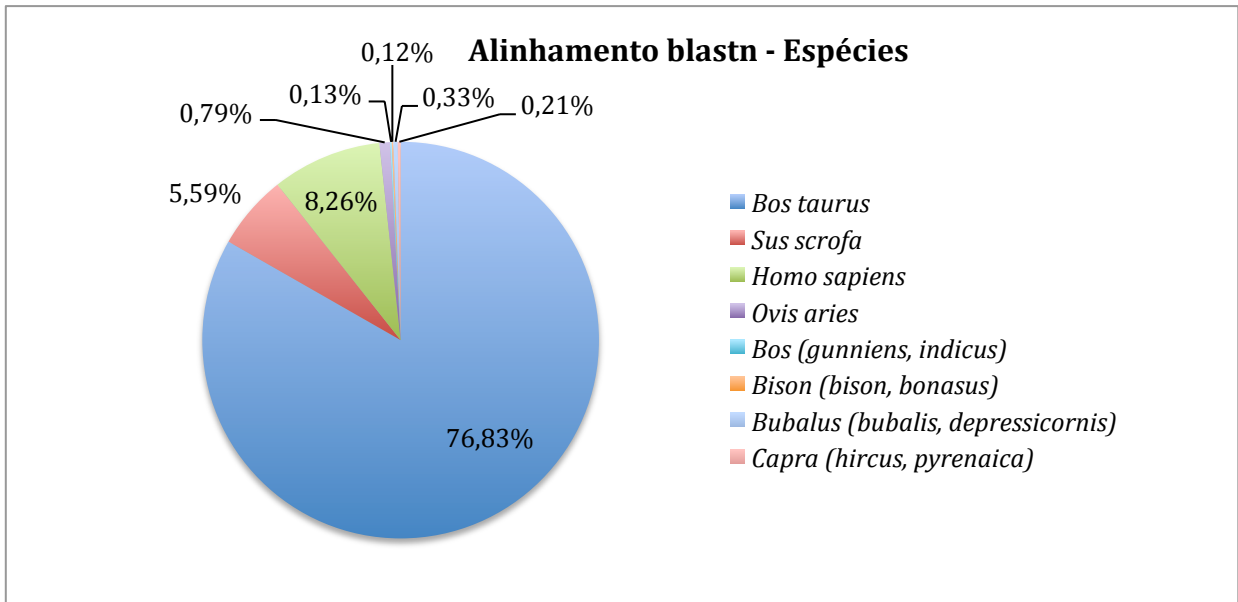
A quantidade de genes expressos a partir das duas fitas de DNA foi bem homogênea. A Figura 8 apresenta a quantidade de genes expressos em cada uma das fitas.



**Figura 8.** Quantidade de genes expressos por fita da molécula de DNA.

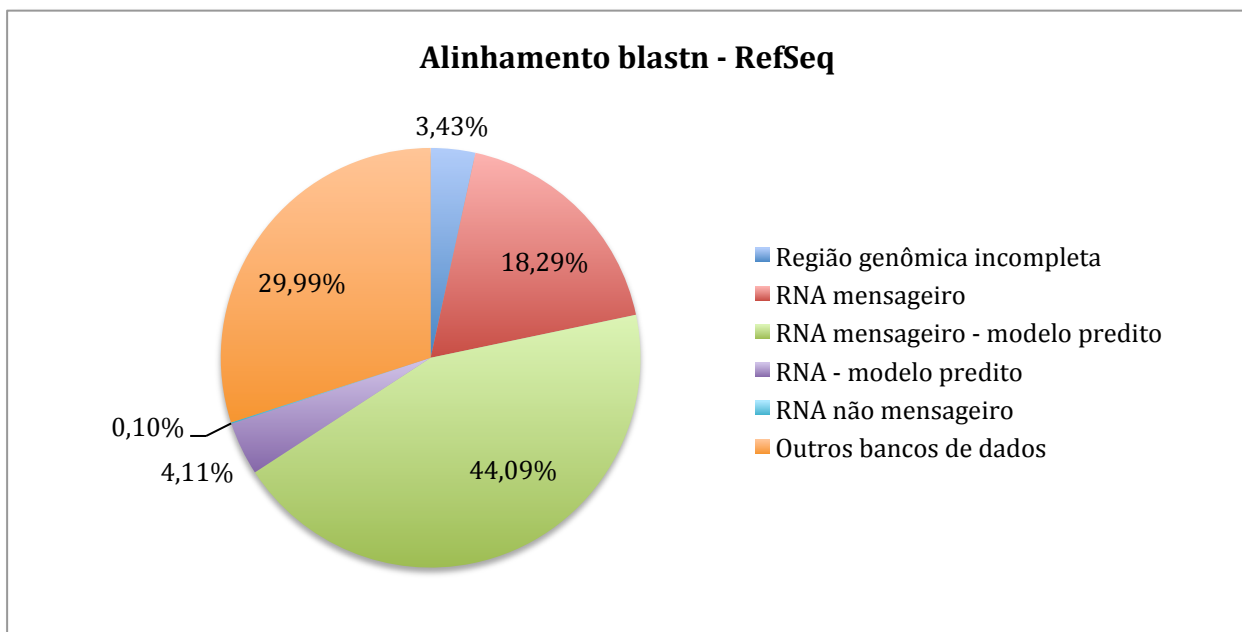
#### *Identificação e análise de potenciais novos genes*

Por meio do software Cufflinks foi detectado um grande número de genes não anotados, denominados potenciais novos genes (PNGs). No total, foram obtidos 45.938 PNGs. Alguns desses PNGs (4,83%) apresentaram mais do que 1 transcrito e o gene com maior quantidade apresentou 35 transcritos. As sequências de nucleotídeos de todos os PNGs foram submetidas a uma busca de similaridade a outras sequências de nucleotídeos depositadas no banco de dados *database Nucleotide collection (nr/nt)* por meio do algoritmo de alinhamento blastn. Dos 45.938 PNGs, apenas 6.065 (12,14%) foram alinhados a outras sequências com valor significativo de similaridade. Na Figura 9 estão discriminadas as porcentagens dos PNGs alinhados a sequências de nucleotídeos de diferentes espécies.



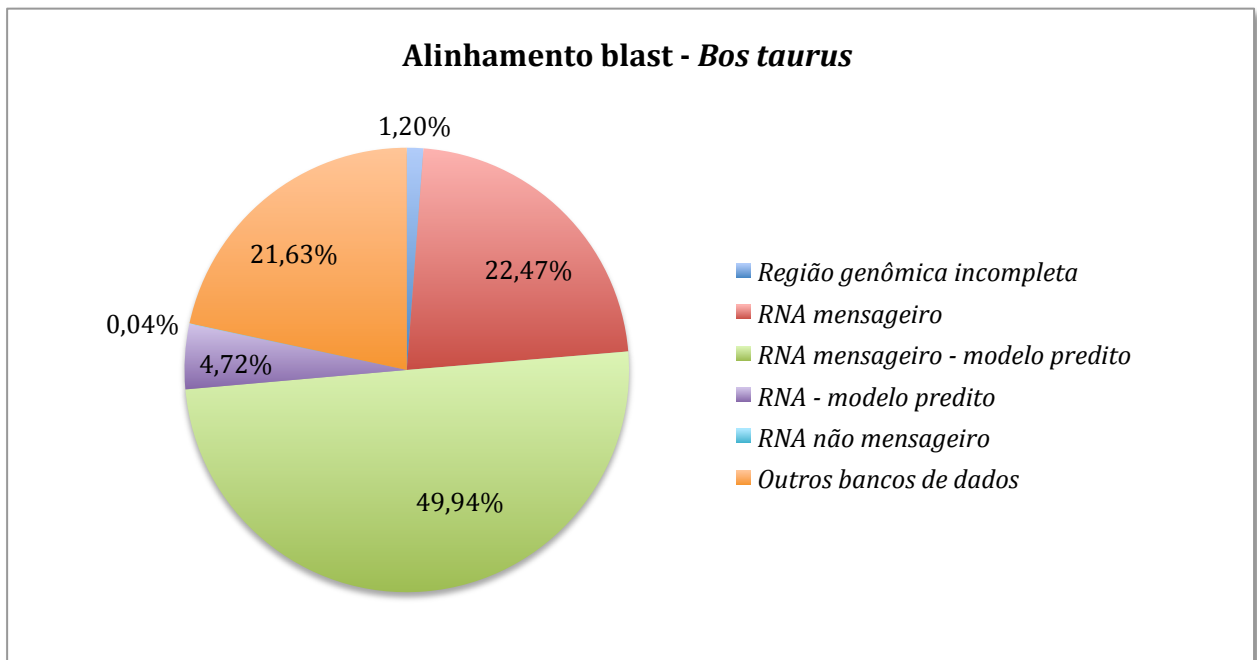
**Figura 9.** Porcentagem de PNGs alinhados a sequências de nucleotídeos depositadas no banco de dados *database Nucleotide collection (nr/nt)* de diferentes espécies, por meio do algoritmo de alinhamento blastn.

Os 6065 resultados de alinhamento foram divididos em PNGs alinhados ao banco de dados curado RefSeq e PNGs alinhados a outros bancos de dados integrantes do *database Nucleotide collection (nr/nt)*. Em seguida, os PNGs alinhados a sequências do RefSeq foram subdivididos de acordo com as classes das sequências depositadas no RefSeq. A Figura 10 mostra o resultado destas divisões.



**Figura 10.** Porcentagem de PNGs alinhados a sequências de nucleotídeos depositadas no banco de dados RefSeq e outros bancos de dados integrantes do *database Nucleotide collection (nr/nt)*. Os PNGs alinhados a sequências do RefSeq foram classificadas de acordo com as classes de sequências depositadas no RefSeq.

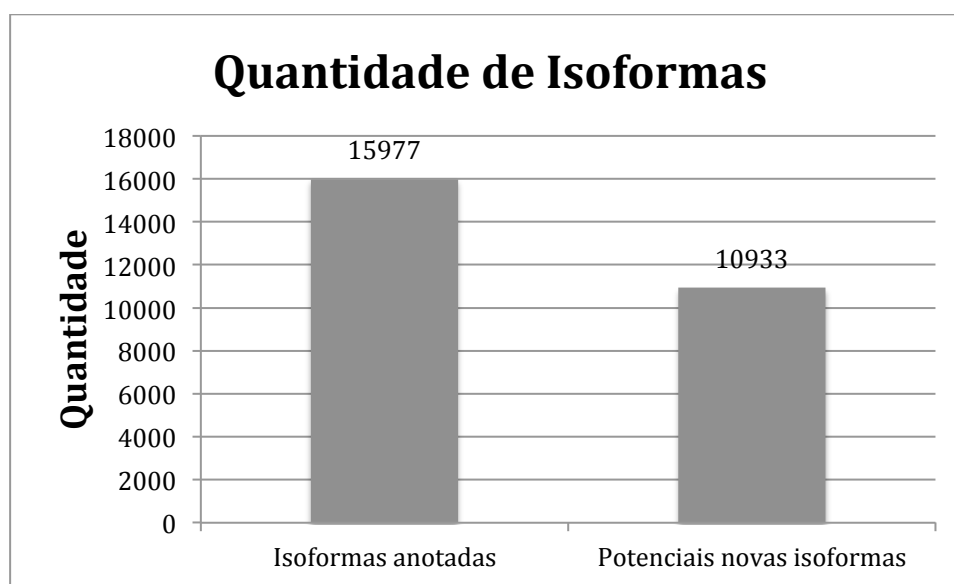
Setenta e sete por cento dos PNGs foram alinhados a sequências de nucleotídeos de *Bos taurus*. Estes foram divididos de acordo com as classes do RefSeq e outros bancos de dados. Quase metade (49,94%) dos PNGs alinhados eram modelos preditos de mRNA (Figura 11).



**Figura 11.** Porcentagem de PNGs alinhados a sequências de nucleotídeos de *Bos taurus* depositadas no banco de dados RefSeq e outros bancos de dados integrantes do *database Nucleotide collection (nr/nt)*. Os PNGs alinhados a sequências do RefSeq foram classificadas de acordo com as classes de sequências depositadas no RefSeq.

### Eventos de *splice*

No total foram identificados 79.427 transcritos. Destes, 51.195 (64,46%) foram provenientes de PNGs e 28.232 (35,54) provenientes de genes anotados. Em média os genes anotados tiveram 1,9 ( $\pm 1,42$ ) transcritos e os PNGs tiveram 1,1 ( $\pm 0,7$ ). Foram identificadas 10.960 potenciais novas isoformas, das quais mais de 99% eram provenientes de genes anotados. Foram consideradas potenciais novas isoformas os transcritos sem anotação e que tinham pelo menos uma junção de splice compartilhada com um transcrito de referência. Na Figura 12 está a quantidade total de isoformas anotadas e a quantidade total de potenciais novas isoformas, ambas provenientes de genes anotados.



**Figura 12.** Quantidade total de isoformas anotadas e a quantidade total de potenciais novas isoformas, ambas provenientes de genes anotados

### *Identificação de genes diferencialmente expressos*

Foi realizada análise de expressão gênica diferencial entre os dois grupos por meio do pacote DEseq2 para ambiente R. Nenhum dos genes teve valor de  $p$  ajustado menor que 0,05. No entanto quando apenas o valor de  $p$  foi considerado, dos 60.814 genes, 391 tiveram valores de  $p$  menores que 0,01. Destes, 110 eram genes anotados (Tabela 3). A Figura 13 mostra a variação do  $-\log_{10}$  dos valores de  $p$  não ajustados em relação ao  $\log_2$  da diferença de expressão para cada um dos genes.

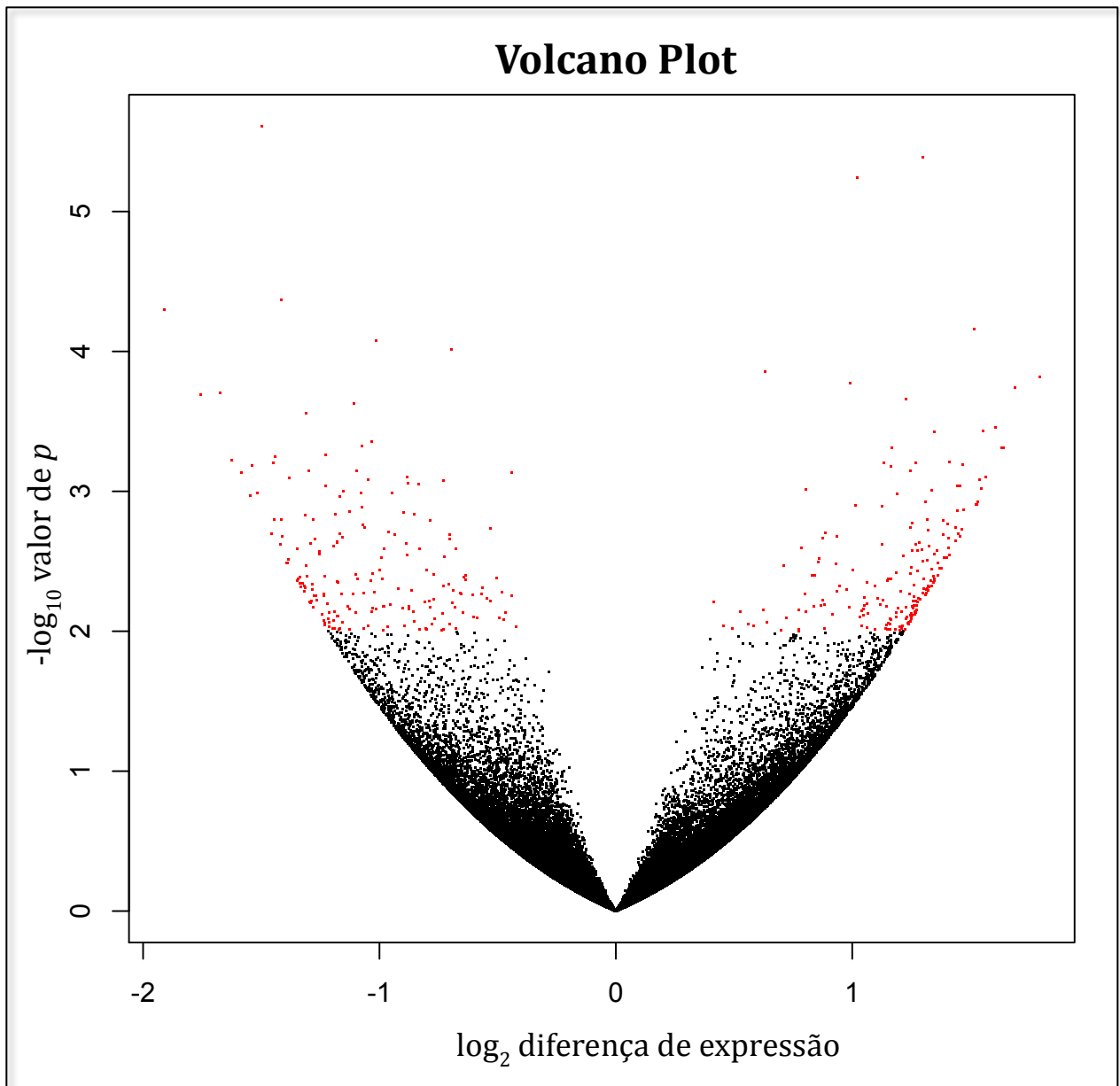
**Tabela 3. Genes anotados e diferencialmente expressos entre os dois grupos**

Numerode Acesso	Símbolo do gene	Nome do Gene	$\log_2$ Fold Change	Erro Padrão - Fold Change	Valor de p
NM_001046474	PPM1K	protein phosphatase, Mg <sup>2+</sup> /Mn <sup>2+</sup> dependent, 1K	-1,499	0,318	2,45E-06
NM_001080217	MFAP4	microfibrillar-associated protein 4	-1,414	0,346	4,30E-05
NM_001076883	GFPT2	glutamine-fructose-6-phosphate transaminase 2 hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit	-1,014	0,258	8,34E-05
NM_174336	HADHB	hydratase (trifunctional protein), beta subunit	0,632	0,166	1,40E-04
NM_001193080	SULT1C3	sulfotransferase family, cytosolic, 1C, member 3	1,793	0,473	1,52E-04
NM_001099706	ECM1	extracellular matrix protein 1	-1,309	0,360	2,75E-04
NM_001191148	OPRD1	opioid receptor, delta 1	1,347	0,379	3,76E-04
NM_001076164	PPP1R3C	protein phosphatase 1, regulatory subunit 3C	-1,031	0,293	4,40E-04
NM_001034765	THY1	Thy-1 cell surface antigen	-1,075	0,308	4,72E-04
NM_001191339	WDR86	WD repeat domain 86	1,631	0,468	4,85E-04
NM_001099091	TTC30B	tetratricopeptide repeat domain 30B	1,169	0,335	4,91E-04

NM_001080363	LOC785567	patatin-like phospholipase domain containing 4-like	-1,226	0,355	5,50E-04
NM_001098989	HIST1H1E	histone cluster 1, H1e	-1,450	0,424	6,22E-04
NM_001034671	CALML4	calmodulin-like 4	1,135	0,332	6,27E-04
NM_001034394	CD1E	CD1e molecule	-1,585	0,469	7,31E-04
NM_174766	TNS1	tensin 1	-0,440	0,130	7,34E-04
NM_001024545	LRRC51	leucine-rich repeat-containing protein 51	-0,729	0,218	8,35E-04
NM_001143743	BOLA	MHC class I antigen	-0,879	0,264	8,75E-04
NM_001045903	DPT	dermatopontin	-0,945	0,288	1,02E-03
NM_001075779	DLL4	delta-like 4 (Drosophila)	-1,076	0,328	1,02E-03
NM_001038212	PMF1	polyamine-modulated factor 1 solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	-1,168	0,358	1,09E-03
NM_001081577	SLC1A4	sideroflexin 3	-1,074	0,334	1,30E-03
NM_001101946	SFXN3	sideroflexin 3	-0,897	0,281	1,42E-03
NM_174241	ANXA8L1	annexin A8-like 1	-1,186	0,373	1,45E-03
NM_001205971	WNT5A	wingless-type MMTV integration site family, member 5A	-0,855	0,269	1,46E-03
NM_001192390	ADAMTS15	ADAM metalloproteinase with thrombospondin type 1 motif, 15	-0,787	0,250	1,63E-03
NM_001205675	SUSD2	sushi domain containing 2	-1,062	0,341	1,81E-03
NM_001046174	SLC35C2	solute carrier family 35, member C2	-0,531	0,171	1,83E-03
NM_001080729	ISLR	immunoglobulin superfamily containing leucine-rich repeat	-0,960	0,310	1,95E-03
NM_174604	SLC2A4	solute carrier family 2 (facilitated glucose transporter), member 4	-0,702	0,228	2,04E-03
NM_001206803	VASH1	vasohibin 1	-0,937	0,304	2,05E-03
NM_001206494	ZNF18	zinc finger protein 18	-0,704	0,230	2,19E-03
NM_001076493	ASF1A	ASF1 anti-silencing function 1 homolog A (S. cerevisiae)	-1,179	0,386	2,29E-03
NM_001075183	LAT2	linker for activation of T cells family, member 2	-1,179	0,387	2,34E-03
NM_001037816	TYMS	thymidylate synthetase	-0,887	0,292	2,37E-03
NM_001163936	APOBEC3A	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A solute carrier family 1 (glial high affinity glutamate transporter), member 3	1,127	0,371	2,39E-03
NM_174600	SLC1A3	neuroigin 2	0,784	0,260	2,53E-03
NM_001191242	NLGN2	neuroigin 2	-0,980	0,325	2,58E-03
NM_001035463	GPC3	glypican 3	-1,254	0,418	2,70E-03
NM_001035345	DNAAF1	dynein, axonemal, assembly factor 1	1,396	0,470	2,98E-03
NM_001192160	TUB	tubby homolog (mouse)	-0,992	0,334	2,99E-03
NM_001206178	SLC12A3	solute carrier family 12 (sodium/chloride transporters), member 3	1,298	0,440	3,16E-03
NM_001075586	LOC512899	ankyrin repeat domain 40-like	0,709	0,242	3,38E-03
NM_001206455	BRAT1	BRCA1-associated ATM activator 1	-0,801	0,276	3,67E-03
NM_174055	FGF1	fibroblast growth factor 1 (acidic)	-1,028	0,356	3,89E-03
NM_001102546	PDE4B	phosphodiesterase 4B, cAMP-specific	0,842	0,292	3,97E-03
NM_001045897	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	-0,637	0,221	4,00E-03
NM_001102002	DOK2	docking protein 2, 56kDa	-1,231	0,429	4,07E-03
NM_173918	NPC2	Niemann-Pick disease, type C2	-0,642	0,224	4,10E-03
NM_001075681	CCDC107	coiled-coil domain containing 107	-0,504	0,176	4,16E-03
NM_001191391	ALDH1L2	aldehyde dehydrogenase 1 family, member L2	-0,989	0,346	4,25E-03
NM_001035451	TMEM222	transmembrane protein 222	-0,636	0,223	4,30E-03
NM_001192792	ALOX5	arachidonate 5-lipoxygenase	-1,038	0,364	4,33E-03
NM_174371	KCNA4	potassium voltage-gated channel, shaker-related subfamily, member 4	-1,177	0,414	4,42E-03
NM_001035379	RASGRP4	RAS guanyl releasing protein 4	-1,176	0,414	4,48E-03
NM_001083449	CIDEA	cell death-inducing DFFA-like effector a	-0,706	0,250	4,64E-03
NM_001038672	UBXN8	UBX domain protein 8	-0,728	0,258	4,73E-03
NM_001206695	B4GALNT3	beta-1,4-N-acetyl-galactosaminyl transferase 3	1,323	0,469	4,75E-03
NM_001035330	GPA33	glycoprotein A33 (transmembrane)	1,097	0,388	4,75E-03
NM_001099720	RMI2	RMI2, RecQ mediated genome instability 2, homolog (S. cerevisiae)	-0,865	0,309	5,11E-03
NM_001103283	LRRTM3	leucine rich repeat transmembrane neuronal 3	1,204	0,430	5,16E-03
NM_001083446	ASAP3	ArfGAP with SH3 domain, ankyrin repeat and PH domain 3	-1,023	0,367	5,26E-03
NM_001014853	TOMM40L	translocase of outer mitochondrial membrane 40 homolog (yeast)-like	-0,482	0,173	5,27E-03
NM_001075830	MXRA8	matrix-remodelling associated 8	-0,659	0,237	5,34E-03
NM_174503	APLN	apelin	-1,161	0,417	5,38E-03
NM_001083763	NPAS2	neuronal PAS domain protein 2	-1,112	0,400	5,39E-03

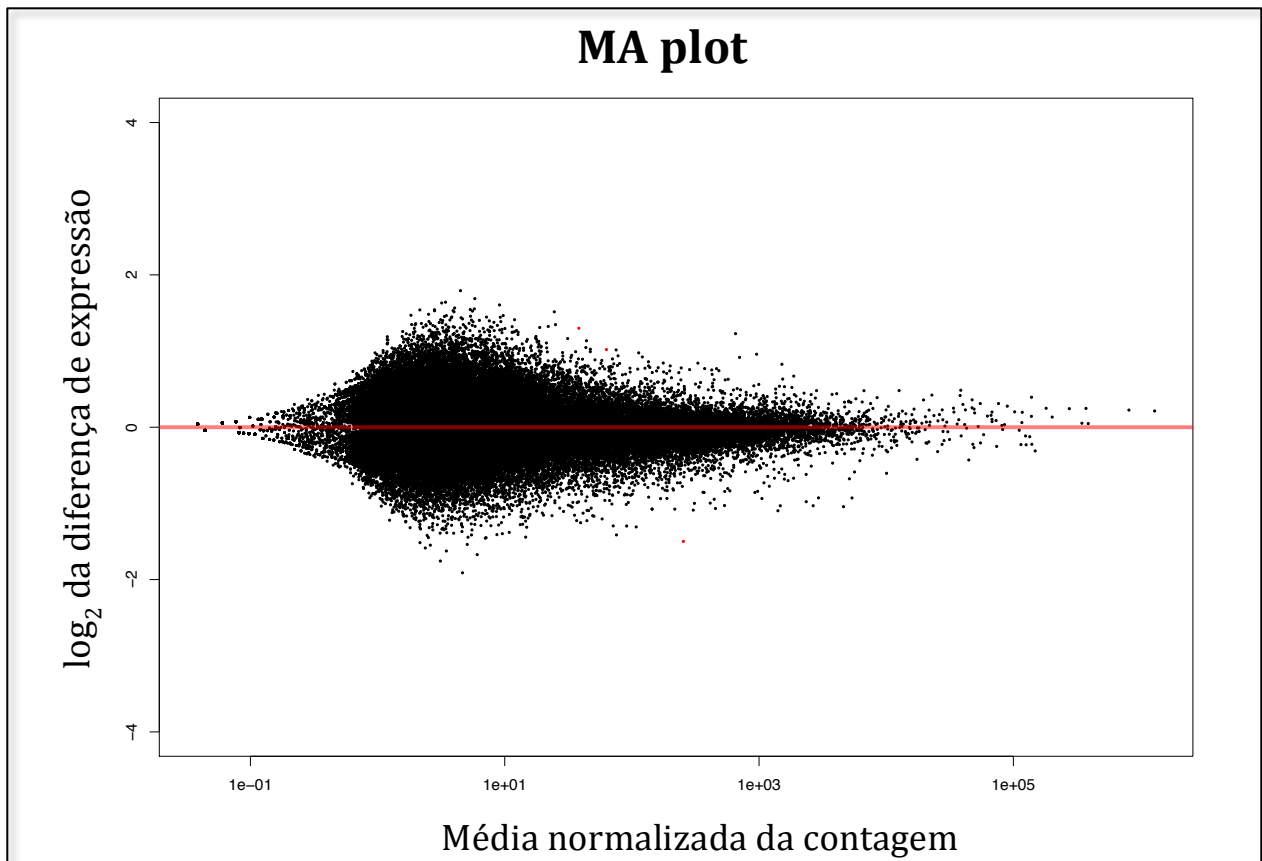
NM_001082473	EXTL3	exostoses (multiple)-like 3	-0,604	0,218	5,47E-03
NM_001083638	MEF2A	myocyte enhancer factor 2A	-0,440	0,159	5,58E-03
NM_001205724	DNASE1L3	deoxyribonuclease I-like 3	-1,276	0,460	5,58E-03
NM_001075636	RBM23	RNA binding motif protein 23	-0,547	0,198	5,78E-03
NM_001076407	RNFT1	ring finger protein, transmembrane 1	-0,532	0,194	6,06E-03
NM_174556	IGFBP3	insulin-like growth factor binding protein 3	-0,790	0,288	6,07E-03
NM_001206030	MIB1	mindbomb homolog 1 (Drosophila)	0,415	0,152	6,13E-03
NM_001122729	GBE1	glucan (1,4-alpha-), branching enzyme 1	-0,808	0,295	6,18E-03
NM_001046252	LITAF	lipopolysaccharide-induced TNF factor	-0,692	0,253	6,26E-03
NM_001145035	C15H11orf96	chromosome 15 open reading frame, human C11orf96	-0,877	0,323	6,60E-03
NM_001098989	HIST1H1E	histone cluster 1, H1e	-1,133	0,417	6,62E-03
NM_001034246	STAP2	signal transducing adaptor family member 2	-0,981	0,361	6,65E-03
NM_001034564	GPR133	G protein-coupled receptor 133	-0,874	0,323	6,89E-03
NM_001191206	CDH23	cadherin-related 23	-1,014	0,376	6,92E-03
NM_001145756	CD209	CD209 molecule	-1,128	0,418	6,97E-03
NM_001102216	ZNF565	zinc finger protein 565	0,623	0,231	7,00E-03
NM_174723	RHBG	Rh family, B glycoprotein	1,274	0,473	7,04E-03
NM_001076053	FAM73B	family with sequence similarity 73, member B	0,525	0,195	7,20E-03
NM_001034801	CEBPG	CCAAT/enhancer binding protein (C/EBP), gamma	-0,857	0,319	7,21E-03
NM_001083765	RMI1	RMI1, RecQ mediated genome instability 1, homolog (S. cerevisiae)	0,803	0,299	7,26E-03
NM_001078100	NR1D1	nuclear receptor subfamily 1, group D, member 1	-0,645	0,241	7,39E-03
NM_174477	TRPC2	transient receptor potential channel 2	1,046	0,392	7,64E-03
NM_001083655	PCDH1	protocadherin 1	-0,600	0,226	7,94E-03
NM_001077991	RECQL	RecQ protein-like (DNA helicase Q1-like)	0,722	0,272	7,97E-03
NM_001192754	LIMA1	LIM domain and actin binding 1	-0,586	0,221	8,02E-03
NM_001075484	MPP1	membrane protein, palmitoylated 1, 55kDa	-0,491	0,185	8,06E-03
NM_001206190	HAUS4	HAUS augmin-like complex, subunit 4	-0,767	0,290	8,10E-03
NM_001191209	VWCE	von Willebrand factor C and EGF domains	-1,215	0,459	8,18E-03
NM_174152	PPIB	peptidylprolyl isomerase B (cyclophilin B)	-0,473	0,179	8,31E-03
NR_031313	MIR365-2	microRNA mir-365-2	-0,852	0,324	8,47E-03
NM_001206624	CCDC69	coiled-coil domain containing 69	0,562	0,215	8,83E-03
NM_001033612	RAB34	RAB34, member RAS oncogene family	-0,703	0,269	8,84E-03
NM_001191366	BCL6B	B-cell CLL/lymphoma 6, member B	-0,777	0,297	8,86E-03
NM_001206781	VIPR2	vasoactive intestinal peptide receptor 2	-1,046	0,401	9,10E-03
NM_174134	OXTR	oxytocin receptor	-0,965	0,371	9,31E-03
NM_001205457	ZNF687	zinc finger protein 687	-0,423	0,163	9,34E-03
NM_001100359	SERPINE3	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 3	0,492	0,190	9,50E-03
NM_001192087	ARHGAP32	Rho GTPase activating protein 32	-0,675	0,261	9,57E-03
NM_001038521	PLAC9	placenta-specific 9	-0,805	0,311	9,62E-03
NM_001099077	PLXDC1	plexin domain containing 1	-0,801	0,310	9,71E-03
NM_001081589	NECAB3	N-terminal EF-hand calcium binding protein 3	-0,830	0,321	9,73E-03
NM_001024486	SLC37A2	solute carrier family 37 (glycerol-3-phosphate transporter), member 2	-0,984	0,381	9,77E-03
NM_001078163	DDIT3	DNA-damage-inducible transcript 3	-0,736	0,285	9,83E-03
NM_001077905	CPA5	carboxypeptidase A5	0,697	0,270	9,85E-03

Análise de expressão gênica diferencial entre os dois grupos de animais com diferentes perfis de crescimento muscular. Diferença de expressão estatisticamente significativa ( $p \leq 0,01$ ). Número de acesso: código de acesso do gene no banco de dados GenBank. Log<sub>2</sub> Fold Change: log<sub>2</sub> do valor da diferença de expressão gênica do Grupo 1 em relação ao Grupo 2. Erro Padrão - Fold Change: erro padrão do valores de diferença de expressão gênica.



**Figura 13.** *Volcano plot.* O eixo y é referente ao  $-\log_{10}$  dos valores de  $p$  não ajustados e o eixo x é referente ao  $\log_2$  da diferença de expressão para cada um dos genes. Os pontos vermelhos representam os genes que tiveram valores de  $p$  não ajustado menores que 0,01.

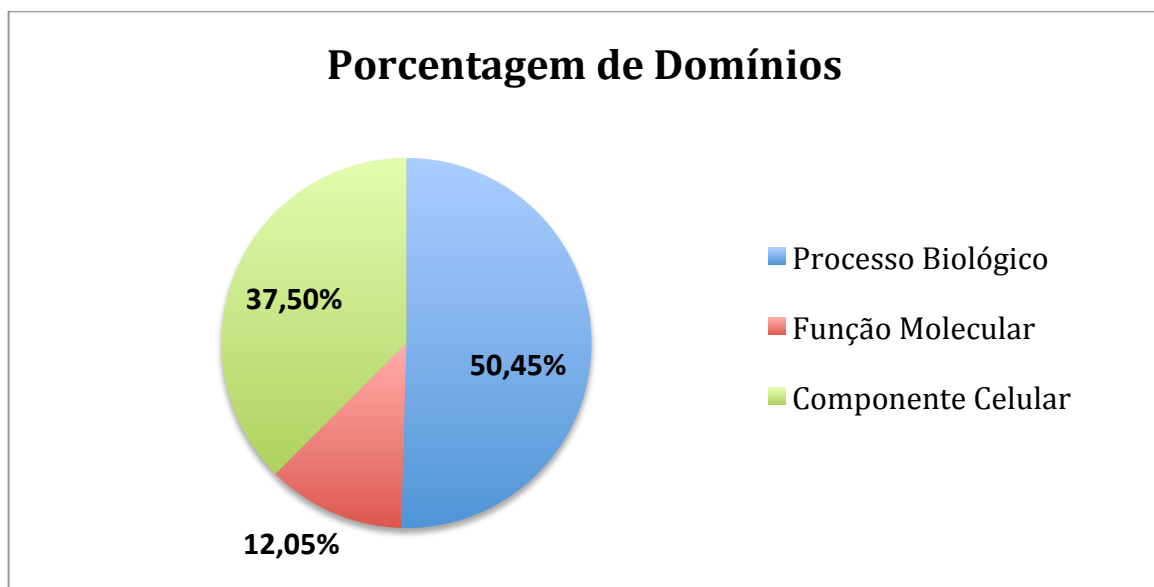
A Figura 14 mostra a variação da diferença de expressão em relação a média normalizada da contagem para cada um dos genes.



**Figura 14.** *MA plot*. O eixo y é referente ao  $\log_2$  da diferença de expressão e o eixo x é referente a média de contagem normalizada para cada um dos genes

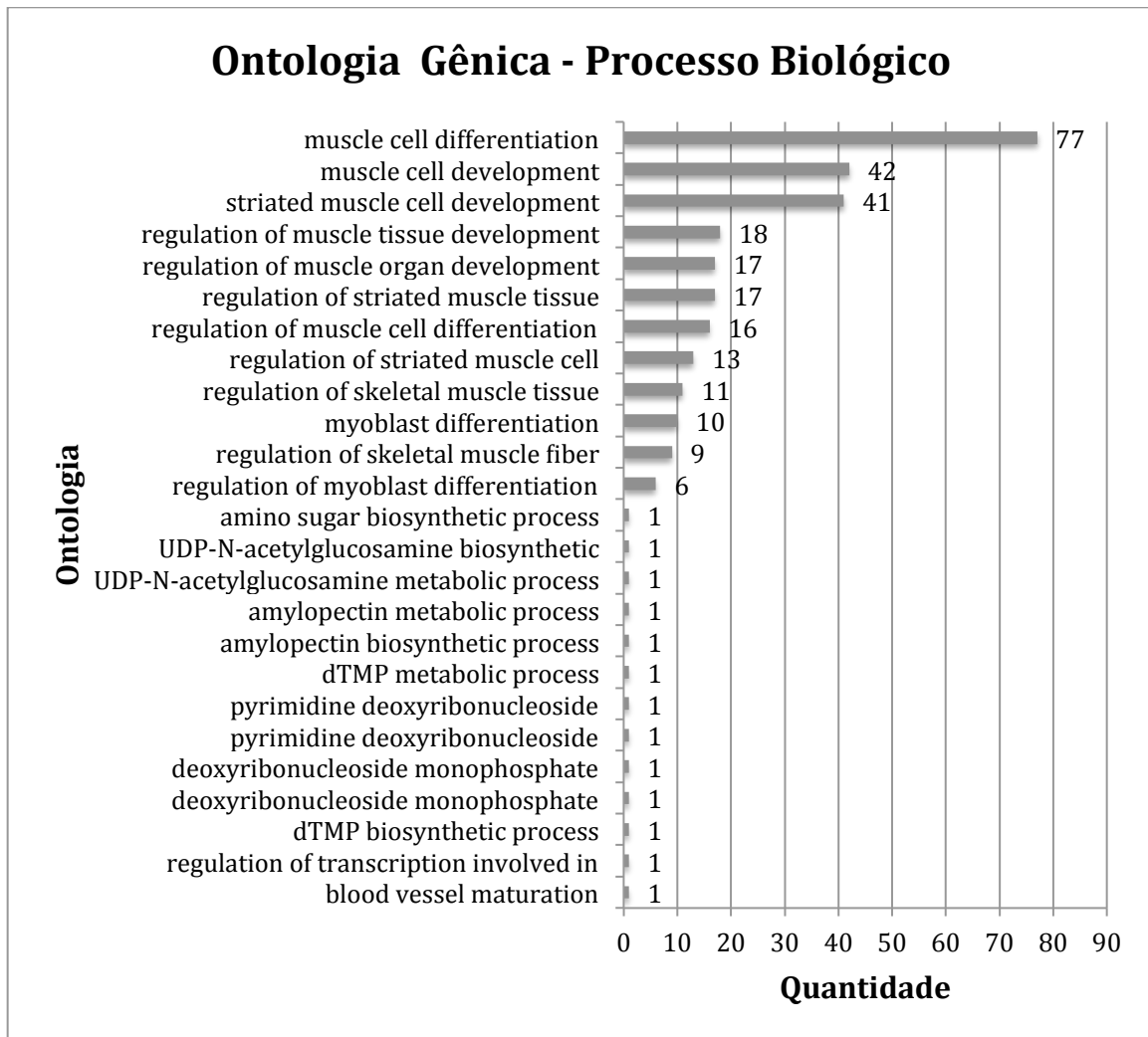
### *Ontologia gênica*

O projeto *Gene Ontology* fornece um vocabulário controlado (termo ontológico) que descreve o gene e os atributos dos produtos gênicos em qualquer organismo. A ontologia gênica cobre três domínios, cada um representando um conceito chave na biologia molecular: “Processos Biológicos”, “Função Molecular” e “Componente Celular”. Após a identificação dos 14.876 genes conhecidos por meio da análise de anotação, foi verificado quais eram os termos ontológicos atribuídos a cada um deles. No total 493.886 termos ontológicos foram atribuídos aos genes, dos quais 11.768 eram termos diferentes entre si. O domínio mais comum entre todos os termos ontológicos de todos os genes foi o “Processo Biológico” representando mais de 50% (Figura 15).

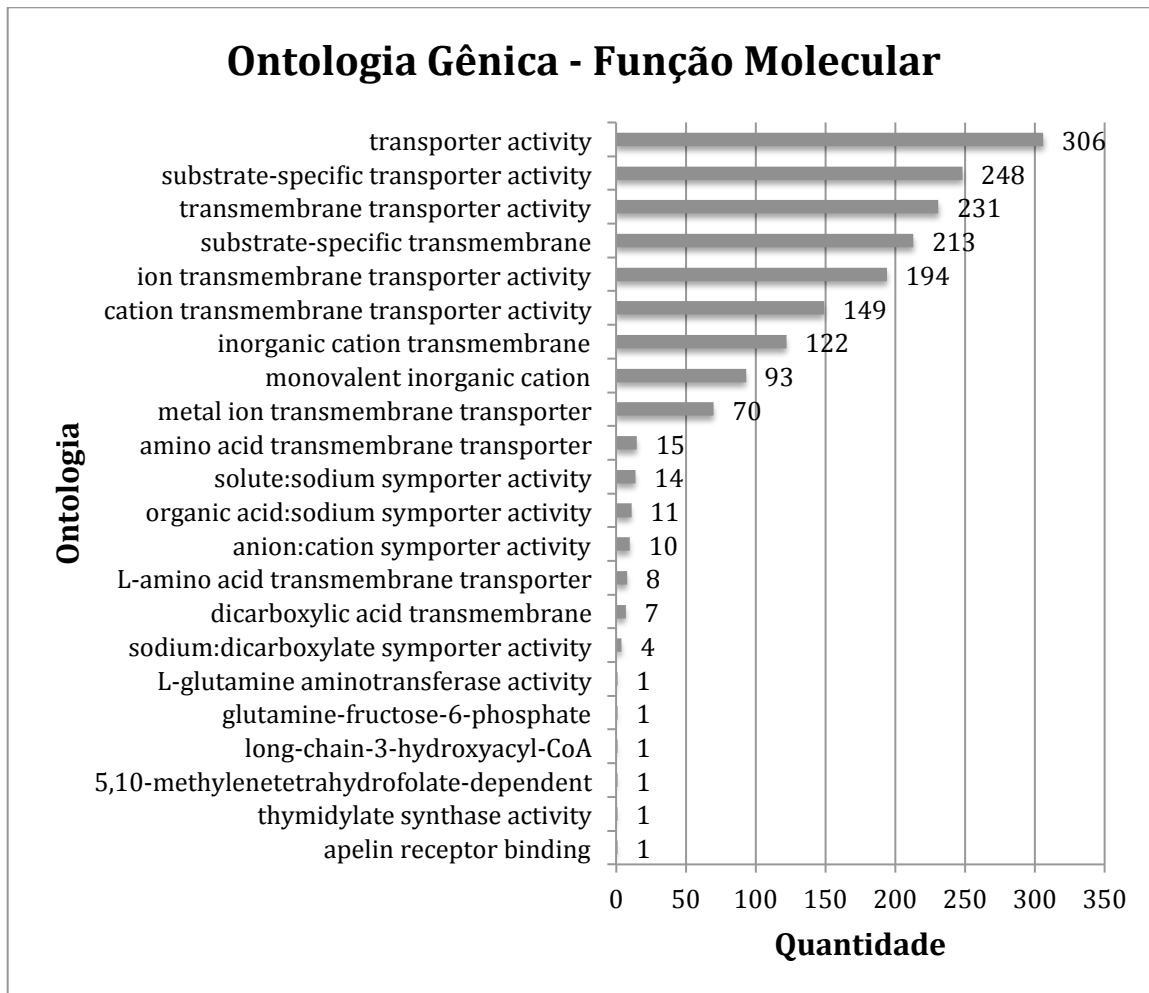


**Figura 15.** Porcentagem de domínios referentes a todos os termos ontológicos atribuídos a todos os genes anotados

A partir do resultado da análise de expressão gênica diferencial, foi calculado quais eram os termos ontológicos mais e menos relevantes entre os genes diferencialmente expressos. No total foram encontrados 48 termos ontológicos super representados e 4 termos ontológicos sub representados. Dos 48 termos super representados, 25 eram referentes ao domínio “Processo Biológico” (Figura 16), 22 eram referentes ao domínio “Função Molecular” (Figura 17) e 1 referente ao domínio “Componente Celular”. Dos 4 termos sub representados, 3 eram referentes a “Função Molecular” e 1 era referente ao “Componente Celular”. Os termos *insulin-responsive compartment* e *macromolecular complex* foram os únicos termos super e sub representados, respectivamente, do domínio “Componente Celular”. Os três termos sub representados do domínio “Função Molecular” foram: *catalytic activity*, *nucleotide binding* e *nucleoside phosphate binding*.



**Figura16.** Ontologia gênica – processo biológico. Termos ontológicos super representados entre os genes diferencialmente expressos. No eixo y estão os nomes dos termos ontológicos que tiveram valores de  $p \leq 0,01$  e no eixo x estão as quantidades de genes que tiveram um determinado termo atribuído a ele.



**Figura17.** Ontologia gênica – função molecular. Termos ontológicos super representados entre os genes diferencialmente expressos. No eixo y estão os nomes dos termos ontológicos que tiveram valores de  $p \leq 0,01$  e no eixo x estão as quantidades de genes que tiveram um determinado termo atribuído a ele.

## 5. Discussão

A análise global do padrão de expressão gênica pode fornecer maior compreensão das características do transcriptoma. Este trabalho contribui com uma visão mais abrangente da complexidade do transcriptoma do músculo de bovinos, mais especificamente do músculo LD de bovinos da raça Nelore. Devido ao alto rendimento, acurácia e repetibilidade, a técnica de RNA-seq, não somente permite a análise de características do transcriptoma, mas também contribui muito para a melhoria de

anotação gênica em uma resolução extremamente específica e assim, fornece dados complementares e de grande valor para outros estudos.

Para as dez bibliotecas geradas nesse estudo, mais de 40 bilhões de bases foram obtidas, o que equivale aproximadamente a treze genomas do boi. No sequenciamento das dez amostras do nosso estudo, foram obtidos mais de 450 milhões de fragmentos. Concomitante a isso, a porcentagem de alinhamento deste trabalho (aproximadamente 64% de alinhamento ao genoma de referência) foi semelhante aos dados de alinhamento de outros trabalhos com bovinos (WICKRAMASINGHE *et al.*, 2012; HUANG & KHATIB, 2010) e suínos (CHEN *et al.*, 2011), nos quais 61,4 a 65,6% de fragmentos foram alinhados ao genoma de referência. Além disso, a densidade gerada de fragmentos foi suficiente para uma análise qualitativa e acurada da expressão gênica. Outros fortes indícios de que a construção da biblioteca e o sequenciamento foram bem sucedidos, consistem no fato de que a quantidade de transcritos provenientes de ambas as fitas da molécula de DNA foi semelhante e o outro fator importante foi a distribuição uniforme da expressão dos genes ao longo dos 31 cromossomos.

Neste estudo foi utilizada a variação do valor de FPKM para a caracterização dos níveis de expressão gênica (XIONG *et al.*, 2012, WICKRAMASINGHE *et al.*, 2012). Poucos genes foram predominantes em todas as amostras, provavelmente compondo o cerne do transcriptoma do músculo LD. Os 20 e 50 genes mais abundantes representaram respectivamente 54,88% e 63,13% do total de genes anotados quantificados. O gene mais abundante foi o LBH (*limb bud and heart development homolog – mouse* - NM\_001099152) representando 13,66% do transcriptoma. LBH é um cofator transcricional tecido específico altamente conservado, sendo que seu gene é um alvo direto da via de sinalização WNT, que por sua vez é uma rede genética fundamental para o desenvolvimento embrionário e para a homeostase de tecidos adultos (RIEGER *et al.*, 2010). Entre os genes mais abundantes encontrados estavam a alfa actina 1 (NM\_174225), diferentes isoformas de cadeia leve de miosina, diferentes isoformas de cadeia pesada de miosina, diferentes isoformas de troponina e desmina (NM\_001081575).

Apesar do número de experimentos para análise de expressão gênica diferencial por meio de RNA-seq em bovinos ter aumentado nos últimos anos (XIONG *et al.*, 2012, WICKRAMASINGHE *et al.*, 2012; HE & LIU, 2013), poucos artigos avaliaram o transcriptoma de tecidos musculares de bovinos, principalmente bovinos de raças zebuínas. Neste trabalho foram desenvolvidas análises preliminares de genes diferentemente expressos em dois grupos de animais da raça Nelore com diferentes perfis de crescimento muscular. Apesar de nenhum valor de  $p$  ajustado para taxa de falsos positivos ter sido considerado estatisticamente significativo, 391 genes tiveram valor de  $p$  não ajustado abaixo de 0,01. Semelhante ao nosso estudo, BALDWIN e colaboradores (2012) analisaram o transcriptoma do músculo de bovinos com diferentes perfis de crescimento muscular e encontraram dois genes diferentemente expressos entre os grupos estudados. Eles verificaram menor valor de expressão do gene PPM1K (*protein phosphatase, Mg<sup>2+</sup>/Mn<sup>2+</sup> dependent, 1K*) e maior valor de expressão do gene CIDEA (*cell death-inducing DFFA-like effector A*) no músculo *longissimus lumborum* em novilhos da raça Angus com maior taxa de crescimento em relação a novilhos com menor taxa de crescimento. Em nosso trabalho, os genes PPM1K e CIDEA também tiveram diferença significativa de expressão no músculo LD, no entanto, ambos os genes tiveram maior valor de expressão no grupo com perfil para maior crescimento muscular do que no grupo com perfil para menor crescimento muscular.

Nossos resultados indicam que a anotação atual do genoma bovino está longe de ser completa. Assim como demonstrado em outros trabalhos (XIONG *et al.*, 2012, WICKRAMASINGHE *et al.*, 2012; HE & LIU, 2013; ZHANG *et al.*, 2013), nossos resultados mostraram diversas novas unidades transcricionais e eventos de *splice* desconhecidos. Além dos 10.933 potenciais novos eventos de *splice* (potenciais novas isoformas) para genes anotados, foram identificados 6.065 PNGs que tiveram alinhamento significativo com o blastn. Interessantemente, 2.547 dos PNGs foram significativamente alinhados a sequências de modelos preditos de RNA de *Bos taurus*. Estes dados podem contribuir significativamente com a descoberta de novos genes e no aperfeiçoamento da atual anotação gênica de bovinos. No entanto, embora o RNA-seq tenha a capacidade de detectar atividades transcricionais não anotadas, vários

filtros são necessários para evitar a detecção de “genes” espúrios. Diversos pacotes de softwares foram desenvolvidos para ajudar na identificação de junção de *splice* (TRAPNELL *et al.*, 2009) e na reconstrução de transcritos de dados de RNA-seq (TRAPNELL *et al.*, 2010; GUTTMAN *et al.*, 2010), mas ainda a profundidade do sequenciamento e tamanho dos fragmento são extremamente importantes. Como tecnologias de sequenciamentos continuam a evoluir para leituras mais longas e acuradas, e softwares mais poderosos continuam a ser desenvolvidos, a necessidade de uma anotação completa é um grande limitante.

Para maior compreensão dos processos biológicos que são regulados nos diferentes perfis de crescimento muscular, os genes diferentemente expressos foram testados para verificar quais os termos ontológicos mais relevantes entre os grupos. Para o domínio “Função Molecular” foram encontrados 22 termos ontológicos relevantes, no qual 18 deles eram relacionados com o transporte de moléculas nas células. Já para o domínio “Processo Biológico”, 12 dos 25 termos mais relevantes eram relacionados com a regulação e desenvolvimento muscular, o que fortalece a relevância da relação dos genes diferentemente expressos (associados a estes 12 termos) com o crescimento muscular. Estes termos estavam associados a quatro genes com diferença de expressão estatisticamente diferente. São eles: proteína de ligação do hormônio de crescimento tipo insulina 3 (IGFBP3), fator potenciador de miócitos 2A (MEF2A), dano induzível do DNA, transcrito 3 (DDIT3) e ASF1 função anti-silenciamento 1 homólogo A (*S. cerevisiae*) (ASF1A). No tópico “Transcriptoma e a qualidade da carne” da introdução foi realizada uma revisão bibliográfica dos genes MEF2A e IGFBP3. No entanto, não encontramos informações relacionadas com os genes DDIT3 e ASF1A e quaisquer características relacionadas com produção animal. ASF1 é uma chaperona específica para os heterodímeros de histonas H3 e H4. Esta proteína foi identificada por sua capacidade de induzir a expressão de genes teloméricos silenciados em levedura (LE *et al.*, 1997). Mais recentemente, foi verificado a participação da ASF1 na deposição e remoção de histonas H3 e H4 durante vários processos biológicos como a duplicação de DNA, transcrição e reparo de dano no DNA (MOUSSON *et al.*, 2007). Em nosso trabalho um termo ontológico (GO:0042083) relacionado com transporte de grupos metil também foi considerado super-

representado. DDIT3 é um membro de uma família de genes relacionados com a supressão de crescimento celular e dano induzido do DNA. A grande maioria dos trabalhos disponíveis relacionados a esta família de genes são referentes a tumores.

Alguns genes considerados potenciais marcadores para qualidade da carne foram destacados neste estudo com o intuito de melhor esclarecer suas características no transcriptoma do músculo LD (Tabela 4). Apesar de detectados, alguns destes genes não tiveram seus valores de expressão calculados pelo Cufflinks. No entanto, o nível de expressão destes genes foi calculado pelo HTseq e as suas diferenças relativas de expressão entre os grupos foi testada. Os genes DES, TNNC 1 e 2, FABP3 e TTN estavam entre os 100 genes mais abundantes do transcriptoma. Não foi encontrada anotação do gene TTN no banco de dados USCS, apesar dele estar anotado no banco de dados ENSEMBL e ter sequências preditas de mRNA no RefSeq. Foram identificados 17 eventos de *splice* no gene da calpastatina (CAST), 4 deles foram considerados conhecidos, referentes as 4 isoformas da calpastatina anotadas para bovinos até o momento. E interessante, 13 eventos foram considerados como potenciais novas isoformas. Os genes LEPR, MFE2A e IGFBP3 tiveram valores significativos de expressão gênica diferencial entre os grupos. Os três genes apresentaram menores valores de expressão no grupo com perfil para menor crescimento muscular em relação ao grupo com perfil para maior crescimento muscular. Estes resultados estão de acordo com a função das proteínas codificadas por estes genes que, nos três casos são proteínas relacionadas com o aumento do crescimento muscular.

**Tabela 4. Resumo dos genes relevantes para qualidade carne**

Símbolo	Número de acesso	FPKM	Posição	Eventos	Anotado	PNI	DE	Expressão
LEP	NM_173928	NA	NA	1	1	0	Não	-
LEPR	NM_001012285	NA	NA	1	1	0	Sim	↓
DGAT1	NM_174693	30,58	7252	5	2	3	Não	-
CAPN1	NM_174259	13,2	11321	5	1	4	Não	-
CAPN2	NM_001103086	144,14	2352	1	1	0	Não	-
CAST	NM_001030318	321,83	1199	17	4	13	Não	-
SDC	NM_173959	110,8	2894	1	1	0	Não	-

PGM	NM_001076903	2447,95	421	5	1	4	Não	-
TG	NM_173883	NA	NA	1	1	0	Não	-
PKM	NM_001205727	1312,33	396	6	1	5	Não	-
GHD	NM_001144852	NA	NA	1	1	0	Não	-
FABP3	NM_174313	7750,85	88	2	1	1	Não	-
FABP4	NM_174314	42,61	5884	1	1	0	Não	-
FABP5	NM_174315	94,55	3275	2	1	1	Não	-
MSTN	NM_001001525	27,27	7756	1	1	0	Não	-
PRKAG3	NM_001162421	109,58	2918	10	8	2	Não	-
PRKAB2	NM_001192328	41,06	6034	4	1	3	Não	-
DES	NM_001081575	15999,22	39	1	1	0	Não	-
TTN		813,94	86*	1	0	0	Não	-
CACNA2D1	NM_001205987	131,79	2525	6	1	5	Não	-
GH	NM_180996	NA	NA	1	1	0	Não	-
GHR	NM_176608	28,56	7548	3	2	1	Não	-
IGF1	NM_00107782	11,73	11866	2	2	0	Não	-
IGF2	NM_174087	118,95	2733	6	1	5	Não	-
MyF5	NM_174116	12,4	11545	1	1	0	Não	-
MyF6	NM_181811	232,8	1601	1	1	0	Não	-
MYOD1	NM_001040478	37,08	6475	2	2	0	Não	-
MYOG	NM_001111325	25,67	8046	1	1	0	Não	-
PPARG	NM_181024	27,54	7707	7	6	1	Não	-
PPARGC1A	NM_177945	20,22	9145	1	1	0	Não	-
TNNC1	NM_001034351	23382,3	25	1	1	0	Não	-
TNNC2	NM_001076373	161456,28	6	3	1	2	Não	-
MEF2A	NM_001083638	47,21	5484	5	1	4	Sim	↓
IGFBP3	NM_174556	54,69	4902	1	1	0	Sim	↓

FPKM: soma dos valores de FPKM para as 12 corridas das 10 amostras. NA: FPKM não calculado para esses genes. Posição: posição do nível de expressão entre os genes anotados. O gene TTN não tem anotação no banco de dados UCSC, portanto não estava na lista dos genes anotados. No entanto, considerando o seu valor de FPKM ele seria o octagésimo sexto gene mais expresso. Eventos: quantidade de eventos de *splice*. Anotado: quantidade transcritos anotados. PNI: quantidade de potenciais novas isoformas. Devido o gene TTN não estar anotado no banco de dados UCSC, o seu único transcrito foi considerado desconhecido pelo Cufflinks neste estudo. DE: gene com valor de expressão significativamente diferente entre os grupos. Expressão: diferença de expressão entre os grupos. ↓ genes com diferença de expressão negativa no Grupo 2 em relação ao Grupo 1.

## 6. Conclusão

Em conclusão, este primeiro trabalho de sequenciamento do transcriptoma do músculo *longissimus dorsi* de bovinos da raça Nelore, revelou um grande número de genes de funções conhecidas e desconhecidas, expandindo a quantidade de

informação genética disponível para esta raça e para a espécie, fornecendo maior esclarecimento sobre a complexidade de sua arquitetura gênica.

## **7. Considerações finais**

Desde o desenvolvimento da metodologia de Sanger até a finalização do projeto do genoma humano, a genética fez grandes avanços no entendimento gênico e funcional. E apesar dos grandes avanços alcançados pelos projetos Genoma Humano, HapMap e ENCODE (BIRNEY, et al., 2007; COLLINS, 2004 et al., 2004; INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2005), nós ainda estamos longe da compreensão completa da diversidade do transcriptoma humano e ainda mais longe da compreensão da diversidade do transcriptoma de animais de produção. Com o surgimento das plataformas de sequenciamento de nova geração em 2004 e a sua rápida evolução, pesquisadores passaram a ter a possibilidade de gerar dados sobre variação genética e padrões de expressão gênica em uma escala sem precedentes, o que vem se tornando uma ferramenta de rotina tanto em laboratórios de pesquisa como em laboratórios de diagnóstico clínico. A desvantagem desta rápida evolução é o aumento do desequilíbrio entre o ritmo das inovações tecnológicas introduzidas nas plataformas de sequenciamento e as descobertas biológicas provenientes delas. Isso porque existe a possibilidade do grande volume de dados gerado ser subutilizado, e em poucos meses deixar de ser útil devido a geração de novos dados. É necessário investir em desenvolvimento equivalente em novas estratégias computacionais para lidar com o volume de dados criados pela nova geração de instrumentos de sequenciamento e assim, maximizar seu potencial. Estas plataformas estão criando um novo mundo para ser explorado, não somente na definição de procedimentos técnicos e experimentais para análise em larga escala, mas também nas análises computacionais subsequentes necessárias para gerar dados de alta qualidade e com interpretação biológica correta. Na prática, o gargalo está na manipulação e análises estatísticas e computacionais dos dados e não mais na geração de dados experimentais.

Protocolos de RNAseq são extremamente sensíveis e é necessário controle de qualidade em cada passo no laboratório. Por exemplo, a contaminação dos reagentes com RNase e mesmo a degradação parcial do RNA tem que ser evitada em todos os procedimentos técnicos. A qualidade do RNA total isolado é o primeiro, e provavelmente o ponto crucial de um experimento com RNA-seq. Baixo rendimento da purificação do mRNA também é um passo crítico para uma biblioteca de qualidade. Além disso, com o objetivo de determinar corretamente a direcionalidade da transcrição dos genes e facilitar a detecção de transcritos opostos e/ou sobrepostos em regiões genômicas com alta densidade de genes, cuidados especiais devem ser tomados para preservação da informação da fita durante a construção da biblioteca. Para obter uma cobertura uniforme ao longo de todo o transcrito, deve-se utilizar primers randômicos na síntese do cDNA ao invés de primers oligo(dT) (que possui o viés da baixa cobertura da extremidade 5' do transcrito). Finalmente, deve-se levar em consideração quais são os tipos de vieses relacionados com o método do sequenciamento. No caso da tecnologia Ion Torrent, deve-se dar atenção especial a regiões homopoliméricas e erros relacionados a inserções e deleções.

Em relação a análise dos dados, sobre os tópicos mencionados acima, é importante considerar que a maioria dos softwares disponíveis para alinhamento de fragmentos foram desenhados inicialmente para fazer mapeamentos genômicos, e portanto, não são completamente capazes de descobrir todas as junções de éxons de uma amostra. No caso do TopHat por exemplo, a partir dos dados preliminares de alinhamento, é pré-construída uma biblioteca de junções que pode reduzir a possibilidade de descoberta de novas junções. Além disso, é necessário maior desenvolvimento para analisar “novas regiões de transcrição”, a “construção” de novos genes e a quantificação precisa de cada uma das novas isoformas, para o qual ainda faltam metodologias estatísticas. Para detecção de genes diferencialmente expressos, as metodologias atuais não foram ainda completamente validadas em dados biológicos e também não foram comparadas entre si em termos de especificidade e sensibilidade. Por último, e potencialmente de grande impacto, é a falta de replicatas biológicas que pode impedir a medição dos efeitos individuais e relação aos efeitos técnicos.

Replicatas biológicas são fundamentais em experimentos de RNA-seq para se calcular diferenças “reais” observadas entre grupos biológicos.

## 8. Referências

ALBERTS, B.; JOHNSON, A.; LEWIS J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Biologia Molecular da Célula*. **Nome capítulo**. 4. ed. Porto Alegre: Artmed, 2004. cap. X, p.

ANDERSSON, L. Genetic dissection of phenotypic diversity in farm animal. **Nature Reviews Genetics**, v. 2, p. 130-138, 2001.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome Biololgy**, v. 11, p. R106, 2010.

ANDERSON, P. T.; BERGEN, W. G., MERKEL, R. A.; ENRIGHT, W. J.; ZINN, S. A.; REFSAL, K. R.; *et al.* The relationship between composition of gain and circulating hormones in grow- ing beef bulls fed three dietary crude protein levels. **Journal of Animal Science**, v. 66, p. 3059-3067, 1988.

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

AKERS, R. M. Major advances associated with hormone and growth factor regulation of mammary growth and lactation in dairy cows. **Journal of Dairy Science**, v. 89, p. 1222-1234, 2006.

ASTIER, C.; LABBE, J. P.; ROUSTAN, C.; BENYAMIN, Y. Effects of different enzymatic treatments on the release of titin fragments from rabbit skeletal myofibrils - Purification of an 800 kDa titin polypeptide. **The Biochemical Journal**, v. 290, p. 731-734, 1993.

AVILÉS, C.; POLVILLO, O.; PEÑA, F.; JUÁREZ, M.; MARTÍNEZ, AL.; MOLINA, A. Associations between DGAT1, FABP4, LEP, RORC, and SCD1 gene polymorphisms and fat deposition in Spanish commercial beef. **Journal of Animal Science**, v. 91, p. 4571-4577, 2013.

BARB, C. R.; KRAELING, R. R.; RAMPACEK, G. B.; HAUSMAN, G. J. The role of neuropeptide Y and interaction with leptin in regulating feed intake and luteinizing hormone and growth hormone secretion in the pig. **Reproduction**, v. 131, p. 1127-1135, 2006.

BARENDSE, W. **Assessing lipid metabolismo**. Patent WO9923248, Patent US6383751. 1997. Disponível em: <<http://ep.espacenet.com>>. Acesso em 21 dez. 2013.

BALDWIN, R. L.; LI, R. W.; LI, C. J.; THOMSON, J. M.; BEQUETTE, B. J. Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. **Physiological Genomics**, v. 44, p. 543-550, 2012.

BEISSBARTH, T.; SPEED, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. **Bioinformatics**, v. 20, p. 1464-1465, 2004.

BELL, C. G.; WALLEY, A. J.; FROGUEL, P. The genetics of human obesity. **Nature Reviews Genetics**, v. 6, p. 221-234, 2005.

BERGVELD, P. Thirty years of isfetology: What happened in the past 30 years and what may happen in the next 30 years. **Sensors and Actuators B: Chemical**, v. 88, p. 1-20, 2003.

BIRNEY, E.; STAMATOYANNOPOULOS, J. A.; DUTTA, A.; GUIGÓ, R.; GINGERAS, T. R.; MARGULIES, E. H.; *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature**, v. 447, p. 799-816, 2007.

BOTTOMLY, D.; WALTER, N. A.; HUNTER, J. E.; DARAKJIAN, P.; KAWANE, S.; BUCK, K. J.; *et al.* Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. **PLoS One**, v. 6, p. e17820, 2011.

BURROWS, M.; WHEELER, D. J. A. Block Sorting Lossless Data Compression Algorithm. **Technical Report 124**. Palo Alto: Digital Equipment Corporation, 1994.

CAMERON, N. D. Genetic and phenotypic parameters for carcass traits, meat and eating quality traits in pigs. **Livestock Production Science**, v. 26, p. 119-135, 1990.

CABILI, M. N.; TRAPNELL, C.; GOFF, L.; KOZIOL, M.; TAZON-VEGA, B.; REGEV, A.; *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. **Genes & Development**, v. 25, p. 1915-1927, 2011.

CAPETANAKI, Y.; BLOCH, R. J.; KOULOUMENTA, A.; MAVROIDIS, M.; PSARRAS, S. Muscle intermediate filaments and their links to membranes and membranous organelles. **Experimental Cell Research**, v. 313, p. 2063-2076, 2007.

CARLING, D. The AMP-activated protein kinase cascade - a unifying system for energy control. **Trends in Biochemical Sciences**, v. 29, p. 18-24, 2004.

CARLIN, K. R. M.; HUFF-LONERGAN, E.; ROWE, L. J.; LONERGAN, S. M. Effect of oxidation, pH, and ionic strength on calpastatin inhibition of mu- and m-calpain. **Journal of Animal Science**, v. 84, p. 925-937, 2006.

CARNAC, G.; RICAUD, S.; VERNUS, B.; BONNIEU, A. Myostatin: biology and clinical relevance. **Mini Reviews in Medicinal Chemistry**, v. 6, p. 765-770, 2006.

CASAS, E.; STONE, R. T.; KEELE, J. W.; SHACKELFORD, S. D.; KAPPES, S. M.; KOOHMARAIE, M. A comprehensive search for quantitative trait loci affecting growth and carcass composition of cattle segregating alternative forms of the myostatin gene. **Journal of Animal Science**, v. 79, p. 854-860, 2001.

CAWLEY, S.; BEKIRANOV, S.; NQ, H. H.; KAPRANOV, P.; SEKINGER, E. A.; KAMPA, D.; *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. **Cell**, v. 116, p. 499-509, 2004.

CIOBANU, D. C.; BASTIAANSEN, J. W.; LONERGAN, S. M.; THOMSEN, H.; DEKKERS, J. C.; PLASTOW, G. S.; *et al.* New alleles in calpastatin gene are associated with meat quality traits in pigs. **Journal of Animal Science**, v. 82, p. 2829-2839, 2004.

CHEN, C.; AI, H.; REN, J.; LI, W.; LI, P.; QIAO, R.; *et al.* A global view of porcine transcriptome in three tissues from a full-sib pair with extreme phenotypes in growth and fat deposition by paired-end RNA sequencing. **BMC Genomics**, v. 12, p. 448-464, 2011.

CHEN, F.; CHEN, H.; WANG, J.; NIU, H.; LAN, X.; HUA, L.; *et al.* MEF2A gene polymorphisms are associated with growth traits in Chinese indigenous cattle breeds. **Journal of Animal and Veterinary Advances**, v. 9, p. 814-819, 2010.

CHEREL, P.; HERAULT, F.; VINCENT, A.; LE ROY, P.; DAMON, M. Genetic variability of transcript abundance in pig skeletal muscle at slaughter: Relationships with meat quality traits. **Journal of Animal Science**, v. 90, p. 699-708, 2012.

CLOONAN, N.; FORREST, A. R. R.; KOLLE, G.; GARDINER, B. B. A.; FAULKNER, G. J.; BROWN, M. K.; *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. **Nature Methods**, v. 5, p. 613-619, 2009.

COLLINS, F. S.; LANDER, E. S.; ROGERS, J.; WATERSON, R. H. Finishing the euchromatic sequence of the human genome. **Nature**, v. 431, p. 931-945, 2004.

CORIN, S. J.; JUHASZ, O.; ZHU, L.; CONLEY, P.; KEDES, L.; WADE, R. Structure and expression of the human slow twitch skeletal muscle troponin I gene. **Journal of Biological Chemistry**, v. 269, p. 10651-10659, 1994.

CREEMERS, E. E.; SUTHERLAND, L. B.; OH, J.; BARBOSA, A. C.; OLSON, E. N. Coactivation of MEF2 by the SAP domain proteins myocardin and MASTR. **Molecular Cell**, v. 23, p. 83-96, 2006.

D-B, L.; K-Y, Z.; D-W, C.; X-M, D.; YU, B. Effects of active immunization against myostatin on carcass quality and expression of the myostatin gene in pigs. **Journal of Animal Science**, v. 80, p. 585-590, 2009.

DENOYELLE, C. Evolution de la flaveur de la viande bovine en fonction de la teneur en lipides intra-musculaires. **Viande et Produits Carnés**, v. 16, p. 89-92, 1995.

DI STASIO, L.; DESTEFANIS, G.; BRUGIAPAGLIA, A.; ALBERA, A.; ROLANDO, A. Polymorphism of the GHR gene in cattle and relationships with meat production and quality. **Animal Genetics**, v. 36, p. 138-140, 2005.

DRANSFIELD, E. Consumer issues and acceptance of meat. In: PROCEEDINGS OF THE 47TH INTERNATIONAL CONGRESS OF MEAT SCIENCE AND TECHNOLOGY, 47., 2001. **Resumos**. Kraków. ICoMST, 2001. p. 72-79.

DUNNER, S.; SEVANE, N.; GARCIA, D.; LEVÉZIEL, H.; WILLIAMS, J. L.; MANGIN, B.; GeMQual Consortium. Genes involved in muscle lipid composition in 15 European *Bos taurus* breeds. **Animal Genetics**, v. 44, p. 493-501, 2013.

ENGSTRÖM, P. G.; SUZUKI, H.; NINOMIYA, N.; AKALIN, A.; SESSA, L.; LAVORGNA, G.; *et al.* Complex loci in human and mouse genomes. **PLoS Genetics**, v. 2, p. 564-577, 2006.

FARBER, C. R.; CORVA, P. M.; MEDRANO, J. F. Genome-wide isolation of growth and obesity QTL using mouse speed congenic strains. **BMC Genomics**, v. 7, p. 102, 2006.

FONTANESI, L.; DAVOLI, R.; NANNI COSTA, L.; SOTTI, E.; RUSSO, V. Study of candidate genes for glycolytic potential of porcine skeletal muscle: Identification and analysis of mutations, linkage and physical mapping and association with meat quality traits in pigs. **Cytogenetic and Genome Research**, v. 102, p. 145-151, 2003.

FRITZ, J. D.; GREASER, M. L. Changes in titin and nebulin in postmortem bovine muscle revealed by gel-electrophoresis, western blotting and immunofluorescence microscopy. **Journal of Food Science**, v. 56, p. 607, 1991.

FRITZ, J. D.; MITCHELL, M. C.; MARSH, B. B.; GREASER, M. L. Titin content of beef in relation to tenderness. **Meat Science**, v. 33, p. 41-50, 1993.

GEARY, T. W., MCFADIN, E. L., MACNEIL, M. D., GRINGS, E. E., SHORT, R. E., FUNSTON, R. N., *et al.* Leptin as a predictor of carcass composition in beef cattle. **Journal of Animal Science**, v. 81, p. 1-8, 2003.

GEAY, Y.; BAUCHART, D.; HOCQUETTE, J. F.; CULIOLI, J. Effect of nutritional factors on biochemical; structural and metabolic characteristics of muscles in ruminants; consequences on dietetic value and sensorial qualities of meat. **Reproduction Nutrition Development**, v. 41, p. 1-26. Erratum, v. 41, p. 377, 2001.

GOMEZ, L. C.; REAL, S. M.; OJEDA, M. S.; GIMENEZ, S.; MAYORGA, L. S.; ROQUÉ, M. Polymorphism of the FABP2 gene: a population frequency analysis and an association study with cardiovascular risk markers in Argentina. **BMC Medical Genetics**, v. 8, p. 39, 2007.

GRISART, B.; COPPIETERS, W.; FARNIR, F.; KARIM, L.; FORD, C.; BERZI, P.; *et al.* Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. **Genome Research**, v. 12, p. 222-231, 2002.

GUIMARAES, S. E. F.; STAHL, C. H.; LONERGAN, S. M.; GEIGER, B.; ROTHSCHILD, M. F. Myostatin promoter analysis and expression pattern in pigs. **Livestock Science**, v. 112, p. 143-150, 2007.

GUTIÉRREZ-GIL, B.; WILLIAMS, J. L.; HOMER, D.; BURTON, D.; HALEY, C. S.; WIENER, P. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. **Journal of Animal Science**, v. 87, p. 24-36, 2009.

GUTTMAN, M.; GARBER, M.; LEVIN, J. Z.; DONAGHEY, J.; ROBINSON, J.; ADICONIS, X.; *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. **Nature Biotechnology**, v. 28, p. 503-510, 2010.

HADJIPAVLOU, G.; MATIKA, O.; CLOP, A.; BISHOP, S. C. Two single nucleotide polymorphisms in the myostatin (GDF8) gene have significant association with muscle depth of commercial Charollais sheep. **Animal Genetic**, v. 39, p. 346-353, 2008.

HALE, C. S.; HERRING, W. O.; SHIBUYA, H.; LUCY, M. C.; LUBAHN, D. B.; KEISLER, D. H.; *et al.* Decreased growth in angus steers with a short TG-microsatellite allele in the P1 promoter of the growth hormone receptor gene. **Journal of Animal Science**, v. 78, p. 2099-2104, 2000.

HAN, X.; JIANG, T.; YANG, H.; ZHANG, Q.; WANG, W.; FAN, B. Investigation of four porcine candidate genes (H-FABP, MYOD1, UCP3 and MASTR) for meat quality traits in Large White pigs. **Molecular Biology Reports**, v. 39, p. 6599-6605, 2012.

HOCQUETTE, J. F.; BAS, P.; BAUCHART, D.; VERMOREL, M.; GEAY, Y. Fat partitioning and biochemical characteristics of fatty tissues in relation to plasma metabolites and hormones in normal and double-muscled young growing bulls. **Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology**, v. 123, p. 311-312, 1999.

HUANG, W.; KHATIB, H. Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. **BMC Genomics**, v. 11, p. 711-720, 2010.

HUFF-LONERGAN, E.; LONERGAN, S. M. Postmortem mechanisms of meat tenderization: The roles of the structural proteins and the calpain system. In: XIONG, Y. L.; HO, C. T.; SHAHIDI, F. (Eds.), *Quality attributes of muscle foods*, 1999. **Resumos**. New York: Kluwer Academic/Plenum Publishers, 1999. p. 229-251.

HUFF-LONERGAN, E.; MITSUHASHI, T.; BEEKMAN, D. D.; PARRISH, F. C.; OLSON, D. G.; ROBSON, R. M. Proteolysis of specific muscle structural proteins by u-calpain at low pH and temperature is similar to degradation in postmortem bovine muscle. **Journal of Animal Science**, v. 74, p. 993-1008, 1996.

HUFF-LONERGAN, E.; PARRISH, F. C.; ROBSON, R. M. Effects of postmortem aging time, animal age, and sex on degradation of titin and nebulin in bovine longissimus muscle. **Journal of Animal Science**, v. 73, p. 1064-1073, 1995.

HAMILL, R. M.; MCBRYAN, J.; MCGEE, C.; MULLEN, A. M.; SWEENEY, T.; TALBOT, A.; *et al.* Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. **Meat Science**, v. 92, p. 440-450, 2012.

HARDCASTLE, T. J.; KELLY, K. A. BaySeq: empirical Bayesian methods for identifying differential expression in sequence count data. **BMC Bioinformatics**, v. 11, p. 422, 2010.

HE, H.; LIU, X. Characterization of transcriptional complexity during longissimus muscle development in bovines using high-throughput sequencing. **PLoS One**, v. 8, p. e64356, 2013.

HINITS, Y.; HUGHES, S. M. Mef2s are required for thick filament formation in nascent muscle fibres. **Development**, v. 134, p. 2511-2519, 2007.

HOUSEKNECHT, K. L.; BAILE, C. A.; MATTERI, R. L.; SPURLOCK, M. E. The biology of leptin: A review. **Journal of Animal Science**, v. 76, p. 1405-1420, 1998.

HUANG, W.; SHERMAN, B. T.; LEMPICKI, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocol**, v. 4, p. 44-57, 2009.

International Human Genome Sequencing Consortium, A haplotype map of the human genome. **Nature**, v. 437, p. 1299-1320, 2005.

IOANNIDIS, J. P. Why most published research findings are false. **PLoS Medicine**, v. 2, p. e124, 2005.

JAKOBSEN, K. S. Direct mRNA isolation using magnetic Oligo(dT) beads: A protocol for all types of cell cultures, animal and plant tissues. **Advances in Biomagnetism**, p. 61-71, 1994.

JIANG, H.; WONG, W. H. Statistical inferences for isoform expression in RNA- Seq. **Bioinformatics**, v. 25, p. 1026-1032, 2009.

JURIE, C.; CASSAR-MALEK, I.; BONNET, M.; LEROUX, C.; BAUCHART, D.; BOULESTEIX, P.; *et al.* Adipocyte fatty acid-binding protein and mitochondrial enzyme activities in muscles as relevant indicators of marbling in cattle. **Journal of Animal Science**, v. 85, p. 2660-2669, 2007.

KAMBADUR, R.; SHARMA, M.; SMITH, T. P.; BASS, J. J. Mutations in myostatin (GDF8) in double-musled Belgian Blue and Piedmontese cattle. **Genome Research**, v. 7, p. 910-916, 1997.

KATO, K. Impact of the next generation DNA sequencers. **International Journal of Clinical and Experimental Medicine**, v. 2, p. 193-202, 2009.

KERSTEN, S. Mechanisms of nutritional and hormonal regulation of lipogenesis. **EMBO Reports**, v. 2, p. 282-286, 2001.

KIM, D.; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, v. 14, p. R36, 2003.

KIM, J. B.; PORRECA, G. J.; SONG, L.; GREENWAY, S. C.; GORHAM, J. M.; CHURCH, G. M.; *et al.* Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic cardiomyopathy. **Science**, v. 316, p. 1481-1484, 2007.

KLONT, R. E.; BROCKS, L.; EIKELNBOOM, G. Muscle fibre type and meat quality. **Meat Science**, v. 49, p. S219-S229, 1998.

KOOHMARAIE, M.; GEESINK, G. H. Contribution of postmortem muscle biochemistry to the delivery of consistent meat quality with particular focus on the calpain system. **Meat Science**, v. 74, p. 34-43, 2006.

KOOHMARAIE, M.; KILLEFER, J.; BISHOP, M. D.; SHACKELFORD, S. D.; WHEELER, T.L.; ARBONA, S. R. Calpastatin-based methods for predicting meat tenderness. In: A. OUALI, A.; DEMEYER, D.; SMULDERS, F. (EDS.), Expression of tissue proteinases and regulation of protein degradation as related to meat quality, 1995. **Resumos**. Utrecht: EECEAMST, 1995. p. 395-412.

KOOHMARAIE, M.; KENT, M. P.; SHACKELFORD, S. D.; VEISETH, E.; WHEELER, T. L. Meat tenderness and muscle growth: is there any relationship. **Meat Science**, v. 62, p. 345-352, 2002.

LAGONIGRO, R.; WIENER, P.; PILLA, F.; WOOLLIAMS, J. A.; WILLIAMS, J. L. A new mutation in the coding region of the bovine leptin gene associated with feed intake. **Animal Genetics**, v. 34, p. 371-374, 2003.

LANGMEAD, B.; TRAPNELL, C.; POP, M.; SALZBERG, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v. 10, p. R25, 2009.

LE, S.; DAVIS, C.; KONOPKA, J. B.; STERNGLANZ, R. Two new S-phase-specific genes from *Saccharomyces cerevisiae*. **Yeast**, v. 13, p. 1029-1042, 1997.

LEE, C. Y.; LEE, H. P.; JEONG, J. H.; BAIK, K. H.; JIN, S. K.; LEE, J. H.; *et al.* Effects of restricted feeding, low- energy diet, and implantation of trenbolone acetate plus estradiol on growth, carcass traits, and circulating concentrations of insulin-like growth factor (IGF)-1 and IGF-binding protein-3 in finishing barrows. **Journal of Animal Science**, v. 80, p. 84-93, 2002.

L'HONORE, A.; RANA, V.; ARSIC, N.; FRANCKHAUSER, C.; LAMB, N. J.; FERNANDEZ, A. Identification of a new hybrid serum response factor and myocyte enhancer factor 2-binding element in MyoD enhancer required for MyoD expression during myogenesis. **Molecular Biology Cell**, v.18, p. 1992-2001, 2007.

LI, C.; BASARAB, J.; SNELLING, W. M.; BENKEL, B. Identical by Descent Haplotype Sharing analysis: Application in Fine Mapping of QTLs for Birth Weight in Commercial Lines of *Bos taurus*. **Proceedings of 7th World Congress of Genetics Applied Livestock Production**, Montpellier, p. 481-484, 2002a.

LI, C.; BASARAB, J.; SNELLING, W. M.; BENKEL, B. The identification of common haplotypes on bovine chromosome 5 within commercial lines of *Bos taurus* and their associations with growth traits. **Journal of Animal Science**, v. 80: p. 1187-1194, 2002b.

LI, J.; TIBSHIRANI, R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. **Statistical Methods in Medical Research**, v. 22, p. 519-536, 2013.

LI, J. H.; YANG, M. S.; XU, N. Y. Single nucleotide polymorphism in 30 UTR of porcine CACNA2D1 gene and its genetic effect on meat quality trait. **Acta Agriculturae Boreali**, Sin v. 22, p. 76-79, 2007.

LI, J. H.; YANG, M. S.; XU, N. Y. Mapping of porcine CACNA2D1 gene and the integrative analysis of correlative genetic maps. **Journal of Northwest A&F University, Natural Sciences Education**, v. 36, p. 49-53, 2008.

LISOWSKI, P.; KOŚCIUCZUK, E. M.; GOŚCIK, J.; PIERZCHAŁA, M.; ROWIŃSKA, B.; ZWIERZCHOWSKI, L. Hepatic transcriptome profiling identifies differences in

expression of genes associated with changes in metabolism and postnatal growth between Hereford and Holstein-Friesian bulls. **Animal Genetics**, 2013. [versão impressa ainda não disponível]

LIU, R.; MA, X.; XU, L.; WANG, D.; JIANG, X.; ZHU, W.; *et al.* Differential microRNA expression in peripheral blood mononuclear cells from Graves' disease patients. **Journal of Clinical Endocrinology & Metabolism**, v. 97, p. E968-E972, 2012.

MAAK, S.; NEUMANN, K.; SALVE, H. H. Identification and analysis of putative regulatory sequences for the MYF5/ MYF6 locus in different vertebrate species. **Gene**, v. 379, p. 141-147, 2006.

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends in Genetics**, v. 24, p. 133-141, 2008.

MARIONI, J. C.; MASON, C. E.; MANE, S. M.; STEPHENS, M.; GILAD, Y. RNA- seq: An assessment of technical reproducibility and comparison with gene expression arrays. **Genome Research**, v. 18, p. 1509-1517, 2008.

MCFADIN, E. L., KEISLER, D. H., SCHMIDT, T. B., LORENZEN, C. L., BERG, E. P. Correlations between serum concentrations of leptin and beef carcass composition and quality. **Journal of Muscle Foods**, v. 14, p. 81-87, 2003.

MCPHERRON, A. C.; LEE, S. J. Suppression of body fat accumulation in myostatin-deficient mice. **Journal of Clinical Investigation**, v. 109, p. 595-601, 2002.

MEMISOGLU, A.; HU, F. B.; HANKINSON, S. E.; MANSON, J. E.; DE VIVO, I.; WILLETT, W. C.; *et al.* Interaction between a peroxisome proliferator-activated receptor  $\gamma$  gene polymorphism and dietary fat intake in relation to body mass. **Human Molecular Genetics**, v. 12, p. 2923-2929, 2003.

MELTON, L.; AMIRI, M.; DAVIS, W.; BACKUS, R. Flavor and chemical characteristics of ground beef from grass-, forage-grain- and grain-finished steers. **Journal of Animal Science**, v. 55, p. 77-87, 1982.

MORTAZAVI, A.; WILLIAMS, B. A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, v. 5, p. 621-628, 2008.

MOUSSON, F.; OCHSENBEIN, F.; MANN, C. The histone chaperone Asf1 at the crossroads of chromatin and DNA checkpoint pathways. **Chromosoma**, v. 116, p. 79-93, 2007.

MULLEN, A. J.; BARTON, P. J. Structural characterization of the human fast skeletal muscle troponin I gene (TNNI2). **Gene**, v. 242, p. 313-320, 2000.

NAGALAKSHMI, U.; WANG, Z.; WAERN, K.; SHOU, C.; RAHA, D.; GERSTEIN, M.; *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. **Science**, v. 320, p. 1344-1349, 2008.

OLSON, E. N.; PERRY, M.; SCHULZ, R. A. Regulation of muscle differentiation by the MEF2 family of MADS box transcription factors. **Developmental Biology**, v. 172, p. 2-14, 1995.

OUALI, A. Meat tenderization: possible causes and mechanisms. A review. **Journal of Muscle Foods**, v. 50, p. 129-165, 1990.

OUALI, A.; GAGAOUA, M.; BOUDIDA, Y.; BECILA, S.; BOUDJELLAL, A.; HERRERA-MENDEZ, C. H.; *et al.* Biomarkers of meat tenderness: Present knowledge and perspectives in regards to our current understanding of the mechanisms involved. **Meat Science**, v. 95, p. 854-870, 2013.

OWENS, P. C.; GATFORD, K. L.; WALTON, P. E.; MORLEY, W.; CAMPELL, R. G. The relationship between endogenous insulin-like growth factors and growth in pigs. **Journal of Animal Science**, v. 77, p. 2098-2103, 1999.

PAGE, B. T., CASAS, E., HEATON, M. P., CULLEN, N. G., HYNDMAN, D. L., MORRIS, C. A., *et al.* Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. **Journal of Animal Science**, v. 80, p. 3077-3085, 2002.

PANNIER, L.; SWEENEY, T.; HAMILL, R. M.; IPEK, F.; STAPLETON, P. C.; MULLEN, A. M. Lack of an association between single nucleotide polymorphisms in the bovine lep- tin gene and intramuscular fat in *Bos taurus* cattle. **Meat Science**, v. 81, p. 731-737, 2009.

PICARD, B.; BARBOIRON, C.; DURIS, M. P.; GAGNIÉRE, H.; JURIE, C.; GEAY, Y. Electrophoretic separation of bovine muscle myosin heavy chain isoforms. **Meat Science**, v. 53, p. 1-7, 1999.

PICARD, B.; JURIE, C.; DURIS, M. P.; RENAND, G. Consequences of selection for higher growth rate on muscle fibre development in cattle. **Livestock Science**, v. 102, p. 107-120, 2006.

PÖSÖ, A. R.; PUOLANNE, E. Carbohydrate metabolism in meat animals. **Meat Science**, v. 70, p. 423-434, 2005.

POTTHOFF, M. J.; ARNOLD, M. A.; MCANALLY, J.; RICHARDSON, J. A.; BASSELDUBY, R.; OLSON, E. N. Regulation of skeletal muscle sarcomere integrity and postnatal muscle function by Mef2c. **Molecular Cell Biology**, v. 27, p. 8143-8151, 2007.

PUIGSERVER, P.; SPIEGELMAN, B. M. Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): Transcriptional coactivator and metabolic regulator. **Endocrine Reviews**, v. 24, p. 78-90, 2003.

RAJARAM, S.; BAYLINK, D. J.; MOHAN, S. Insulin-like growth factor-binding protein in serum and other biological fluids: regulation and functions. **Endocrine Reviews**, v. 18, p. 801-831, 1997.

REARDON, W.; MULLEN, A. M.; SWEENEY, T.; HAMILL, R. M. Association of polymorphisms in candidate genes with colour, water-holding capacity, and composition traits in bovine *M. longissimus* and *M. semimembranosus*. **Meat Science**, v. 86, p. 270-275, 2010.

REECY, J. M.; SPURLOCK, D. M.; STAHL, C. H. Gene expression profiling: Insights into skeletal muscle growth and development. **Journal of Animal Science**, v. 84, p. E150, 2006.

RENAND, G.; BERGE, P.; PICARD, B.; ROBELIN, J.; GEAY, Y.; KRAUSS, D.; *et al.* Genetic parameters of young Charolais bulls progeny of divergently selected sires. In: PROCEEDINGS OF THE 5TH WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 5., 1994. **Resumos**. Guelph: WCGALP, 1994. p. 446-449.

RENAND, G.; PICARD, B.; TOURAILLE, C.; BERGE, P.; LEPETIT, J. Relationship between muscle characteristics and meat quality traits of young Charolais bulls. **Meat Science**, v. 59, p. 49-60, 2001.

RIEGER, M.E.; SIMS, A. H.; COATS, E. R.; CLARKE, R. B.; BRIEGEL, K. J. The embryonic transcription cofactor LBH is a direct target of the Wnt signaling pathway in epithelial development and in aggressive basal subtype breast cancers. **Molecular and Cellular Biology**, v. 30, p. 4267-4279, 2010.

REINARTZ, J.; BRUYNS, E.; LIN, J. Z.; BURCHAM, T.; BRENNER, S.; BOWEN, B.; *et al.* Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. **Briefings in Functional Genomics Proteomics**, v. 1, p. 95-104, 2002.

ROBINSON, J. T.; THORVALDSDOTTIR, H.; WINCKLER, W.; GUTTMAN, M.; LANDER, E. S.; GETZ, G. MESIROV, J. P. Integrative Genomics Viewer. **Nature Biotechnology**, v. 29, p. 24-26, 2011.

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, p. 139-140, 2010.

- ROBINSON, M. D.; SMYTH, G. K. Moderated statistical tests for assessing differences in tag abundance. **Bioinformatics**, v. 23, p. 2881-2887, 2007.
- ROBINSON, M. D.; SMYTH, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. **Biostatistics**, v. 9, p. 321-332, 2008.
- ROBLES, J. A.; QURESHI, S. E.; STEPHEN, S. J.; WILSON, S. R.; BURDEN, C. J.; TAYLOR, J. M. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. **BMC Genomics**, v. 13, p. 484, 2012.
- ROSEN, E. D.; MACDOUGALD, O. A. Adipocyte differentiation from the inside out. **Nature Reviews Molecular Cell Biology**, v. 7, p. 885-896, 2006.
- ROTHSCHILD, M. F.; HU, Z. L.; JIANG, Z. Advances in QTL mapping in pigs. **International Journal of Biological Sciences**, v. 10, p. 192-197, 2007.
- RUDEL, L.; PARK, S.; SAWYER, K. Compared with dietary monounsaturated and saturated fat, polyunsaturated fat protects African green monkeys from coronary artery athero-sclerosis. **Arteriosclerosis, Thrombosis, and Vascular Biology**, v. 15, p. 2101-2110, 1995.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of National Academy Science**, v. 74, p. 5463-5467, 1977.
- SCHENKEL, F. S.; MILLER, S. P.; YE, X.; MOORE, S. S.; NKRUMAH, J. D.; LI, C.; *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, p. 2009-2049, 2005.
- SCHUSTER, S. C. Next-generation sequencing transforms today's biology. **Nature Methods**, v. 5, p. 16-18, 2008.
- SELLIER, P. Genetics of meat and carcass traits. In: Rotschild MF, Ruvinsky A (eds). THE GENETICS OF THE PIG, 1998. **Resumos**. Wallingford, UK: CAB International, 1998, p. 463-510.
- SENTANDREU, M. A.; COULIS, G.; OUALI, A. Role of muscle endopeptidases and their inhibitors in meat tenderness. **Trends in Food Science and Technology**, v. 13, p. 400-421, 2002.
- SEVANE, N.; ARMSTRONG, E.; CORTÉS, O.; WIENER, P.; WONG, R. P.; DUNNER, S.; GemQual Consortium. Association of bovine meat quality traits with genes included in the PPARG and PPARGC1A networks. **Meat Science**, v. 94, p. 328-335, 2013.
- SHACKELFORD, S. D.; KOOHMARAIE, M.; MILLER, M. F.; CROUSE, J. D.; REAGAN, J. O. An evaluation of tenderness of the longissimus muscle of Angus by Hereford

versus Brahman crossbred heifers. **Journal of Animal Science**, v. 69, p. 171-177, 1991.

SHIRAKI, T.; KONDO, S.; KATAYAMA, S.; WAKI, K.; KASUKAWA, T.; KAWAJI, H.; *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. **Proceedings of National Academy Science**, v. 100, p. 15776-15781, 2003.

SIECZKOWSKA, H.; ZYBERT, A.; KRZECIO, E.; ANTOSIK, K.; KOĆWIN-PODSIADŁA, M.; PIERZCHAŁA, M. The expression of genes PKM2 and CAST in the muscle tissue of pigs differentiated by glycolytic potential and drip loss, with reference to the genetic group. **Meat Science**, v. 84, p. 137-142, 2010.

SMITH S. The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. **The FASEB Journal**, v. 8, p. 1248-1259, 1994.

STONE, R. T.; KAPPES, S. M.; BEATTIE, C. W. The bovine homolog of the obese gene maps to chromosome 4. **Mammalian Genome**, v. 7, p. 399-400, 1996.

SULTAN, M.; SCHULZ, M. H.; RICHARD, H.; MAGEN, A.; KLINGENHOFF, A.; SCHERF, M.; *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. **Science**, v. 321, p. 956-960, 2008.

SUN, W. B.; CHEN, H.; LEI, X. Q.; LEI, C. Z.; ZHANG, Y. H.; LI, R. B.; *et al.* Polymorphisms of insulin-like growth factor binding protein 3 gene and its associations with several carcass traits in Qinchuan cattle. **Yi Chuan**, v. 25, p. 511-516, 2003.

TANG, F.; CATALIN, B.; WANG, Y.; NORDMAN, E.; LEE, C.; XU, N.; *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. **Nature Methods**, v. 6, p. 377-382, 2009.

TARRANT, P. V. Some recent advances and future priorities in research for the meat industry. **Meat Science**, v. 49, p. S1-S16, 1998.

TARAZONA, S.; GARCIA-ALCALDE, F.; DOPAZO, J.; FERRER, A.; CONESA, A. Differential expression in RNA-seq: a matter of depth. **Genome Research**, v. 21, p. 2213-2223, 2011.

THALLER, G.; KÜHN, C.; WINTER, A.; EWALD, G.; BELLMANN, O.; WEGNER, J.; *et al.* DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. **Animal Genetics**, v. 34, p. 354-357, 2003.

THOMAS, M. G.; ENNS, R. M.; HALLFORD, D. M.; KEISLER, D. H.; OBEIDAT, B. S.; MORRISON, C. D.; *et al.* Relationships of metabolic hormones and serum glucose to growth and reproductive development in performance-tested Angus, Brangus, and Brahman bulls. **Journal of Animal Science**, v. 80, p. 757-767, 2002.

TIZIOTO, P. C.; DECKER, J. E.; TAYLOR, J. F.; SCHNABEL, R. D.; MUDADU, M. A.; SILVA, F. L.; *et al.* Genome scan for meat quality traits in Nelore beef cattle. **Physiology Genomics**, v. 45, p. 1012-1020, 2013.

TOUNG, J. M.; MORLEY, M.; LI, M.; CHEUNG, V. G. RNA-sequence analysis of human B-cells. **Genome Research**, v. 21, p. 991-998, 2011.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. TopHat: discovering splice junctions with RNA-Seq. **Bioinformatics**, v. 25, p. 1105-1111, 2009.

TRAPNELL, C.; WILLIAMS, B. A.; PERTEA, G.; MORTAZAVI, A.; KWAN, G.; BAREN, M. J.; *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, p. 511-515, 2010.

UIMARI, P.; SIRONEN, A.; SEVÓN-AIMONEN, M. L. Evidence for three highly significant QTL for meat quality traits in the Finnish Yorkshire pig breed. **Journal of Animal Science**, v. 91, p. 2001-2011, 2013.

VERNER, J.; HUMPOLICEK, P.; KNOLL, A. Impact of MYOD family genes on pork traits in Large White and Landrace pigs. **Journal of Animal Breeding and Genetics**, v. 124, p. 81-85, 2007.

VELCULESCU, V. E.; ZHANG, L.; VOGELSTEIN, B.; KINZLER, K. M. Serial analysis of gene expression. **Science**, v. 270, p. 484-487, 1995.

WALTON, P. E.; GOPINATH, R.; ETHERTON, T. D. Porcine insulin-like growth factor (IGF) binding protein blocks IGF-I action on porcine adipose tissue. **Proceedings of the Society for Experimental Biology and Medicine**, v. 190, p. 315-319, 1989.

WANG, E. T.; SANDBERG, R.; LUO, S.; KHREBTUKOVA, I.; ZHANG, L.; MAYR, C.; *et al.* Alternative isoform regulation in human tissue transcriptomes. **Nature**, v. 456, p. 470-476, 2008.

WANG, L.; FENG, Z.; WANG, X.; WANG, X.; ZHANG, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. **Bioinformatics**, v. 26, p. 136-138, 2010.

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

WHEELER, T. L.; CUNDIFF, L. V.; KOCH, R. M.; CROUSE, J. D. Characterization of biological types of cattle (Cycle IV): Carcass traits and Longissimus palatability. **Journal of Animal Science**, v. 74, p. 1023-1035, 1996.

- WHITAKER, L. On the Poisson law of small numbers. **Biometrika**, v. 10, p. 36-71, 1914.
- WICKRAMASINGHE, S.; RINCON, G.; ISLAS-TREJO, A.; MEDRANO, J. F. Transcriptional profiling of bovine milk using RNA sequencing. **BMC Genomics**, v. 13, p. 45-59, 2012.
- WINTER, A.; KRÄMER, W.; WERNER, F. A. O.; KOLLERS, S.; KATA, S.; DURSTEWITZ, G.; *et al.* Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, p. 9300-9305, 2012.
- WOLD, B.; MYERS, R. M. Sequence census methods for functional genomics. **Nature Methods**, v. 5, p.19-21, 2007.
- WYLIE, A. R. G. Leptin in farm animals: where are we and where can we go? **Animal**, v. 5, p. 246-267, 2011.
- XIONG, J.; LU, X.; ZHOU, Z.; CHANG, Y.; YUAN, D.; *et al.* Transcriptome analysis of the model protozoan, *tetrahymena thermophila*, using deep RNA sequencing. **PLoS One**, v. 2, p. e30630, 2012.
- XU, H. E.; LAMBERT, M. H.; MONTANA, V. G.; PARKS, D. J.; BLANCHARD, S. G.; BROWN, P. J.; *et al.* Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. **Molecular Cell**, v. 3, p. 397-403, 1999.
- YANG, A.; LARSEN, T. W.; SMITH, S. B.; TUME, R. K. D9 desaturase activity in bovine subcutaneous fatty acid composition. **Lipids**, v. 34, p. 971-978, 1999.
- YERLE, M.; ECHARD, G.; ROBIC, A.; MAIRAL, A.; DUBUT-FONTANA, C.; RIQUET, J.; *et al.* A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. **Cytogenetics Cell Genetics**, v. 73, p. 194-202, 1996.
- YOUNG, M. D.; WAKEFIELD, M. J.; SMYTH, G. K.; OSHLACK, A. Gene ontology analysis for RNA-seq: accounting for selection bias. **Genome Biology**, v. 11, p. R14, 2010.
- YUAN, Z.,; LI, J.; LI, J.; GAO, X.; GAO, H.; XU, S. Effects of DGAT1 gene on meat and carcass fatness quality in Chinese commercial cattle. **Molecular Biology Reports**, v. 40, p. 1947-1954, 2013.
- ZEEBERG, B. R.; FENG, W.; WANG, G.; WANG, M. D.; FOJO, A. T.; SUNSHINE, M.; *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. **Genome Biology**, v. 4, p. R28, 2003.

ZHANG, C.; WANG, G.; WANG, J.; JI, Z.; LIU, Z.; PI, X.; *et al.* Characterization and comparative analyses of muscle transcriptomes in Dorper and small-tailed Han sheep using RNA-Seq technique. **PLoS One**, v. 8, p. e72686, 2013.

ZHANG, Y. J.; PROENCA, R.; MAFFEI, M.; BARONE, M.; LEOPOLD, L.; FRIEDMAN, J. M. Positional cloning of the mouse obese gene and its human homologue. **Nature**, v. 372, p. 425-432, 1994.

ZHAO, X.; MO, D.; LI, A.; GONG, W.; XIAO, S.; ZHANG, Y.; *et al.* Comparative analyses by sequencing of transcriptomes during skeletal muscle development between pig breeds differing in muscle growth rate and fatness. **PLoS One**, v. 6, p. e19774, 2011.

ZHOU, X.; SU, Z. EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. **BMC Genomics**, v. 8, p. 246, 2007.

ZHOU, G.; DUDGEON, C.; LI, M.; CAO, Y.; ZHANG, L.; JIN, H. Molecular cloning of the HGD gene and association of SNPs with meat quality traits in Chinese red cattle. **Molecular Biology Reports**, v. 37, p. 603-611, 2010.