

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**CARACTERIZAÇÃO DE VARIAÇÃO NO NÚMERO DE
CÓPIAS (CNV) NO GENE *HMGA2* ASSOCIADO COM
TAMANHO DE PREPÚCIO EM BOVINOS NELORE (*BOS
INDICUS*)**

Tamiris Sayuri Aguiar

Médica

Veterinária

2018

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**CARACTERIZAÇÃO DE VARIAÇÃO NO NÚMERO DE
CÓPIAS (CNV) NO GENE *HMGA2* ASSOCIADO COM
TAMANHO DE PREPÚCIO EM BOVINOS NELORE (*BOS
INDICUS*)**

**Tamíris Sayuri Aguiar
Orientador: José Fernando
Garcia**

Dissertação apresentada à
Faculdade de Ciências Agrárias e
Veterinárias – Unesp, Câmpus de
Jaboticabal, como parte das exigências
para obtenção do título de Mestre em
Medicina Veterinária (Reprodução Animal).

2018

Ficha Catalográfica

A282c Aguiar, Tamíris Sayuri
Caracterização de variação no número de cópias (CNV) no gene
HMG2 associado com tamanho de prepúcio em bovinos nelore (*Bos indicus*) / Tamíris Sayuri Aguiar. – – Jaboticabal, 2018
vii, 40p. : 6 il.; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2018
Orientador: José Fernando Garcia
Banca examinadora: Flávia Lombardi Lopes, Silvana de Cássia
Paulan
Bibliografia

1. Bovinos. 2. *HMG2*. 3. CNV. 4. Umbigo. 5. *Bos taurus*. 6. *Bos indicus*. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 619:612.6:636.2

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

Certificado de aprovação



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: CARACTERIZAÇÃO DO GENE HMGA2 EM BOVINOS NELORE (*Bos indicus*)

AUTORA: TAMÍRIS SAYURI AGUIAR

ORIENTADOR: JOSÉ FERNANDO GARCIA

Aprovada como parte das exigências para obtenção do Título de Mestre em MEDICINA VETERINÁRIA, área: REPRODUÇÃO ANIMAL pela Comissão Examinadora:

Prof. Dr. JOSÉ FERNANDO GARCIA
Departamento de Apoio, Produção e Saúde Animal / FMVA-Araçatuba/SP

Dra. SILVANA DE CÁSSIA PAULAN
Bióloga Autônoma / Santa Fé do Sul/SP

Dra. FLÁVIA LOMBARDI LOPES
Departamento de Apoio, Produção e Saúde Animal / FMVA-Araçatuba/SP

Jaboticabal, 21 de junho de 2018

DADOS CURRICULARES DO AUTOR

TAMÍRIS SAYURI AGUIAR - Nascida em São Paulo – SP em 25 de setembro de 1990, filha de Flávio Lourenço Aguiar e Maria das Graças Mitsue Nishikawa Aguiar. Em 2009 ingressou no curso de Medicina Veterinária pela Faculdade de Medicina Veterinária de Araçatuba, UNESP campus de Araçatuba, obtendo título de Médica Veterinária em 2013. Trabalhou de dezembro de 2013 até setembro de 2014 na empresa Sexing Technologies, onde realizava as análises e processamento de sêmen sexado.

No período de junho a setembro de 2015, esteve na Università Cattolica Del Sacro Cuore, UCSC, Itália para a realização de um estágio de treinamento em Análises básicas de bioinformática em genômica, sob a orientação e supervisão do pesquisador professor Dr. Paolo Ajmone Marsan e professor Dr. José Fernando Garcia. Em Março de 2016 ingressou no Programa de Pós-Graduação em Medicina Veterinária, área de concentração em Reprodução Animal, da Faculdade de Ciências Agrárias e Veterinárias, UNESP campus de Jaboticabal como bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob orientação do Professor Dr. José Fernando Garcia.

“It is our choices that show what we truly are, far more than our abilities.”

J.K. Rowling

AGRADECIMENTOS

A Deus, por sempre me dar força e saúde.

Aos meus pais, pelo amor e apoio incondicional, e por nunca medirem esforços para me ajudar e orientar.

Ao meu orientador, professor Fernando Garcia por todas as oportunidades e ensinamentos. Além da paciência desprendida.

À todas as pessoas que estiveram ao meu lado e me ajudaram durante essa jornada em especial Vantuir, Fernando I., Adam, Marco, Carol de Marchi, Amanda Miyuki, Beatriz, Sarita, Larissa, Silvana, Fernanda, Pier.

À todos os colegas do LBBMA, pelo companheirismo.

A Professora Cárís, pelos conselhos e paciência.

À amiga Anirene, por compartilhar sempre as experiências comigo, e me dar força para nunca desistir

Ao Yuri, por ter sido minha luz ao fim do túnel, por ter tido paciência e disposição para me ajudar e ensinar em todos os “planos”.

Ao Marco Milanesi, pelos conselhos, ensinamentos e ajuda.

A professora Flávia Lombardi, Silvana Paulan, e professor Hans por terem gentilmente aceitado fazer parte da qualificação e defesa desse trabalho, contribuindo para melhorias.

Ao Caíque, que nunca duvidou da minha capacidade, e acima de tudo me fez acreditar nela.

Ao professor Paolo-Ajmone, pela oportunidade, pelos conselhos e por ser um exemplo para mim.

À todos que, de alguma maneira, contribuíram para a minha formação profissional com orientações, críticas ou elogios.

Jaboticabal, Julho, 2018

APOIO FINANCEIRO

Este projeto obteve bolsa do país concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob processo nº 132809/2016-8 166419/2017-6.

SUMÁRIO

RESUMO	iii
ABSTRACT	iv
LISTA DE ABREVIACÕES E SIMBOLOS	v
LISTA DE FIGURAS	vii
Introdução.....	1
Revisão de Literatura.....	2
Referências.....	10
CAPÍTULO 2: Association of a copy number variant with navel in <i>Bos indicus</i> reinforces the relevance of the HMGA2-PLAG1-IGF2 pathway in cattle breeding	19
Abstract.....	20
Introduction	21
Material and Method	22
Genotypes.....	22
Phenotypes.....	23
Genome-wide association analysis.....	23
Analysis of whole genome sequence data.....	25
SNP probe intensity analysis.....	26
qPCR.....	26
Results.....	27
GWAS identifies a <i>HMGA2</i> haplotype associated with navel at yearling.....	27

Nellore sequence data suggests a CNV intronic to <i>HMGA2</i> as the causal variant	28
Genome sequences from nine cattle breeds indicate a <i>B. indicus</i> origin for the <i>HMGA2</i> -CNVR	28
qPCR validates haplotype-based prediction of CNV genotypes.....	29
Conventional GWAS and SNP probe intensity data fail to detect the CNV	30
Discussion	32
Author Contributions.....	33
Funding.....	34
Reference.....	34

CARACTERIZAÇÃO DE VARIAÇÃO NO NÚMERO DE CÓPIAS (CNV) NO GENE *HMGA2* ASSOCIADO COM TAMANHO DE PREPÚCIO EM BOVINOS NELORE (*BOS INDICUS*)

RESUMO - O gene High Mobility Group AT-hook2 (*HMGA2*) apresentou fortes evidências de estar associado com o tamanho de umbigo em bovinos da raça Nelore através das análises de associação genômica ampla (GWAS). Diversos relatos de associação desse gene a fenótipos do âmbito da morfologia corporal existem para diferentes espécies, tais como altura em humanos, cães e equinos, e tamanho da orelha em suínos. Descobertas recentes demonstraram que o gene *HMGA2* está associado avia metabólica de grande importância fisiológica e biológica que tem como um dos principais fatores o *PLAG1* (Pleomorphic adenoma gene1), que está associado ao fator de crescimento semelhante à insulina 2 (*IGF2*), importante regulador do crescimento e da reprodução em bovinos. No presente trabalho, foi descrita a identificação e caracterização de variação no número de cópias (CNV) cromossomo 5 do cromossomo bovino, na região do gene *HMGA2* que apresenta associação a característica de tamanho de umbigo. Análises da sequência completa do genoma de indivíduos *Bos taurus* e *Bos indicus* foram empregadas para caracterizar o CNV, sendo sua validação realizada através de PCR quantitativo (qPCR). Além disso, os resultados foram comparados com dados de sequência de animais africanos *B. indicus* evidenciando a origem zebuína do CNV.

Palavras-chave: Bovinos, *HMGA2*, CNV, Umbigo, *Bos taurus*, *Bos indicus*.

CHARACTERIZATION OF VARIATION IN THE NUMBER OF COPIES (CNV) IN THE HMGA2 GENE ASSOCIATED TO NAVEL SIZE IN NELORE BOVINES (BOS INDICUS)

ABSTRACT –The High Mobility Group AT-hook 2 (HMGA2) gene presented strong evidence of being associated with navel size in Nelore cattle through genome association analysis (GWAS). Several reports of association of this gene with phenotypes in the scope of body morphology exist for different species, such as height in humans, dogs and horses, and ear size in swine. Recent discoveries have shown that the HMGA2 gene is associated with a metabolic pathway of great physiological and biological importance that has as one of its main factors PLAG1 (Pleomorphic adenoma gene 1), which is associated with insulin-like growth factor 2 (IGF2), important regulator of growth and reproduction in cattle. In the present work, the identification and characterization of copy number variation (CNV) on chromosome 5 of the bovine chromosome in the region of the HMGA2 gene that has association with the navel size trait was described. Genome sequence analysis of *Bos taurus* and *Bos indicus* individuals was used to characterize the CNV, and its validation was performed using quantitative PCR (qPCR). In addition, the results were compared with sequence data from a African *B.indicus* animals evidencing the indicine origin of the CNV.

Keywords: Bovine; CNV; Navel; *Bos taurus*; *Bos indicus*.

LISTA DE ABREVIACÃO

3'-UTR	3'-Untranslated Region
bp	Base pairs
CHR	Chromosome
CNV	Copy Number Variation
DNA	Deoxyribonucleic Acid
dEBV	Deregressed Estimated Breeding Value
EBV	Estimated Breeding Value
GRM	Genomic Relationship Matrix
GWAS	Genome Wide Association Study
BovineHD	Illumina® BovineHD Bead Chip Assay
<i>HMGA2</i>	High Mobility Group AT-hook 2 gene
<i>IGF2</i>	Insulin-like Growth Factor 2
<i>IGF</i>	Insulin-like Growth Factor
IGV	Integrative GenomicsViewer
kbp	Kilobasepairs (10^3 base pairs)
LOCO	Leave-one-chromosome-out
LD	Linkage Disequilibrium
LRR	Total Signal Intensity
Mbp	Mega base pairs (10^6 base pairs)
miRNA	Micro Ribonucleic Acid
NY	Navel at Yearling
mRNA	Messenger RibonucleicAcid
PCR	Polymerase ChainReaction
<i>PLAG1</i>	Pleomorphic Adenoma gene1

qPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Trait Locus
RNA	Ribonucleic Acid
SHAPEIT2	Segmented HAPlotype Estimation & ImputationTool SNP Single NucleotidePolymorphism
SNP	Single NucleotidePolymorphism
	<i>TSSC4</i> Tumor Suppressing Subtransferable Candidate 4 gene
VEP	Variant Effect Predictor

LISTA DE FIGURAS

CÁPITULO 2—Association of a copy number variant with navel in *Bos indicus* reinforces the relevance of the *HMGA2-PLAG1-IGF2* pathway in cattle breeding

Figure 1. *HMGA2-PLAG1-IGF2* pathway scheme. Protein tertiary structures displayed in this figure were build using SWISS-MODEL

Figure 2. Haplotype-based GWAS maps navel at yearling associations to *HMGA2*. (a) Each point in the Manhattan plot corresponds to a 6-markers long haplotype. The dashed horizontal line corresponds to the Bonferroni threshold ($p < 2.44 \times 10^{-8}$). (b) Distributions of dEBVs according to number of copies of the leading haplotype. The *HMGA2* locus accounted for 2.04% of the variance in dEBVs.

Figure 3. Discovery of a *B. indicus* specific CNV on *HMGA2* affecting navel at yearling. (a) Relative fold increase in sequence coverage of a segment of *HMGA2* intron 3 correlates with TCCTCCAAC haplotype counts in Nellore cattle. Each smoothed curve corresponds to sequence coverage averaged across samples with same haplotype count. (b) Inspection of the CNV region in additional European *B. taurus* and African *B. indicus* (breeds reveals specificity of copy gains in *B. indicus* (marked with *).

Figure 4. Navel at yearling associations with varying haplotype sizes (1, 5 and 10 SNPs). Dashed horizontal lines correspond to the Bonferroni-corrected significance levels.

Figure 5. Average probe intensity data from the BovineHD assay for the *HMGA2*-CNVR segment (vertical dashed lines). Each curve corresponds to average LRR values of samples with the same TCCTCCAAC haplotype count. Points in the curves indicate positions of SNP markers.

CÁPITULO 1 – Considerações Gerais

1. RESUMO

O Brasil possui o segundo maior rebanho efetivo do mundo, porém no qual apenas 12% das fêmeas em idade reprodutiva são inseminadas (ASBIA, 2016). Isso faz com que o uso de touros na monta natural revista-se de grande importância o que consequentemente implica na necessidade de atenção para com a sanidade detouros reprodutores.

Em termos genéticos, os touros possuem grande importância, especialmente quando comparados com vacas, uma vez que são capazes de deixar centenas de descendentes no rebanho. Entre diversas características a serem observadas na funcionalidade de um touro, a ocorrência de desordens físicas na parte externa do sistema reprodutivo masculino, especialmente no prepúcio penduloso que é mais susceptível a injúrias, trata-se daquela que pode levar a danos irreversíveis na capacidade reprodutiva consequentemente ocasionando perdas econômicas significativas. Nesse sentido, nos processos de seleção e melhoramento genético de bovinos de corte, a característica prepúcio/umbigo é levada em consideração por tratar-se de fenótipo que apresenta alta herdabilidade (VIU, et al.,2002, BIGNARDI,et al.,2011).

Atualmente, diversos programas de melhoramento genético da raça Nelore utilizam avaliação visual conhecida genericamente como CPMU (Conformação; Precocidade; Musculatura e Umbigo), na qual se atribuem escores visuais individuais de um a cinco para os fenótipos: conformação, precocidade, musculatura, e umbigo (KOURY FILHO, 2005).

Entretanto, apesar de ser um dos fenótipos de maior importância econômica na raça, o tamanho de bainha prepucial para machos e o tamanho do umbigo para fêmeas é muitas vezes negligenciado.

Estudos de associação genômica ampla (GWAS) têm por finalidade analisar variações na sequência de DNA genômico, buscando associações com características fenotípicas definidas, permitindo a localização da região cromossômica e, eventualmente, dos genes envolvidos na expressão do fenótipo de

interesse, auxiliando na compreensão dos mecanismos biológicos e fisiológicos desses genes (KU et al., 2010).

A presença de um QTL de alta significância associado a característica umbigo identificado por GWAS foi encontrado na região do cromossomo 5 bovino que contém o gene *HMGA2* (Porto-Neto et al., 2014) que recentemente foi relacionado como fenótipo altura corporal em humanos (Weedon et al., 2007; Yang et al., 2010).

O principal objetivo da presente dissertação de mestrado foi de identificar a variante causal para o fenótipo tamanho de umbigo de bovinos da raça Nelore através do GWAS, e avançar na compreensão e determinação da variante causal para essa característica que possam auxiliar a seleção genética principalmente de reprodutores.

No Capítulo 2, descreve-se a identificação de uma região candidata para o fenótipo umbigo em bovinos, na qual está localizado o gene *HMGA2*. Além disso, através das técnicas de análise de dados de sequência completa de DNA, GWAS e qPCR, foi possível determinar CNV no gene *HMGA2* como provável responsável pela determinação do fenótipo, bem como identificar via pleiotrópica de grande importância.

2.REVISÃO DE LITERATURA

2.1. PANORAMA DA PECUÁRIA BOVINA NO BRASIL E NO MUNDO

A pecuária bovina possui forte representatividade nas exportações no mercado interno brasileiro, gerando aproximadamente seis bilhões de reais no ano de 2015. Esse setor representa de 6% do PIB brasileiro e 30% do PIB do agronegócio (GOMES et al., 2017). Segundo dados do USDA (USDA, 2018) em 2018 houve aumento de 5% na exportação global de carne mundial *in natura*, sendo que o Brasil ocupa posição de destaque nesse panorama, sendo o maior exportador de carne bovina em 2017 (2.02 milhões de toneladas). O Brasil ocupa também o segundo lugar em produção global de carne, sendo essa expansão devida ao aumento do peso de carcaça, da demanda doméstica e por exportações (9.9

milhões de toneladas), além do terceiro lugar como mercado consumidor, ficando atrás apenas da China e dos Estados Unidos.

Diante desse contexto de liderança mundial, é de crucial importância para o Brasil o desenvolvimento e aplicação de métodos de seleção genética para as características associadas à eficiência produtiva, com o objetivo de aperfeiçoar a cadeia da carne como um todo.

2.2. RAÇA NELORE

As raças zebuína (*Bos indicus*), originárias da Índia, foram introduzidas no Brasil a partir do século XIX, difundindo-se por todo território, sendo a raça Nelore predominante no cenário pecuário brasileiro atual.

Uma das características da raça Nelore que possibilitou sua rápida expansão foi a alta adaptabilidade aos sistemas de produção, tanto extensivos quanto intensivos, uma vez que possui a capacidade de digerir alimentos de baixa qualidade e convertê-lo sem proteína animal com maior eficiência quando comparados a outras raças, tornando-a atrativa para os produtores brasileiros (KOURY FILHO et al., 2003).

Outros fatores importantes que devem ser considerados com relação à raça são a ocorrência de pelagem branca e de pele preta (o que confere maior tolerância ao clima tropical e subtropical), a presença do cupim, a rusticidade e resistência a endo e ectoparasitas e, por fim, a eficiência na dissipação do calor endógeno por condução, convecção, irradiação e evaporação (HANSEN, 2004; TAKAHASHI et al., 2009).

2.3. UMBIGO/PREPÚCIO: PROBLEMAS REPRODUTIVOS

O tamanho do prepúcio é uma característica morfológica de importante relevância nos bovinos, principalmente nos reprodutores zebuínos, visto que estes possuem prepúcio mais penduloso do que aqueles dos animais taurinos

(NASCIMENTO e SANTOS, 2011) tornando-os mais predispostos a injúrias e traumatismos (RABELO et al., 2008).

Práticas de manejo equivocadas como a falta de controle de endo e ectoparasitas, uso de pastagens de baixa qualidade, mal preparadas e manejadas, são fatores que podem contribuir para o desenvolvimento de traumas no sistema reprodutivo do macho e, conseqüentemente, desencadear enfermidades como a acrobustite, inflamação na extremidade externa do prepúcio que, quando associada ao estreitamento do óstio prepucial dificulta a exposição peniana devido à dor, causando a incapacidade da cópula, além de processos inflamatórios e lesões penianas, contribuindo assim para redução da fertilidade geral do rebanho (RABELO et al., 2012).

Sugere-se que aqueles animais que apresentem prepúcio longo, sejam eliminados do rebanho em razão da alta herdabilidade genética desta característica. A herdabilidade para tamanho do prepúcio/umbigo varia entre 0,29 a 0,70 (LAGOS & FITZHUGH Jr., 1970; LIMA et al., 1989; CARDOSO et al., 1998; KOURY FILHO et al., 2003; VIU et al., 2002; BIGNARDI et al., 2011; LIMA et al., 2013).

Diante do peso da produção animal para a economia brasileira e de como asperdas impostas pelas falhas reprodutivas, através da diminuição da vida reprodutiva do touro, falha na cobertura, impotência *coeundi* e altos custos de tratamentos que muitas vezes se mostram ineficazes, essa condição é um limitante importante da eficiência produtiva (ASHDOWN, 2006; RABELO et al., 2006; RABELO et al., 2008, BOLIGON, 2016).

Sendo assim, o principal método utilizado para atenuar os problemas advindos da ocorrência de umbigo penduloso é a seleção direta através dos escores de bainha prepucial (machos) e umbigo (fêmeas) de acordo com seu tamanho e posicionamento. Salienta-se que a seleção de tamanho do umbigo para as fêmeas não possui valor funcional, servindo apenas como um indicativo de seleção para sua progênie.

2.4. COLETA DE DADOS FENOTÍPICOS

Programas de melhoramento genético de bovinos de corte utilizam características fenotípicas coletadas na desmama (sete meses) e ao sobreano (18 meses) como parâmetros para a seleção uma vez que essas possuem baixo custo e alta acurácia. O programa CPMU (Conformação, Precocidade, Musculosidade e Umbigo) desenvolvido por Koury Filho et al. (2003), atribui escore visual linear para cada animal de um grupo em relação a aspectos de sua estrutura corporal, os quais são integrados através de cálculos matemáticos que estabelecem valor chamado DEP (diferença esperada entre progênie) que leva em consideração as informações de desempenho individuais e da progênie, além do pedigree.

Essas avaliações visuais são realizadas em animais pertencentes ao mesmo grupo de contemporâneos (animais do mesmo sexo, ano de nascimento, e mesmo lote no rebanho) por dois avaliadores capacitados, ocasião em que é atribuída nota individual em escala de 1 a 5. Para a avaliação de umbigo/bainha prepucial, os maiores escores correspondem aos animais com a bainha prepucial mais pendulosa, e os menores para animais com a bainha mais retraída.

2.5. ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)

No Brasil, onde bovinos de raças zebuínas desempenham papel importante na cadeia produtiva, o uso de ferramentas genômicas para melhorar a eficiência de produção vem sendo adotado na última década (UTSUNOMIYA et al., 2013; SANTANA et al., 2014; PEREIRA et al., 2016).

Uma das mais importantes perspectivas para aplicação dos conhecimentos genômicos na prática, consiste na abordagem conhecida como Estudo de Associação Genômica Ampla (GWAS – “Genome Wide Association Study”), que utiliza informações sobre o tipo e frequência alélica de marcadores genômicos do tipo SNP, os quais consistem das variações pontuais (mutações de um nucleotídeo) que ocorrem e se acumulam no DNA através das gerações, permitindo seu uso na diferenciação de indivíduos e na determinação das causas de diferenças fenotípicas observadas.

Os resultados do GWAS permitem a identificação de regiões ou variantes associadas com características complexas de interesse, aumentando o entendimento sobre importantes vias moleculares e biológicas além de indicar genes candidatos para determinados fenótipos (SCOTT et al., 2007).

Os primeiros estudos desse tipo foram realizados em humanos, na área epidemiológica (MCCARTHY et al., 2008), no entanto, atualmente essa abordagem é utilizada em diversas espécies, principalmente em espécies de interesse agropecuário.

Do ponto de vista das alterações na forma e estrutura dos cromossomos, de forma geral as variações genômicas podem ser divididas em variantes neutras, como é o caso dos SNP, e variantes estruturais, quando ocorrem em mais de um par de bases e podem ser subdivididas em: variações no número de cópias (CNV), inserções-deleções (INDELS), inversões e substituições (KORN et al., 2008; BICKHART et al., 2012).

Recentemente, haplótipos (conjuntos específicos de alelos SNP em tandem e contidos em região específica de um determinado cromossomo) tem sido utilizado em diversos estudos de associação genômica obtendo-se resultados semelhantes aos obtidos como uso de SNPs (HAYES et al., 2007).

2.6. HMGA2

O gene High Mobility Group AT-hook 2" (HMGA2) codifica proteínas pertencentes ao grupo *HMG* (do inglês: non-histone chromosomal high mobility group protein family). Essas proteínas atuam regulando a transcrição gênica por alterações de cromatina e estrutura dos cromossomos.

O *HMGA2* é expresso principalmente durante o desenvolvimento embrionário nos rins, fígado e útero, sendo que em tecidos adultos é quase indetectável, contudo, quando expresso é normalmente associado com formas tumorais malignas e benignas, além de regular direta e indiretamente a expressão de diversos genes.

Deleções, amplificações e rearranjos nesse gene estão diretamente associados a lipomas, evidenciando seu papel na adipogênese, e diferenciação

mesenquimal (ASHAR et al., 1995). Além disso, estudos *in vitro* demonstraram que a presença da proteína *HMGA* leva ao aumento da atividade proliferativa dos condrócitos (RICHTER et al., 2009).

Estudos com camundongos nocaute demonstraram que esse gene está associado à obesidade induzida por dieta. Zhou et al. (1995) identificou o fenótipo *pygmy*, onde ocorrem baixos níveis de gordura e falhas no crescimento pré e pós natal. Já Weedon e colaboradores (2007) e Yang et al. (2010) caracterizaram a associação de *HMGA2* com altura em humanos, sendo o mesmo tipo de associação identificado em eqüinos (FRISCHKNECHT et al., 2015; MAKVANDI-NEJAD et al., 2012) e cães (HAYWARD et al., 2016). Em suínos, o tamanho de orelha foi associado com efeitos desse gene (LI et al., 2012) bem como em humanos quanto à erupção dos dentes permanentes (GELLER et al., 2011).

O gene *HMGA2* é altamente expresso durante a fase embrionária e início do desenvolvimento fetal, após essa fase sua expressão começa a ser regulada através do micro RNA let-7 causando sua supressão, no entanto, em casos de neoplasias em adultos há uma inversão dessa regulação, onde há um aumento da expressão do *HMGA2* e a supressão do let-7 (Shell et al., 2007).

O fator de transcrição *HMGA2* possui cinco exons, sendo extremamente conservado entre os mamíferos. Os três primeiros exóns codificam os *AT-hook-domain*, que podem regular grande números de genes alvos. Já os dois últimos exóns codificam as “*acid tail regions*”.

Estudos realizados por KLEMKE et al. (2014) e HABIB et al. (2017), apontaram que o gene do *HMGA2* participa da via metabólica (*HMGA2/PLAG1/IGF2*) através da auto-regulação dos dois genes de extremam importância: *PLAG1* e *IGF2*. O *HMGA2* pode atuar diretamente na regulação do *PLAG1* e do *IGF2*, ou através do aumento da expressão de *PLAG1* que indiretamente influencia no aumento da expressão de *IGF2*.

O *IGF* do inglês “insulin-like growth factor” é responsável pela via do crescimento celular, inibição da apoptose e proliferação celular (ROSENZWEIG e ATREYA, 2010; KING e WONG, 2012), sendo os principais participantes desse complexo sistema o *IGF1* e *IGF2*, que atuam diretamente na deficiência de crescimento em ratos (DeCHIARA et al., 1991). Assim como o *IGF2*, o nocaute do

gene *PLAG1* (Pleomorphic Adenoma Gene 1) em camundongos levou ao retardo no crescimento e diminuição da fertilidade (HENSEN et al., 2004).

2.7. CNV

Alterações no número de cópias (CNV) de um segmento específico da sequência do genoma, através de duplicações ou deleções, são ocorrências comumente descritas em várias espécies, tais como humanos (MILLS et al., 2011; FU et al., 2010); ovinos (FONTANES et al., 2011), caprinos (FONTANES et al., 2010), suínos (CHEN et al., 2012; LI et al., 2012; WANG et al., 2013) aves (SKINNER et al., 2009; WANG et al., 2010; LUO et al., 2013; ABERNATHY et al., 2014) e bovinos (BAE et al., 2010; LIU et al., 2010; BICKHART et al., 2012; JIANG et al., 2013). Essas modificações são essenciais para a diversidade genética e podem também desempenhar funções importantes em diversas doenças em humanos, como o Alzheimer (ROVELET-LECRUX et al., 2006), Parkinson (SIMON-SANCHEZ et al., 2008), autismo (SEBAT et al., 2007), entre outras.

Embora as alterações de SNP ocorram em abundância no genoma, os efeitos de CNV sobre fenótipos tendem a apresentar maior impacto por se tratarem de mudanças estruturais no cromossomo que podem alterar a regulação gênica atingindo regiões codificantes de proteínas ou aquelas implicadas no seu controle (SEBAT et al., 2004; STRANGER et al., 2007; ZHANG et al., 2009).

A existência de múltiplas cópias de um gene no genoma pode ser considerada como mecanismo alternativo para garantir o processo evolutivo ao manter a existência e funcionamento de genes essenciais (KONRAD, 2011), além de também poderem estar associados com diferenças adaptativas entre raças geradas pelos processos de seleção artificial (BICKHART et al., 2012).

Atualmente há quatro mecanismos prováveis para a origem dos CNVs: a recombinação homóloga não alélica, a junção de extremidades não homólogas, ocorrência de um erro durante a replicação, denominado Fork Stalling and "Templateswitching", e a retrotransposição (CLOP et al., 2012).

No Capítulo 2 foi abordada a validação do gene *HMGA2* como possível variante causal para o fenótipo tamanho de umbigo com o auxílio das ferramentas GWAS, análise de dados de sequência genômica e qPCR.

3.REFERÊNCIAS

ABERNATHY, J.; LI, X.; JIA, X.; CHOU, W.; LAMONT, S. J.; CROOIJMANS, R.; ZHOU, H. Copy number variation in Fayoumi and Leghorn chickens analyzed using array comparative genomic hybridization. **Animal Genetics**, v. 45, p. 400-411, 2014.

ABIEC. **Associação Brasileira das Indústrias Exportadoras de Carnes** .
Rebanho Bovino Brasileiro. Disponível em:
http://www.abiec.com.br/download/stat_mercadomundial.pdf Acesso em: 23 de mar de 2018.

ABI-HABIB, W.; BRIOUDE, F.; EDOUARD, T.; BENNETT, J.T.; LIENHARDT-ROUSSIE, A.; TIXIER, F.; SALEM, J.; YUEN, T. Genetic disruption of the oncogenic *HMGA2-PLAG1-IGF2* pathway causes fetal growth restriction. **Genetics in Medicine**. 2017.

ASBIA. **Associação Brasileira de Inseminação Artificial. Relatório índice do ano de 2014**. Disponível em: <http://www.asbia.org.br/novo/relatorios/>
Acesso em: 13 de Ago de 2017

ASHAR, H. R.; SCHOENBERG FEJZO, M.; TKACHENKO, A.; ZHOU, X.; FLETCHER, J. A.; WEREMOWICZ, S.; MORTON, C. C.; CHADA, K. Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains. **Cell** 82: 57-65, 1995.

ASHDOWN, R.R. Functional, developmental and clinical anatomy of the bovine penis and prepuce. CAB Reviews: Perspectives in agriculture, **Veterinary Science, Nutrition and Natural Resources**, London, v.1, n.21, p.29-37, 2006.

BAE J.S.; CHEONG H.S.; KIM L.H.; NAMGUNG, S.; PARK, T.J.; CHUN, J.Y.; KIM, J.Y.; PASAJE, C.F.; LEE, J.S.; SHIN H.D. Identification of copy number variations and common deletion polymorphisms in cattle. **BMC Genomics** v.11. 2010

BICKHART, D. M.; HOU Y, S. S. G.; ALKAN, C.; CARDONE, M. F.; MATUKUMALLI, L. K.; SONG, J.; SCHNABEL, R. D.; VENTURA, M.; TAYLOR, J. F.; GARCIA, J. F.; VAN TASSELL, C.P.; SONSTEGARD, T. S.; EICHLER, E. E.; LIU, G. E. Copy

number variation of individual cattle genomes using next-generation sequencing. **Genome Res**, v.22. p.778 -790. 2012.

BIGNARDI, A.B. et al . Parâmetros genéticos de escore visual do umbigo em bovinos da raça Nelore. **Arq. Bras. Med. Vet. Zootec.**, Belo Horizonte, v. 63, n. 4, p. 941-947, 2011.

BOLIGON, A. A., DE VARGAS, L., SILVEIRA, D.D.; ROSO, V. M.; CAMPOS, G. S.; VAZ, R. Z.; SOUZA, F. R. P.; Genetic models for breed quality and navel development scores and its associations with growth traits in beef cattle. **Trop Anim Health Prod**. v. 48, 2016.

BOYKO, A. R.; QUIGNON, P.; LI, L.; SCHOENEBECK, J. J.; DEGENHARDT, J. D.; LOHMUELLER, K. E.; ZHAO, K.; BRISBIN, A.; PARKER, H. G.; VONHOLDT, B. M.; CARGILL, M.; AUTON, A.; REYNOLDS, A.; ELKAHLOUN, A. G.; CASTELHANO, M.; MOSHER, D. S.; SUTTER, N. B.; JOHNSON, G. S.; NOVEMBRE, J.; HUBISZ, M. J.; SIEPEL, A.; WAYNE, R. K.; BUSTAMANTE, C. D.; OSTRANDER, E. A. A simple genetic architecture underlies morphological variation in dogs. **PLoS Biol** v. 8, p. 49-50, 2010.

CARDOSO, F.F.; CARDELINO, R.A.; CAMPOS, L.T. Utilização de um escore de avaliação visual para seleção do tamanho do umbigo em bovinos da raça Santa Gertrudis. In: **Simpósio nacional de melhoramento animal**, 2, Uberaba. Anais. Uberaba: SBMA, p.385-386.1998.

CHEN, C.; QIAO, R.; WEI, R.; GUO, Y.; AI, H.; MA, J.; REN, J.; HUANG, L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. **BMC Genomics**. v. 13. 2012.

CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. **Anim Genet**.v.45, p. 503-517. 2012.

DECHIARA, T. M., ROBERTSON, E. J., & EFSTRATIADIS, A. Parental imprinting of the mouse insulin-like growth factor II gene. **Cell**, v.64, p. 849–859, 1991.

ESTADOS UNIDOS. Department of Agriculture. USDA.gov - United States Department of Agriculture. Livestock and Poultry: World Markets and Trade Disponível em: https://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf. Acesso em: 25 março 2018.

FONTANESI, L.; MARTELLI, P.L.; BERETTI, F.; RIGGIO, V.; DALL'OLIO, S.; COLOMBO, P.; CASADIO, R.; RUSSO, V.; PORTOLANO, B. An initial comparative map of copy number variations in the goat (*Capra hircus*) genome, **BMC Genomics** v.11, 2010.

FONTANESI, L.; BERETTI, F.; MARTELLI, P. L.; COLOMBO, M.; DALL'OLIO, S.; OCCIDENTE, M.; PORTOLANO, B.; CASADIO, R.; MATASSINO, D.; RUSSO, V. A first comparative map of copy number variations in the sheep genome. **Genomics**. v.97. p. 158-165, 2011.

FU, W.; ZHANG, F.; WANG, Y.; GU, X.; JIN, L. Identification of copy number variation hotspots in human populations. *Am J Hum Genet.* v.87. p. 494 -504. 2010.

FRISCHKNECHT, M.; JAGANNATHAN, V.; PLATTET, P.; NEUDITSCHKO, M.; SIGNER-HASLER, H.; BACHMANN, I.; PACHOLEWSKA, A.; DRÖGEMÜLLER, C.; DIETSCHI, E.; FLURY, C.; RIEDER, C.; LEEB, T. A Non-Synonymous *HMGA2* Variant Decreases Height in Shetland Ponies and Other Small Horses. **PLOS ONE** v.10, 2015.

GOMES, R. C.; FEIJO, G. L. D.; CHIARI, L.; Evolução e Qualidade da Pecuária Brasileira. Campo Grande: **Embrapa**: 2017 (Nota técnica).
HANSEN, P. Physiological and cellular adaptations of zebu cattle to thermal stress. **Animal Reproduction Science**, v. 82-83, p. 349–360. 2004.

HAYWARD, J. J.; CASTELHANO, M. G.; OLIVEIRA, K.C.; COREY, E.; BALKMAN, C.; BAXTER, T.L.; CASAL, M.L.; CENTER, S.A.; FANG, M.; GARRISON, S.J.; KALLA, S.E.; KORNILIEV, P.; KOTLIKOFF, M.I.; MOISE, N. S.; SHANNON, L.M.; SIMPSON, K.W.; SUTTER, N.B.; TODHUNTER, R.J.; BOYKO, A.R. Complex disease and phenotype mapping in the domestic dog. **Nature. Commun.** v.7, p.1–11, 2016.

HENSEN, K.; BRAEM, C.; DECLERCQ, J.; VAN DYCK, F.; DEWERCHIN, M.; FIETTE, L.; DENEFF, C.; VAN DE VEN, W. J. M. Targeted disruption of the murine *Plag1* proto-oncogene causes growth retardation and reduced fertility. **Development Growth and Differentiation**, v. 46, p. 459-470, 2004.

HAYES BJ, CHAMBERLAIN AJ, McPARTLAN H, MACLEOD I, SETHURAMAN L, GODDARD ME. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. **Genet Res.** 2007;89: 215–20.
JIANG, L.; JIANG, J.; YANG, J.; LIU, X.; WANG, H.; DING, X.; LIU, J.; ZHANG, Q. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. **BMC Genomics**, v.14, p.131, 2013.

KING, E. R.; WONG, K. Insulin-like Growth Factor: Current Concepts and New Developments in Cancer Therapy. **Recent Patentson Anti-Cancer Drug Discovery**, v. 7, p. 14-30, 2012.

KOURY FILHO, W.; JUBILEU J.S.; ELER J.P.; FERRAZ, J. B. S.; PEREIRA, E.; CARDOSO, E. P. Parâmetros genéticos para escores de umbigo e características de produção em bovinos da raça Nelore. **Arq. Bras. Med. Vet. Zoot.**, v. 55 n. 5, p. 594-598, 2003.

KOURY FILHO, W. Escores visuais e suas relações com características de crescimento em bovinos de corte. 2005. **Tese (Doutorado em Zootecnia – Produção Animal)** - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Julio de Mesquita Filho”, Jaboticabal, 2005.

KONRAD, A.; TEUFEL, A. I.; GRAHNEN, J. A.; LIBERLES, D. A. Toward a general model for the evolutionary dynamics of gene duplicates. **Genome BiolEvol**, v.3, p.1197-209, 2011.

KORN, J. M.; KURUVILLA, F. G.; MCCARROLL, S. A.; WYSOKER, A.; NEMESH, J.; CAWLEY, S.; HUBBELL, E.; VEITCH, J.; COLLINS, P. J.; DARVISHI, K.; LEE, C.; NIZZARI, M. M.; GABRIEL, S. B.; PURCELL, S.; DALY, M. J.; ALTSHULER, D. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. **Nature Genetics**. v. 40, p. 1253–1260, 2008.

KU, C. S.; LOY, E. Y. The pursuit of genome-wide association studies: where are we now? **Journal of Human Genetics**, v. 55, p.195–206, 2010.

KUBISTA, M.; ANDRADE, J. M.; BENGTSSON, M.; FOROOTAN, A.; JONÁK, J.; LIND, K.; SINDELKA, R.; SJÖBACK, R.; SJÖGREEN, B.; STRÖMBOM, L.; STÅHLBERG, A.; ZORIC, N. The real-time polymerase chain reaction. **Molecular Aspects of Medicine**, v. 27, p. 95-125, 2006.

KLEMKE, K.; MÜLLER, M. H.; WOSNIOK, W.; MARKOWSKI, D. N.; NIMZYK, R.; MARIA HELMKE, B.M.; BULLERDIEK, J. Correlated Expression of HMGA2 and PLAG1 in Thyroid Tumors, Uterine Leiomyomas and Experimental Models. v.9, **PLoS ONE**, 2014.

LAGOS, F.; FITZHUGH Jr, H.A. Factors influencing preputial prolapse in yearling bulls. **J. Anim. Sci.** v.30, n.6, p.949-52, 1970.

LI, Y.; MEI, S.; ZHANG, X.; PENG, X.; LIU, G.; TAO, H.; WU, H.; JIANG, S.; XIONG, Y.; LI, F. Identification of genome-wide copy number variations among diverse pig breeds by array CGH. **BMC Genomics**, v. 13, p. 725, 2012.

LI, P.; XIAO, S.; WEI, N.; ZHANG, Z.; HUANG, R.; GU, Y.; REN, J.; HUANG, L.; CHEN, C. Fine mapping of a QTL for ear size on porcine chromosome 5 and identification of high mobility group AT-hook 2 (*HMGA2*) as a positional candidate gene. **Genetics, Selection, Evolution**, v. 44, 2012.

LIU, G.E.; HOU, Y.; ZHU, B.; CARDONE, M. F.; JIANG, L.; CELLAMARE, A.; MITRA, A.; ALEXANDER, L. J. COUTINHO, L. L.; ELENA, M.; AQUILA, D.; GASBARRE, L. C.; LACALANDRA, G.; LI, R. W.; MATUKUMALLI, L. K.; NONNEMAN, D.; REGITANO, L. C. DE A.; SMITH, T. P. L.; SONG, J.; SONSTEGARD, T. S.; TASSELL, C. P.V.; VENTURA, M.; EICHLER, E. E.; MCDANELD, T. G.; KEELE, J. W. Analysis of copy number variations among diverse cattle breeds. **Genome Research**, p. 693-703. 2010.

LIMA, F.P.; BONILHA NETO, L.M.; RAZOOK, A.G. PACOLA, L. J.; FIGUEIREDO, L. A. D.; PEIXOTO, A. M. Parâmetros genéticos em características morfológicas de bovinos Nelore. **Bol. Indúst. Anim**, Nova Odessa, v. 46, n. 2, p. 249-257, 1989.

LIMA, P. R. M.; PAIVA, S. R.; COBUCCI, J. A.; NETO, J. B.; MACHADO, C. H. C.; MCMANUS, C. Genetic parameters for type classification of Nelore cattle on central performance tests at pasture in Brazil. **Trop Anim Health Prod**. v. 45, p. 1627 – 1634, 2013.

LIVAK, Kenneth J.; SCHMITTGEN, Thomas D. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$ method. **methods**, v. 25, n. 4, p. 402-408, 2001.

LUO, J.; YU, Y.; MITRA, A.; CHANG, S.; ZHANG, H.; LIU, G.; YANG, N.; SONG, J. Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. G3: **Genes - Genomes – Genetics**, v. 3, p. 217–223, 2013.

MILLS, R. E.; WALTER, K.; STEWART, C.; HANDSAKER, R. E.; CHEN, K.; ALKAN, C.; ABYZOV, A.; YOON, S. C.; YE, K.; CHEETHAM, R. K.; CHINWALLA, A.; CONRAD, D. F.; FU, Y.; GRUBERT, F. HAJIRASOULIHA, I.; HORMOZDIARI, F.; IAKOUCHEVA, L. M.; IQBAL, Z.; KANG, S.; KIDD, J. M.; KONKEL, M. K.; KORN, J.; KHURANA, E.; KURA, D.; LAM, H. Y. K.; LENG, J.; LI, R.; LI, Y.; LIN, C. Y.; LUO, R.; MU, X. J.; NEMESH, J.; PECKHAM, H. E.; RAUSCH, T.; SCALLY, A.; SHI, X.

STROMBERG, M. P.; STÜTZ, A. M.; URBAN, A. E.; WALKER, J. A.; WU, J.; ZHANG, Y.; ZHANG, Z. D.; BATZER, M. A.; DING, L.; MARTH, G. T.; MCVEAN, G.; SEBAT, J.; SNYDER, M.; WANG, J.; YE, K.; EICHLER, E. E.; GERSTEIN, M. B.; HURLES, M. E.; LEE, C.; MCCARROLL, S. A.; KORBEL, J. O. Mapping copy number variation by population-scale genome sequencing. **Nature**, v. 470, p. 59–65, 2011

MAKVANDI-NEJAD, S.; HOFFMAN, G. E.; ALLEN, J. J.; CHU, E.; GU, E.; CHANDLER, A. M.; LOREDO, A. I.; BELLONE, R. R.; MEZEY, J. G.; BROOKS, S. A.; SUTTER, N. B. Four loci explain 83% of size variation in the horse. **PLoS ONE**, v.7, p. 1-6, 2012

McCARTHY, M. I.; ABECASIS, G. R.; CARDON, L. R.; GOLDSTEIN, D. B.; LITTLE, J.; IOANNIDIS, J. P. A.; HIRSCHHORN, J. N. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. **NatureReviewsGenetics**, v. 9, p. 356–369, 2008.

NASCIMENTO, E. F.; SANTOS, R.L. **Patologia da reprodução dos animais domésticos**. 3 ed. Rio de Janeiro: Guanabara Koogan, p. 153. 2011.

PEREIRA, ANIRENE G. T.; UTSUNOMIYA, YURI T.; MILANESI, MARCO; TORRECILHA, RAFAELA B. P.; CARMO, ADRIANA S.; NEVES, HAROLDO H. R.; CARVALHEIRO, ROBERTO; AJMONE-MARSAN, PAOLO; SONSTEGARD, TAD S.; SOLKNER, JOHANN; CONTRERAS-CASTILLO, CARMEN. ; GARCIA, JOSE F. Pleiotropic Genes Affecting Carcass Traits in Bos indicus (Nellore) Cattle Are Modulators of Growth. **PLoS One**, v. 11, n. 7, 2016.

PORTO-NETO, L.R., REVERTER, A., PRAYAGA, K. C., CHAN, E. K. F., DAVID, J., JOHNSTON, D. J., HAWKEN, J. R., FORDYCE, G., GARCIA, J. F., SONSTEGARD, T. S, BOLORMAA, S., GODDARD, M. E., BURROW, H, M., HENSHALL, J. M., LEHNERT, A. S., BARENDSE, W. The Genetic Architecture of Climatic Adaptation of Tropical Cattle. **PLOS ONE**, 9(11), 2014.

RABELO, R. E.; SILVA, L. A. F.; VIU, M. A. O.; ROMANI, A. F.; ALVES, C. B.; FERNANDES, J. J. R.; CASTRO, C. F. P. Acrobustite bovina: revisão de literatura. **Revista CFMV – Suplemento Técnico**, Brasília, Ano XII- n. 37, p. 29-36, 2006

RABELO, R.E.; SILVA, L.A.F.; BRITO, L.A.B.; MOURA, M.I; SILVA, O..C.; CARVALHO, V.S.; FRANCO, L.G. Epidemiological aspects of surgical diseases of

the genital tract in a population of 12,320 breeding bulls (1982-2007) in the state of Goiás, Brazil. **Ciência Animal Brasileira**, Goiânia, v.9, n.3, p.705-713, 2008.

RABELO, R. E.; SCALLA, V.; VALCINIR, A.; CARDOSO, L. D.; DUTRAS, H. T.; HELRIGEL, P. A.; VINCENTIN, F. R. Aspectos Anatômicos E Sua Relação Com As Enfermidades Do Prepúcio E Pênis No Touro Anatomical Aspects and Its Relation With the Disease of the Penis and Fore skin of the Bull, **Revista Científica de Medicina Veterinária**, v. 18, 2012

RICHTER, A.; HAUSCHILD, G.; MURUA ESCOBAR, H.; NOLTE, I.; BULLERDIEK, J.; Application of high mobility-group-A proteins increases the proliferative activity of chondrocytes in vitro. **Tissue Eng Part A**. v.15 p.473-477. 2009.

ROSENZWEIG, S. A.; ATREYA, H. S. Defining the pathway to insulin-like growth factor system targeting in cancer. **Biochemical Pharmacology**, v. 80, p. 1115-1124, 2010.

ROVELET-LECRUX, A., HANNEQUIN, D., RAUX, G., et al. APP locus duplication causes autosomal dominant early-onset alzheimer disease with cerebral amyloid angiopathy. **Nature Genetics**, v.38, p.24–26, 2006.

SANTANA, M. H. A.; UTSUNOMIYA, Y. T.; NEVES, H. H. R.; GOMES, R. C.; GARCIA, J. F.; FUKUMASU, H.; SILVA, S. L.; LEME, P. R.; COUTINHO, L. L.; ELER, J. P.; FERRAZ J. B. Genome-wide association study for feedlot average daily gain in Nelore cattle (*Bos indicus*). **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 210-216, (2014).

SEBAT, J.; LAKSHMI, B.; TROGE, J.; ALEXANDER, J.; YOUNG, J.; LUNDIN, P.; MANER, S.; MASSA, H.; WALKER, M.; CHI, M.; NAVIN, N.; LUCITO, R.; HEALY, J.; HICKS, J. Y. E. K.; REINER, A.; GILLIAM, T. C.; TRASK, B.; PATTERSON, N.; ZETTERBERG, A.; WIGLER, M. Large-scale copy number polymorphism in the human genome. **Science**. v.305, p.525–528. 2004.

SEBAT, J., LAKSHMI, B., MALHOTRA, D., TROGE, J., LESE-MARTIN, C., WALSH, T.; YAMROM, B.; YOON, S.; KRASNITZ, A.; KENDALL, J.; LEOTTA, A.; PAI, D.; ZHANG, R.; LEE, Y. H.; HICKS, J.; SPENCE, S. J.; LEE, A. T.; PUURA, K.; LEHTIMÄKI, T.; LEDBETTER, D.; GREGERSEN, P. K.; BREGMAN, J.; SUTCLIFFE, J. S. JOBANPUTRA, V.; CHUNG, W.; WARBURTON, D.; KING, M. C.; SKUSE, D.; GESCHWIND, D. H.; CONRAD, T. G.; YE, K.; WIGLER, M. Strong Association of De Novo Copy Number Mutations with Autism. **Science**. v.316, p.445–449. 2007

SIMON-SANCHEZ, J.; SCHOLZ, S.; MATARIN, M.D.E.L.M.; FUNG, H.C.; HERNANDEZ, D.; GIBBS, J.R.; BRITTON, A.; HARDY, J.; SINGLETON, A. Genome wide SNP assay reveals mutations underlying Parkinson disease. **Human Mutations**, v.29, p.315–322, 2008.

SCOTT, L. J.; MOHLKE, K. L.; BONNYCASTLE, L. L.; WILLER, C. J.; LI, Y.; DUREN, W. L. et al., A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. **Science**. v. 316, p. 1341–1345, 2007.

SHELL, S.; PARK, S.-M.; RADJABI, A. R.; SCHICKEL, R.; KISTNER, E. O.; JEWELL, D. A. FEIG, C.; LENGYEL, E.; PETER, M. E. Let-7 expression defines two differentiation stages of cancer. **Proceedings of the National Academy of Sciences**, v. 104, p. 11400-11405, 2007.

SKINNER, B. M.; ROBERTSON, L. B.; TEMPEST, H. G.; LANGLEY, I. D.; FOWLER, K. E.; CROOIJMANS, R. P.; HALL, A. D.; GRIFFIN, D. K.; VÖLKER, M. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis, **BMC Genomics** v.10. 2009.

STRANGER, B. E.; FORREST, M. S.; DUNNING, M. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. **Science**, v.315, n.5813, p.848-53, 2007.

TAKAHASHI, L.S., BILLER, J.D. & TAKAHASHI, K.M., **Bioclimatologia zootécnica**, 1 ed. Jaboticabal, SP. p. 40-47. 2009.

UTSUNOMIYA, YURI T.; DO CARMO, ADRIANA S.; CARVALHEIRO, ROBERTO; NEVES, HAROLDO H. R.; MATOS, MARCIA C.; ZAVAREZ, LUDMILLA B.; O'BRIEN, ANA M. PEREZ; SOELKNER, JOHANN; MCEWAN, JOHN C.; COLE, JOHN B.; et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC GENETICS**, v. 14, 2013.

VIU, M. A. O.; TONHATI, H.; CERÓN-MUNÓZ, M. F.; FRIEZ, L. A.; TEIXEIRA, R. A. Parâmetros genéticos do peso e escores visuais de prepúcio e umbigo em gado de corte./ Genetic parameters for weight and visual score of sheat and navel in beef cattle. **Ars Veterinária**, Jaboticabal, SP, Vol. 18, nº 2, 179-184, 2002.

WANG, X.; NAHASHON, S.; FEASTER, T. K.; BOHANNON-STEWART, A.; ADEFOPE, N. An initial map of chromosomal segmental copy number variations in the chicken. **BMC Genomics**, v.11, p.351, 2010.

WANG, J.; WANG, H.; JIANG, J.; KANG, H.; FENG, X.; ZHANG, Q.; LIU, J. Identification of genome – wide copy number variations among diverse pig breeds using SNP genotyping arrays. **Plos one**, v.8, e68683, 2013.

WEEDON, M.N.; LETTRE, G.; FREATHY, R.M.; LINDGREN, C.M.; VOIGHT, B.F.; PERRY, J.R.; ELLIOTT, K.S.; HACKETT, R.; GUIDUCCI, C.; SHIELDS, B.; ZEGGINI, E.; LANGO, H.; LYSSENKO, V.; TIMPSON, N.J.; BURTT, N.P.; RAYNER, N.W.; SAXENA, R.; ARDLIE, K.; TOBIAS, J.H.; NESS, A.R.; RING, S.M.; PALMER, C.N.; MORRIS, A.D.; PELTONEN, L.; SALOMAA, V.; DIABETES GENETICS INITIATIVE; WELLCOME TRUST CASE CONTROL CONSORTIUM, DAVEY, S.G.; GROOP, L.C.; HATTERSLEY, A.T.; MCCARTHY, M.I.; HIRSCHHORN, J.N.; FRAYLING, T.M.A common variant of HMGA2 is associated with adult and childhood height in the general population. **Nat Genet** v.39, p.1245–1250, 2007.

WOODWARK, C.; BATEMAN, A.The characterisation of three types of genes that overlie copy number variable regions. **PLoS One**, v.6, n.5, p.e14814, 2011.

YANG, T.L.; GUO, Y.; ZHANG, L.-S.; TIAN, Q.; YAN, H.; GUO, Y.F.; DENG, H.W. HMGA2 Is Confirmed To Be Associated with Human Adult Height. **Annals of Human Genetics**; v.74(1), p. 11–16, 2010.

ZHANG, F.; GU, W.; HURLES, M. E.; LUPSKI, J. R. Copy Number Variation in Human Health, Disease, and Evolution. Rev. **Genomics Hum. Genet**, v.10, p. 451–481, 2009.

ZHOU, X.; BENSON, K. F.; ASHAR, H. R.; CHADA, K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. **Nature**.v.376. p.771- 774. 1995.

CAPÍTULO 2: Association of a copy number variant with navel in *Bos indicus* reinforces the relevance of the HMGA2-PLAG1-IGF2 pathway in cattle breeding

Tamíris Sayuri Aguiar^{1,2}, Rafaela Beatriz Pintor Torrecilha^{1,2}, Marco Milanesi^{2,3,4}, Adam Taiti Harth Utsunomiya^{2,3}, Beatriz Batista Trigo^{2,3}, Abdulfatai Tijjani⁵, Hassan Hussein Musa⁶, Flávia Lombardi Lopes³, Paolo Ajmone-Marsan⁴, Roberto Carneiro⁷, Haroldo Henrique de Rezende Neves⁸, Adriana Santana do Carmo⁹, Olivier Hanotte^{5,10}, Tad Stewart Sonstegard¹¹, José Fernando Garcia^{1,2,3*}, Yuri Tani Utsunomiya^{2,3*}

¹São Paulo State University (Unesp). School of Agricultural and Veterinarian Sciences, Jaboticabal. Department of Preventive Veterinary Medicine and Animal Reproduction. São Paulo, Brazil;

²International Atomic Energy Agency (IAEA) Collaborating Centre on Animal Genomics and Bioinformatics, Araçatuba, São Paulo, Brazil;

³São Paulo State University (Unesp). School of Veterinary Medicine, Araçatuba. Department of Support, Production and Animal Health. São Paulo, Brazil;

⁴Department of Animal Science Food and Nutrition and Biodiversity and Ancient DNA Research Center –BioDNA, Università Cattolica del Sacro Cuore, Piacenza, Italy;

⁵Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom;

⁶Institute of Molecular Biology, University of Nyala, Sudan;

⁷São Paulo State University (Unesp). School of Agricultural and Veterinarian Sciences, Jaboticabal. Department of Animal Science. São Paulo, Brazil;

⁸GenSys Consultores Associados, Porto Alegre, Brazil;

⁹Escola de Veterinária e Zootecnia (EVZ), Universidade Federal de Goiás - UFG, Campus Samambaia, Goiânia, Goiás, Brazil;

¹⁰LiveGene - CTLGH, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia;

¹¹Recombinetics, Inc., St Paul, MN 55104.

* Correspondence:

Yuri TaniUtsunomiya

ytutsunomiya@gmail.com

Keywords: bovine, SNP, haplotype, CNV, *HMGA2*, *PLAG1*, *IGF2*

Abstract

Navel injuries caused by friction against the pasture can promote infection, reproductive problems and costly treatments in beef cattle raised in extensive systems. A haplotype-based genome-wide association study (GWAS) was performed for visual scores of navel length at yearling in Nellore cattle (*Bos indicus*) using data from 1298 animals and 503088 single nucleotide polymorphism (SNP) markers. The strongest signal ($p = 1.01 \times 10^{-9}$) was found on chromosome 5 spanning positions 47.9 - 48.2 Mbp. This region contains introns 3 and 4 and exons 4 and 5 of the high mobility group AT-hook 2 gene (*HMGA2*). Further inspection of the region with whole genome sequence data of 21 animals revealed correlations between counts of the significant haplotype and copy number gains of a ~6.2 kbp segment of intron 3 of *HMGA2*. Analysis of genome sequences from five African *B. indicus* and four European *Bos taurus* breeds revealed that the copy number variant (CNV) is indicine-specific. This intronic CNV was then validated through quantitative polymerase chain reaction (qPCR) using Angus animals as copy neutral controls. Importantly, the CNV was not detectable by means of conventional SNP-based GWAS or SNP probe intensity analyses. Given that *HMGA2* affects the expression of the insulin-like growth factor 2 gene (*IGF2*) together with the pleomorphic adenoma gene 1 (*PLAG1*), and that the latter has been repeatedly shown to be associated with

quantitative traits of economic importance, these findings highlight the relevance of the *HMGA2-PLAG1-IGF2* pathway to cattle breeding.

Introduction

The insulin-like growth factor (IGF) or somatomedin pathway is a complex system that controls the signaling cascades of the PI3K/Akt, MAPK and Ras/Raf/MEK/ERK pathways, resulting in cell growth and proliferation and inhibition of apoptosis (Rosenzweig and Atreya, 2010; King and Wong, 2012). The main effectors of the IGF system, namely *IGF1* and *IGF2*, are preferentially expressed after and before birth, respectively (Bergman et al., 2013). In particular, *IGF2* is maternally imprinted, and disrupting mutations inherited from the male germline cause growth deficiency in mice (DeChiara et al., 1991). Moreover, a mouse knockout model of the pleomorphic adenoma gene 1 (*PLAG1*), a transcription factor for *IGF2*, also presented marked growth retardation and decreased fertility (Hensen et al., 2004). Indeed, a naturally occurring *PLAG1* mutation has been shown to affect a wide range of economically relevant traits in both *Bostaurus* (Karim et al., 2011; Saatchi et al., 2014; Bolormaa et al., 2014) *Bosindicus* (Utsunomiya et al., 2013; Fortes et al., 2013; Pereira et al., 2016) cattle, implicating a major contribution of the IGF system to genetic variance of several traits in the bovine species.

Another transcription factor affecting expression of *IGF2* is the high mobility AT-hook 2 (*HMGA2*). Regulation of *IGF2* by *HMGA2* has been recently demonstrated to occur directly or through increased expression of *PLAG1* (Klemke et al., 2014; Abi Habib et al., 2017) (Figure 1). In fact, *HMGA2* has been found to harbor variants associated with height in humans (Weedon et al., 2007; Yang et al., 2010), horses (Frischknecht et al., 2015) and dogs (Hayward et al., 2016), as well as ear size in pigs (Li et al., 2012). The bovine *HMGA2* orthologue is located on chromosome (CHR) 5 between base pairs (bp) 48053846 and 48199963 in the UMD v3.1.1 assembly (Zimin et al., 2009). Quantitative trait loci (QTL) encompassing this genomic region have been reported in cattle for growth and fatness (Bolormaa et al., 2014), and body condition score and penile sheath (Porto-Neto et al., 2014). However, putative causal variants underlying these QTL remain elusive.

Here we report a haplotype-based genome-wide association study (GWAS) for navel at yearling (NY) in Nellore cattle (*Bos indicus*), which resulted in the discovery of a copy number variation (CNV) in intron 3 of *HMGA2* that was not detectable from conventional GWAS or SNP probe intensity analyses. Moreover, validation of the CNV through quantitative polymerase chain reaction (qPCR) showed that the genome-wide significant haplotype performed well as a predictor of CNV genotype in our samples. Finally, analysis of a panel of whole genome sequences of African *B. indicus* and European *B. taurus* breeds revealed the CNV to be of *B. indicus* origin.

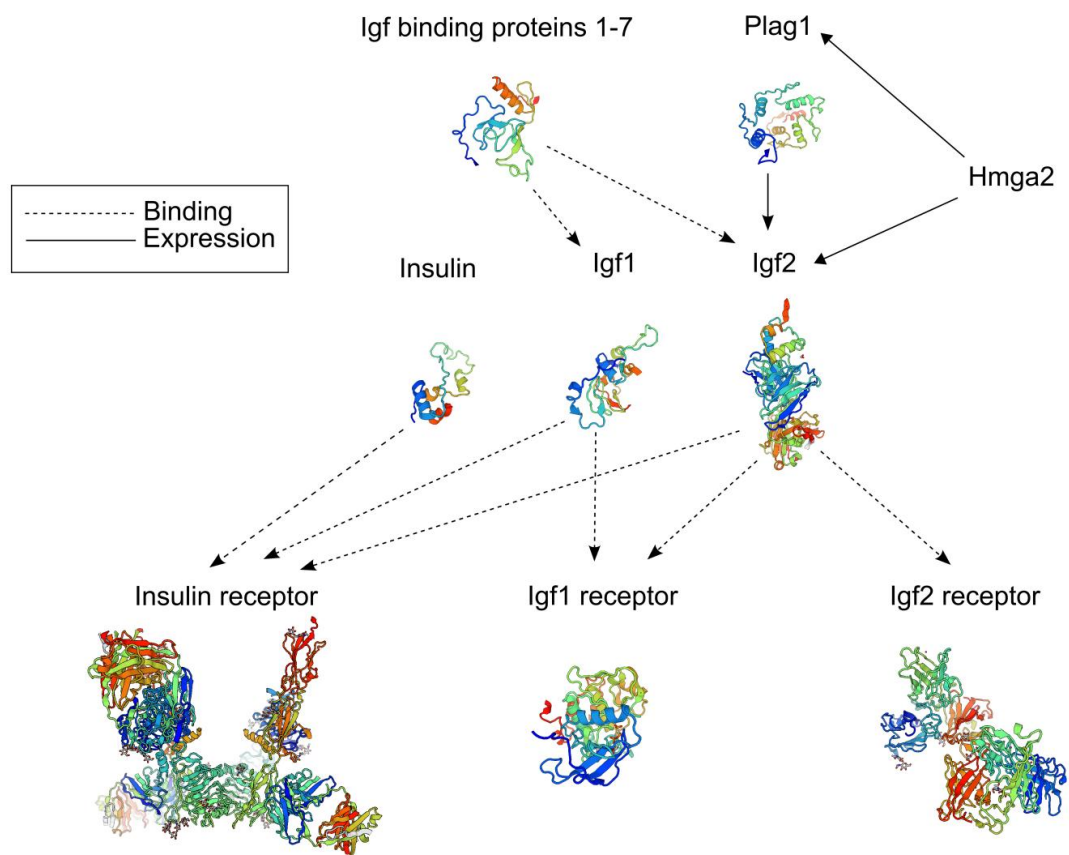


Figure 1. *HMGA2-PLAG1-IGF2* pathway scheme. Protein tertiary structures displayed in this figure were built using SWISS-MODEL (Bienert et al., 2017).

Material and Methods

Genotypes

Illumina BovineHDBeadChip assay (786799 SNPs) genotypes of 953 Nellore bulls and 1278 Nellore cows were available for analysis from previous studies (Carvalho et al., 2014; Zavarez et al., 2015). These animals comprised part of the genomic selection reference population from a commercial breeding program that routinely performs genetic evaluations for weight, carcass and reproductive traits. All genotyped samples had a minimum genotyping rate of 90%. Autosomal markers with unique genomic coordinates were filtered with PLINK v1.90b4.6 (Purcell et al., 2007; Chang et al., 2015) for a minimum call rate of 95%, GenTrain Score of at least 70% and minor allele frequency of at least 2%. The filtered genotypes were then phased with the Segmented HAPlotype Estimation & Imputation Tool (SHAPEIT2) v2.r837 (O'Connell et al., 2014). Phasing was performed with a burn in of 10 iterations, pruning of 10 iterations, 50 main iterations, 200 states, windows of 500 kbp and effective population size of 113. The latter parameter was estimated from genotype the data with SNeP v1.1 (Barbato et al., 2015).

Phenotypes

Pseudo-phenotypes were based on estimated breeding values (EBVs) obtained from an animal model fitted to records of 745,466 animals. Navel at yearling was recorded based on visual evaluation of animals by trained technicians. The navel scores were assigned considering an absolute scale, ranging from 1 to 5, so that the largest scores were attributed to animals with longer or more pendulous navels. For males, the penile sheath was considered part of the navel. Animals pertaining to the same contemporary group (i.e., animals from the same herd, sex, birth year, birth season and management group) were evaluated by the same technicians, so that the fixed effect of contemporary group fitted in the animal model is expected to account for any systematic effects associated to technicians. Prior to GWAS, EBVs were deregressed following Garrick et al., (2009).

Genome-wide association analysis

The GWAS model used here was essentially the leave-one-chromosome-out (LOCO) procedure proposed by Yang et al., (2014), except that haplotypes were tested instead of single SNPs. Model fitting was performed in GCTA v.1.90.2 beta

(Yang et al., 2011) by using haplotype data as input via coding of haplotypes as pseudo-SNP markers. More specifically, haplotype data were provided to GCTA in binary (bed/bim/fam) format with genotypes expressed in terms of number of copies of each haplotype (0, 1 or 2). The association analysis for each haplotype was then based on the following mixed linear model:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{x}b + \mathbf{u} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of deregressed EBVs, $\mathbf{1}_n$ is a vector of ones (length equal to n , the number of animals with deregressed EBV), μ is the intercept, $\mathbf{x}$ is a $n \times 1$ vector of standardized haplotype counts, b is the fixed effect of the haplotype, $\mathbf{u}$ is the $n \times 1$ vector of accumulated contributions of random effects of all SNPs except those located in the same chromosome as the haplotype being tested, and $\mathbf{e}$ is the $n \times 1$ vector of random residual effects. It was assumed that $\mathbf{y} \sim \text{MVN}(\mathbf{1}_n\mu + \mathbf{x}b, \mathbf{G}\sigma_u^2 + \mathbf{I}\sigma_e^2)$, where $\mathbf{G}$ is the SNP-based genomic relationship matrix (GRM), σ_u^2 is the marked additive genetic variance, $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance. The model above was fitted in two steps: first, variance components were estimated using the average information restricted maximum likelihood (AI-REML) algorithm; then, estimates of fixed effects were obtained using generalized least squares equations. In order to reduce computational burden, variance components were estimated only once per chromosome using a reduced model without haplotypes. Also, only animals presenting accuracy of deregressed EBV greater than 0.70 were used to reduce influence of heterogeneity of variance in the analysis. By conditioning all analyses on the accumulated effects of SNPs, associations presumably reflected significant haplotype effects not captured by random SNP effects. Moreover, the SNP-based GRM helped controlling the analyses for putative confounding effects of cryptic relationships and population substructure. Haplotypes were constructed with GHap v1.2.2 (Utsunomiya et al., 2016) using overlapping segments of six consecutive markers, considering that the average intermarker distance was ~5 kbp and that linkage disequilibrium extends up to 30 kbp in the Nellore genome (Espigolan et al., 2013). Only haplotypes with frequency between 5% and 95% were tested (analogous to exclusion of loci with minor allele frequency below 5% in conventional GWAS). Additionally, to assess the sensitivity of the GWAS

procedure to the choice of haplotype size, we repeated the analysis with haplotypes of 1 (equivalent to single marker analysis), 5 and 10 SNPs. Haplotypes were prioritized for investigation if their p-values were lower than a stringent Bonferroni corrected significance level of $0.05/N$, where N is the total number of tested haplotypes.

Analysis of whole genome sequence data

Next-generation sequence data of Nellore animals were obtained from a previous study (Utsunomiya et al, 2017) and consisted of Illumina HiSeq 2000 paired-end reads from 21 bulls aligned to the UMD v3.1.1 *Bos taurus* assembly (Zimin et al., 2009). The average sequencing coverage was 9.25x. For the candidate region detected by the GWAS analysis, single nucleotide variants and short insertions/deletions were identified with the mpileup algorithm from SAMtools v1.3.1 (Li, 2009) and annotated with the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Copy number gains were identified by increase of nucleotide coverage as compared to the average sample coverage. First, for each sample and nucleotide, coverage was computed as N_{reads}/C , where N_{reads} is the number of reads encompassing the nucleotide and C is the sample coverage. Then, bulls were grouped according to their number of copies (0, 1 and 2) of the significant haplotype and nucleotide coverage was averaged within groups. A last normalization step was adopted by taking the median coverage of 1 kbp windows. This procedure was intended to smooth out outlying values and increase the signal-to-noise ratio in the data. Candidate variants and alignment data were also inspected with the Integrative Genomics View (IGV) v2.3 software (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). Occurrence of the candidate causal mutation in other cattle breeds was investigated using re-sequencing data ($n = 13$) from five African zebu *B. indicus* cattle breeds (BioProject PRJNA312138), namely Kenana, Butana, Sudanese Baggara, Ogaden and Kenyan Boran (Kim et al. 2017), as well as publicly available whole genome sequence data ($n = 20$) of four *B. taurus* breeds (Angus, Holstein, Jersey and Simmental, BioProject PRJNA238491) (Daetwyler et al., 2014).

SNP probe intensity analysis

Let X and Y be the normalized signal intensity for the A and B allele, respectively, and $R = X + Y$. The normalized total signal intensity (LRR) was calculated as $\log_2 (R_{\text{observed}}/R_{\text{expected}})$, where R_{expected} was computed from linear interpolation of the canonical genotype clusters. Since LRR is prone to genomic “waves”, supposedly produced by differences in GC-content (%GC) and intensified as the sample deviates from optimal DNA quantity (Diskin et al., 2008), LRR values were further corrected for waviness to avoid false positive copy number variation. Adjustment was achieved by regressing LRR values onto % GC computed over 1 Mbp windows encompassing each marker and taking residuals as the new waviness-normalized LRR values. Calculation of %GC from a FASTA file of the UMD v3.1.1 assembly was performed with the nuc program in BEDTools v2.26.0 (Quinlan, 2014). Values of LRR were then averaged across haplotype classes 0, 1 and 2 and compared with coverage data obtained from whole genome sequences.

qPCR

A qPCR assay was set up in order to validate a CNV that was identified in the whole genome sequence analysis. The pair of primers 5'TCAAAGCCAACTGATTGCTG 3' (forward) and 5'TTTCCTATGGTCCCAAGCG 3' (reverse) was designed considering the limits of the CNV region using the NCBI Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast>). Primers 5' GGTGCCTTGTGGGAGATGTA 3' (forward) and 5' GTGTCAGCTGCCTGAGTCCT 3' (reverse) targeting the tumor suppressing subtransferable candidate 4 gene (TSSC4) were used as a single copy reference. The assay was performed using a 96-samples plate. We selected for qPCR analysis available samples from our DNA bank from 4, 4 and 3 animals with 0, 1 and 2 copies of the significant haplotype, respectively. Additionally, 4 Angus animals were used as copy neutral controls. All samples were assessed in triplicate and no-DNA templates were included for each pair of primers as a control. Each assay consisted of a reaction with a total volume of 20 μ l, which contained 1 μ l of genomic DNA, 0.8 μ l of each primer (10 μ M), 7.4 μ l of pure water and 10 μ l of QuantiTect® SYBER® Green RT-PCR Master Mix (QIAGEN®). Relative increase of copy number was calculated for each sample using the $2^{-\Delta\Delta C_t}$ method

(Livak and Schmittgen, 2001). The $2^{-\Delta\Delta C_t}$ calculation considered as the calibrator sample one Angus animal.

Results

GWAS identifies a *HMGA2* haplotype associated with navel at yearling in Nellore cattle.

The initial data filtering reduced the SNP set to 503088 markers, which were retained for phasing. Haplotyping of six-markers segments was carried out in all 2231 samples, but the GWAS analysis was conducted exclusively on data from 2016 animals whose accuracy of deregressed EBV was at least 0.70. Association results pointed to a single leading QTL region on chromosome 5 mapping between 47.9 and 48.2 Mbp (Figure 2). Four haplotypes from consecutive segments spanning positions 47921750-48069099 bp were tied as the most significant haplotype ($p = 1.01 \times 10^{-9}$), which formed the consensus allele TCCTCCAAC. This consensus haplotype overlapped introns 3 and 4, exons 4 and 5, and the 5' end of *HMGA2*.

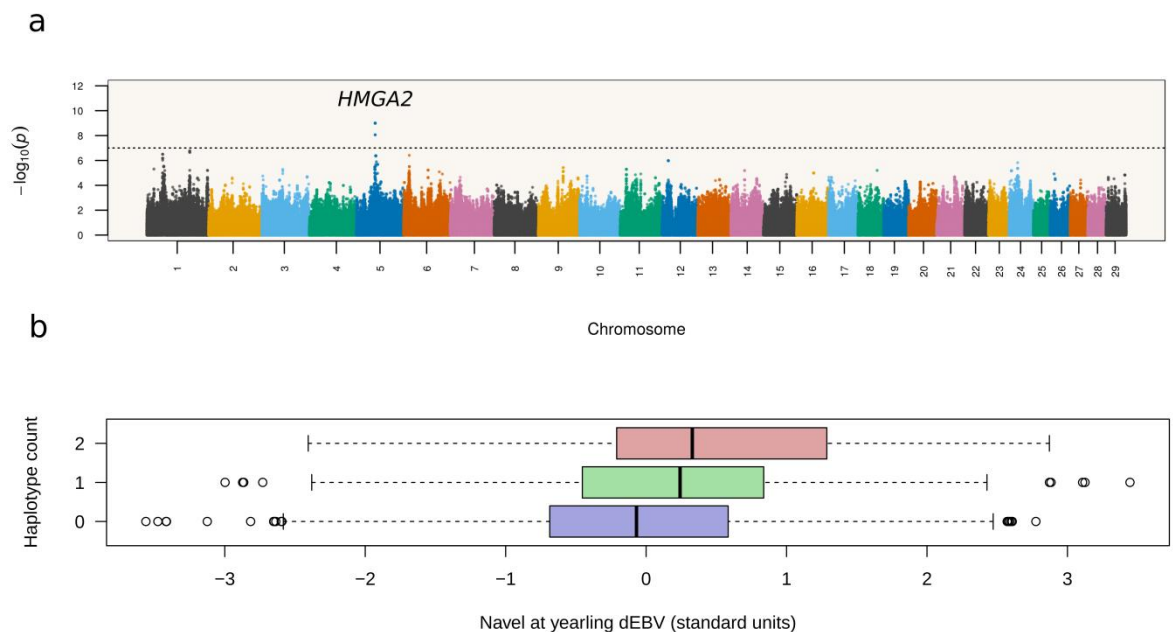


Figure 2. Haplotype-based GWAS maps navel at yearling associations to *HMGA2*. **(a)** Each point in the Manhattan plot corresponds to a 6-markers long haplotype. The dashed horizontal line corresponds to the Bonferroni threshold ($p <$

2.44×10^{-8}). **(b)** Distributions of dEBVs according to number of copies of the leading haplotype. The *HMGA2* locus accounted for 2.04% of the variance in dEBVs.

Nellore sequence data suggests a CNV intronic to *HMGA2* as the causal variant

Assuming that the TCCTCCAAC haplotype tagged an unobserved causal mutation, we used TCCTCCAAC copy number to predict causal mutation genotypes in a sample of 21 Nellore bulls with whole genome sequence data. From a total of 1185 sequence variants detected by SAMtools within the chromosome 5 range 47.9 – 48.2 Mbp, only 6 presented genotype correspondence with haplotype counts, namely: rs209832737 (CHR5:47962011, C>G), rs521422509 (CHR5:47964144, A>G), rs525140201 (CHR5:47965249, G>A), rs521794670 (CHR5:47966055, G>A), rs519732918 (CHR5:47966375, T>C) and rs516281664 (CHR5:47967920, C>T). However, none of these variants were novel and all of them were intergenic. Analysis of coverage data revealed an additional variant comprising a ~6.2 kbp CNV spanning positions 48074233-48080443 (Figure 3a). This CNV was located on intron 3 of *HMGA2*, and TCCTCCAAC haplotype counts 0, 1 and 2 translated into approximate fold changes of 4.0, 4.6 and 5.2 in sequence coverage across the ~6.2 kbp segment in the Nellore samples. Importantly, all sequenced animals presented copy gains at the CNV region (hereafter denoted *HMGA2*-CNVR), regardless of haplotype count, suggesting a baseline copy gain instead of a copy neutral sequence background, and co-segregation of the relevant haplotype with additional copies of the ~6.2 kbp DNA segment.

Genome sequences from nine cattle breeds indicate a *B. indicus* origin for the *HMGA2*-CNVR

In order to determine whether the *HMGA2*-CNVR was a recent derived mutation private to Nellore cattle or an ancient variant, we inspected the candidate chromosomal region in sequence data from five African *B. indicus* and four European *B. taurus* breeds (Figure 3b). Four out of five *B. indicus* breeds presented increased coverage compatible with the *HMGA2*-CNVR previously found in Nellore cattle,

whereas all *B. taurus* breeds were copy neutral at the relevant positions on chromosome 5. These results strongly point to an indicine-specific CNV.

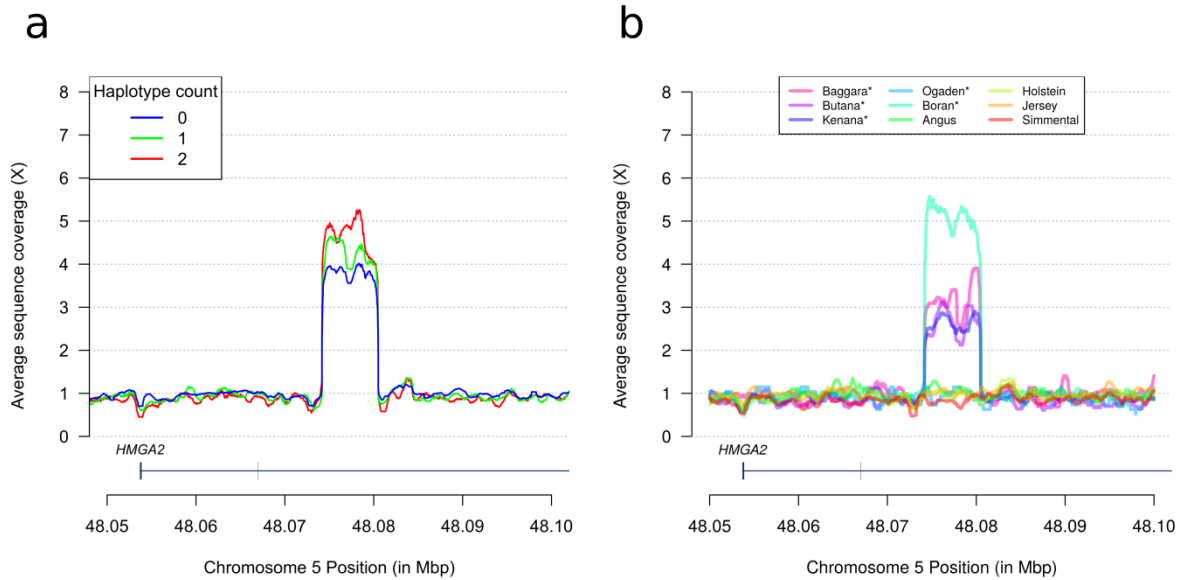


Figure 3. Discovery of a *B. indicus* specific CNV on *HMGA2* affecting navel at yearling. **(a)** Relative fold increase in sequence coverage of a segment of *HMGA2* intron 3 correlates with TCCTCCAAC haplotype counts in Nellore cattle. Each smoothed curve corresponds to sequence coverage averaged across samples with same haplotype count. **(b)** Inspection of the CNVR in additional European *B. taurus* and African *B. indicus* breeds reveals specificity of copy gains in *B. indicus* (marked with *).

qPCR validates haplotype-based prediction of CNV genotypes

The TCCTCCAAC haplotype tag suggested that the frequency of additional copy gains at the *HMGA2*-CNVR in our Nellore sample was 16.5%. The numbers of animals carrying 0, 1 and 2 copies of the tag haplotype were 1554, 619 and 58, respectively. Based on available samples in our DNA bank, we randomly chose 4 animals from each group for validation of the *HMGA2*-CNVR with qPCR. An exception was the group of TCCTCCAAC-homozygote animals, which included 3 samples. We also used DNA from 4 Angus animals as copy neutral controls, assuming that *B. taurus* animals were copy neutral. The average $2^{-\Delta\Delta C_t}$ values for carriers of 0, 1 and 2 copies of the tag haplotype in the Nellore breed were

respectively 7.04 ± 1.57 , 8.79 ± 1.93 and 9.02 ± 2.79 , whereas Angus animals presented an average of 1.39 ± 0.62 . The qPCR results were therefore in agreement with the existence of a CNVR located in intron 3 of *HMGA2* in Nellore cattle, for which relative copy gain could be predicted by the TCCTCCAAC haplotype.

Conventional GWAS and SNP probe intensity data fail to detect the CNV

In order to evaluate sensitivity of GWAS results to the choice of haplotype size, we repeated the genome-wide scan analysis with single SNPs and haplotypes recovered from segments of 5 and 10 markers (Figure 4). Although the topology of the chromosome 5 QTL was fairly preserved across all analyses, the single SNP scan did not present enough power to declare the *HMGA2* region as significant considering a Bonferroni-corrected significance level. This suggests that haplotypes performed better than single SNPs as tags for the putative causal mutation in this particular study. Therefore, a conventional single-marker GWAS analysis using Bonferroni correction would have missed the *HMGA2* association reported here. Indeed, associations were even stronger when the haplotype size was increased to 10 markers. Regarding detection of the *HMGA2*-CNVR via probe intensity data, a previous study using the same Nellore genotypes (Zhou et al., 2016) did not detect this CNV with either the CNAM optimal segmentation method from the Golden Helix SVS v8.3.0 software (Golden Helix Inc., Bozeman, MT, USA) or the Hidden Markov Model (HMM) implemented in PennCNV (Wang et al., 2007). Therefore, we decided to directly inspect LRR values from SNP probes to evaluate whether intensity data could capture information from the CNV. As shown in Figure 5, only two SNP markers mapped to *HMGA2*-CNVR, indicating that the local SNP density in the BovineHD assay was insufficient to reveal the presence of the CNV with confidence. Moreover, only a slight increase in LRR values could be observed for these two SNPs as compared to their neighboring markers, which was probably not high enough to be detectable with either SVS or PennCNV.

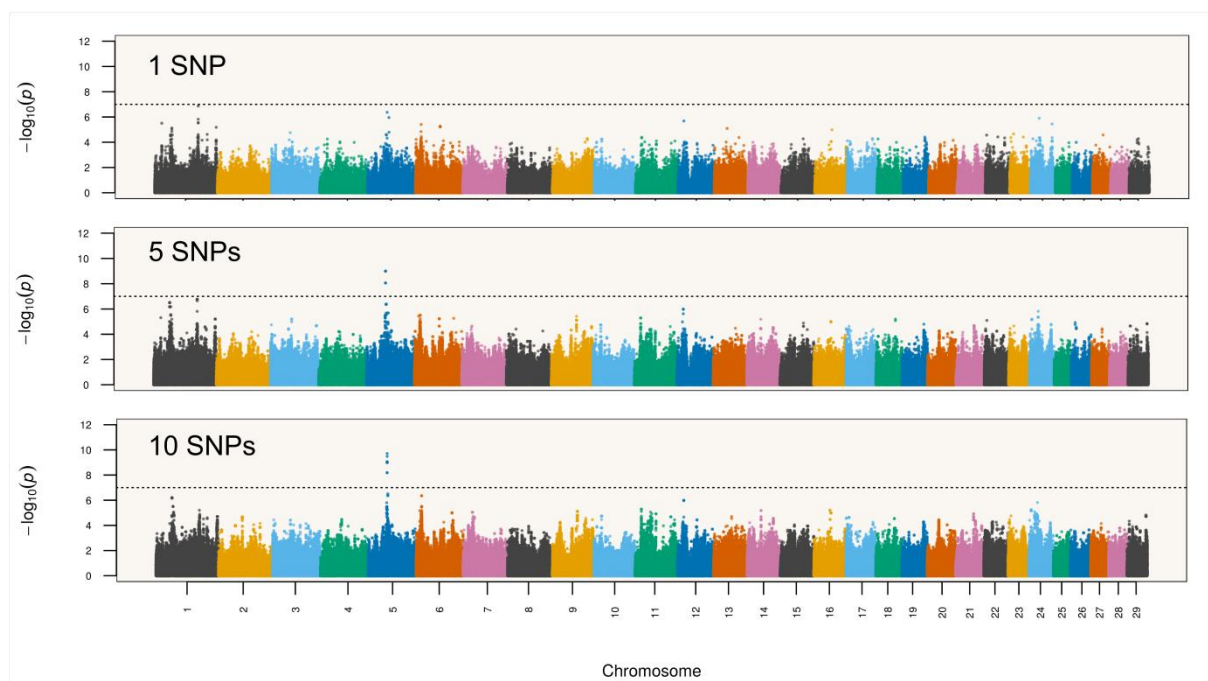


Figure 4. Navel at yearling associations with varying haplotype sizes (1, 5 and 10 SNPs).

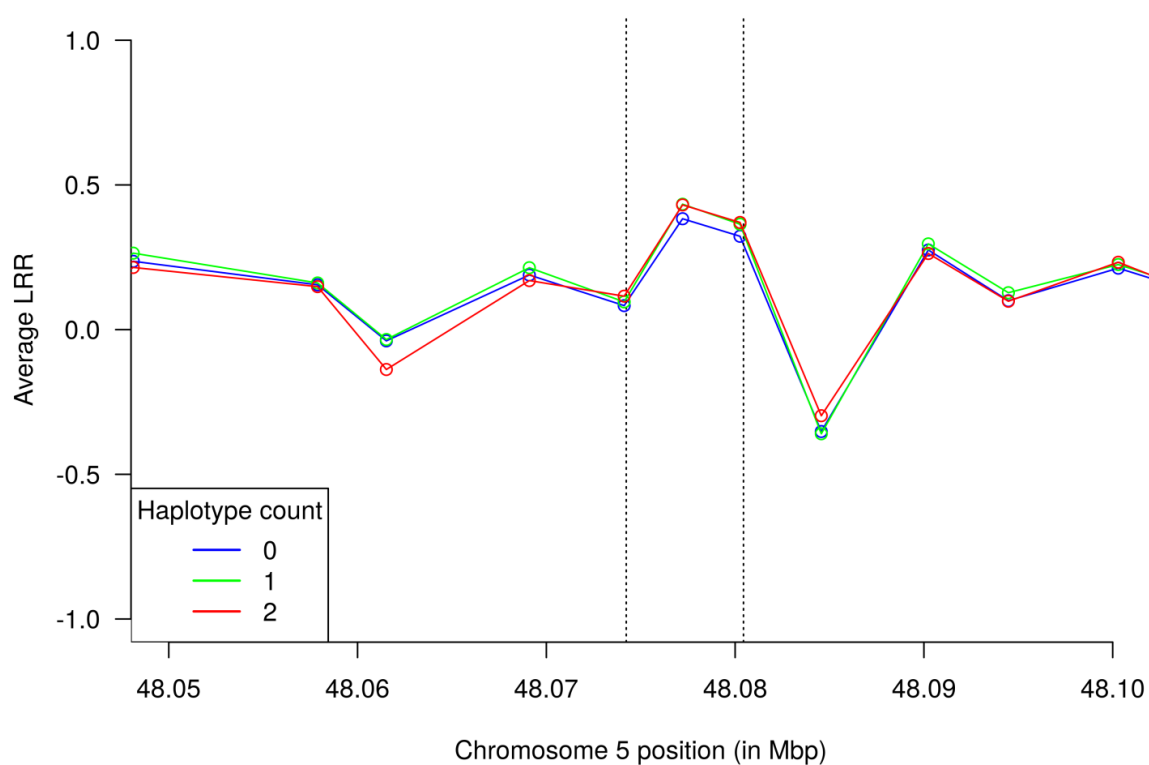


Figure 5. Average probe intensity data from the BovineHD assay for the *HMGA2*-CNVR segment (vertical dashed lines). Each curve corresponds to average LRR values of samples with the same TCCTCCAAC haplotype count. Points in the curves indicate positions of SNP markers.

Discussion

Navel at yearling is an economically important trait in beef cattle raised in extensive production systems. A pendulous navel increases the risk of injuries caused by friction against the pasture, leading to infection, reproductive impairment and treatment expenses (Ashdown, 2006; Rabelo et al., 2008, Boligon, 2016). This trait is especially important to cattle raised in Brazil, the second largest beef exporter of the world (ABIEC, 2015), where Nellore animals compose the majority of the herds and approximately 88% of the females in reproductive age are subjected to natural mating (ASBIA, 2016). Here, we reported the identification of a CNV located on intron 3 of *HMGA2* associated with NY, which was detected through a combination of haplotype-based GWAS and whole genome sequence analysis. We also showed that this putative causal mutation was difficult to detect via conventional GWAS and CNV detection methods. Moreover, we found that other *B. indicus* breeds apart from Nellore cattle also carry copy gains of this *HMGA2*intronic segment.

Interactions among *HMGA2*, *PLAG1* and *IGF2*, and their numerous pleiotropic effects on traits related to growth and reproduction, have only recently started to emerge (Klemke et al., 2014; Habib et al., 2017). However, the importance of *HMGA2* as an oncogene, and its regulation by a series of micro RNAs (miRNA) binding to its 3'-UTR region, has been known for at least a decade. For instance, one of the first characterized miRNAs, let-7, is a major negative regulator of *HMGA2* and explains the suppression of this gene in later stages of development (Lee and Dutta, 2007). Another relevant regulatory miRNA is miR-763, which is encoded by intron 3 of *HMGA2* and possibly co-expressed with its host gene (Von Ahsen et al., 2008; Artzi et al., 2008). Intron 3 of *HMGA2* is also a frequent target of structural and chromosomal abnormalities in human tumors. The *HMGA2*-CNVR identified in the present study occurs on the same intron. The orthologous sequence in humans is predicted to generate an anti-sense, long non-coding RNA, which is yet to be

validated and observed in the bovine species. Nevertheless, given the major regulatory role of this intron on *HMGA2* transcript abundance, we propose an effect of variation of copy number at this CNVR on *HMGA2* expression and navel size in Nellore cattle. This is also supported by associations between penile sheath and SNPs close to *HMGA2* in Brahman (*B. indicus*) and Tropical Composite (*B. indicus* × *B. taurus*) cattle (Porto-Neto et al., 2014).

Apart from the observed effects of the *HMGA2*-CNVR on navel size, the candidate causal variant is potentially involved with other heritable traits considering the pleiotropic nature of *HMGA2*. For instance, *HMGA2* variants have been associated with floppy ears in dogs (Boyko et al., 2010), ear size in pigs (Li et al., 2012), fat deposition in cattle (Bolormaa et al., 2014) and stature in humans (Weedon et al., 2007; Yang et al., 2010), cattle (Bouwman et al., 2018), horses (Frischknecht et al., 2015) and dogs (Hayward et al., 2016). Mice homozygous for a null *HMGA2* allele present a “superpygmy” phenotype, reduced amounts of fat tissue and infertility (Federico et al., 2014). Also, the detection of the *HMGA2*-CNVR in African *B. indicus* cattle is intriguing. The *B. indicus* population in Africa descended from importation of animals from Asia during the Arab invasions between the 7th – 8th centuries (Hanotte et al. 2002), whereas the Brazilian Nellore population derived from imports of Indian animals in the mid 20th century (Ajmone-Marsan et al., 2010). Altogether, these observations suggest that the causal variant is probably a wide spread ancient *B. indicus* mutation, which might have had an adaptive value in this subspecies.

Author Contributions

J.F.G. and Y.T.U. conceived and designed the study. A.S.C. and T.S.S. coordinated genotyping and sequencing of Nellore animals. A.T. and O.H. provided and analyzed African cattle sequence data. R.C. and H.H.R.N. analyzed phenotypes and pedigree data and generated dEBVs. A.T.H.U. performed genotype phasing. T.S.A. and Y.T.U. performed association analyses. M.M., Y.T.U. and P.A.M. designed the qPCR experiment. T.S.A., B.T., R.B.P.T. and F.L.L performed the qPCR assays. T.S.A., R.B.P.T. and M.M. performed Nellore and *B. taurus* sequence data analyses. T.S.A. and Y.T.U. wrote the manuscript. All authors revised and agreed with the contents of the manuscript.

Funding

This research was supported by the Sao Paulo Research Foundation (FAPESP, process 2010/52030-2, 2014/01095-8, 2016/07531-0, 2016/05787-7 and 2017/08373-1) and National Council for Scientific and Technological Development (CNPq, process 560922/2010-8, 483590/2010-0, 132809/2016-8 and 166419/2017-6).

Reference

ABIEC. Associação Brasileira das Indústrias Exportadoras de Carnes. Rebanho Bovino Brasileiro. Disponível em: http://www.abiec.com.br/download/stat_mercadomundial.pdf Acesso em: 23 de mar de 2016.

AJMONE-MARSAN, P., GARCIA, J. F., LENSTRA, J. A., THE GLOBALDIV CONSORTIUM. (2010). On the Origin of Cattle: How Aurochs Became Cattle and Colonized the World. **Evolutionary Anthropology**. 19, 48–157

ARTZI, S., KIEZUN, A., SHOMRON, N. (2008). miRNAMiner: A tool for homologous microRNA gene search. **BMC Bioinformatics**. 9. 39.

ASBIA. Associação Brasileira de Inseminação Artificial. Relatório índice do ano de 2014. Disponível em: <http://www.asbia.org.br/novo/relatorios/> Acesso em: 13 de Ago de 2017.

ASHDOWN, R.R. (2006). Functional, developmental and clinical anatomy of the bovine penis and prepuce. **CAB Reviews: Perspectives in agriculture, Veterinary Science, Nutrition and Natural Resources**. 21, 29-37. doi: 10.1079/PAVSNNR20061021

BARBATO, M., OROZCO-TERWENGEL, P., TAPIO, M., BRUFORD, M.W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome wide SNP data. **Frontiers in Genetics** .6, 109.doi: 10.3389/fgene.2015.00109

BERGMAN, D., HALJE, M., NORDIN, M., ENGSTRÖM, W. (2013). Insulin-Like Growth Factor 2 in Development and Disease: A Mini-Review. **Gerontology**. 59, 240 – 249. doi: 10.1159/000343995

BIENERT, S., WATERHOUSE, A., DE-BEER, T.A., TAURIELLO, G., STUDER, G., BORDOLI, L., et al., (2017). The SWISS-MODEL Repository-new features and functionality. **Nucleic Acids Res.** 45, 313-319. doi: 10.1093/nar/gkw1132

BOUWMAN, A.C., DAETWYLER, H.D., CHAMBERLAIN, A.J., PONCE, C.H., SARGOLZAEI, M., SCHENKEL, F.S., et al., (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. **Nat Genet.** 50, 362-367.

BOLIGON, A.A., DE VARGAS, L., SILVEIRA, D.D., ROSO, V.M.; CAMPOS, G. S.; VAZ, R. Z. et al., (2016). Genetic models for breed quality and navel development scores and its associations with growth traits in beef cattle. **Trop Anim Health Prod.** 48, 1679-1684.

BOLORMAA, S., PRYCE, J. E., REVERTER, A., ZHANG, Y., BARENDSE, W., KEMPER, K., et al., (2014). A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. **PLOS Genetics.** 10.

BOYKO, A. R., QUIGNON, P., LI, L., SCHOENEBECK, J. J., DEGENHARDT, J. D., LOHMUELLER, K. E., et al., (2010) A simple genetic architecture underlies morphological variation in dogs. **PLoS Biol.** 8, 49-50.

CARVALHEIRO, R., BOISON, S. A., NEVES, H. H. R., SARGOLZAEI, M., SCHENKEL, F. S., UTSUNOMIYA, Y. T., et al., (2014). Accuracy of genotype imputation in Nelore cattle. **Genetics Selection Evolution (GSE).** 46-69.

CHANG, C.C., CHOW, C.C., TELLIER, L.C., VATTIKUTI, S., PURCELL, S. M, LEE, J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. (2015) **GigaScience.** 4, 7.

DAETWYLER, H. D., CAPITAN, A., PAUSCH, H., STOTHARD, P., VAN BINSBERGEN, R., BRØNDUM, R. F., et al. (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nat Genet.** 46, 858-865.

DECHIARA, T.M., ROBERTSON, E.J., EFSTRATIADIS, (1991). A Parental imprinting of the mouse insulin-like growth factor II gene. **Cell**. 64, 849–859.

DISKIN, S. J., LI, M., HOU, C., YANG, S., GLESSNER, J., HAKONARSON, H., et al., (2008) Adjustment of Genomic Waves in Signal Intensities from Whole-Genome SNP Genotyping Platforms. **Nucleic Acids Research**. 36.

ESPIGOLAN, R., BALDI, F., BOLIGON, A. A., SOUZA, F. R., GORDO, D. G., TONUSSI, R. L. et al., (2013). Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics**. 14, 305.

FEDERICO, A., FORZATI, F., ESPOSITO, F., ARRA, C., PALMA, G., BARBIERI, A. et al., (2014). Hmga1/Hmga2 double knock-out mice display a "superpygmy" phenotype. **Biol Open**. 11.

FORTES, M.R.S., REVERTER, A., KELLY, M., MCCULLOCH, R., LEHNERT, S.A. (2013). Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. **Andrology**. 1, 644-650.

FRISCHKNECHT, M., JAGANNATHAN, V., PLATTET, P., NEUDITSCHKO, M., SIGNER-HASLER, H., BACHMANN, I., PACHOLEWSKA, A., et al., (2015). A Non-Synonymous HMGA2 Variant Decreases Height in Shetland Ponies and Other Small Horses. **PLOS ONE**. 10.

GARRICK, D.J., TAYLOR, J.F., FERNANDO, R.L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution (GSE)**. 41, 55.

ABI-HABIB, W., BRIOUDE, F., EDOUARD, T., BENNETT, J.T., LIENHARDT-ROUSSIE, A., TIXIER, F., et al., (2018). Genetic disruption of the oncogenic HMGA2–PLAG1–IGF2 pathway causes fetal growth restriction. **Genetics in Medicine**. 20, 250-258.

HANOTTE, O., BRADLEY, D.G., OCHIENG, J., VERJEE, Y., HILL, E.W. AND REGE J.E.O. (2002). African pastoralism: genetic imprints of origins and migrations. **Science**. 296, 336-339.

HAYWARD, J. J., CASTELHANO, M.G., OLIVEIRA, K.C., COREY, E., BALKMAN, C., BAXTER, T.L., et al., (2016). Complex disease and phenotype mapping in the domestic dog. **Nature. Commun.** 7, 1–11.

HENSEN, K., BRAEM, C., DECLERCQ, J., VAN DYCK, F., DEWERCHIN, M., FIETTE, L., et al., (2004). Targeted disruption of the murine *Plag1* proto-oncogene causes growth retardation and reduced fertility. **Development Growth and Differentiation**. 46, 459-470.

KARIM, L., TAKEDA, H., LIN, L., DRUET, T., ARIAS, J.A.C., BAURAIN, D., et al., (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. **Nature Genetics**. 43, 405-413.

KIM, J., HANOTTE, O., MWAI, O.A., DESSIE, T., BASHIR, S., DIALLO, B., et al., (2017). The genome landscape of indigenous African cattle. **Genome Biology**. 18, 34.

KING, E. R., WONG, K. (2012). Insulin-like Growth Factor: Current Concepts and New Developments in Cancer Therapy. Recent Patents on Anti-Cancer **Drug Discovery**. 7, 14-30.

KLEMKE, K., MÜLLER, M.H., WOSNIOK, W., MARKOWSKI, D.N., NIMZYK, R., MARIA HELMKE, B.M., et al., (2014). Correlated Expression of HMGA2 and PLAG1 in Thyroid Tumors, Uterine Leiomyomas and Experimental Models. **PLoS ONE**. 9.

LEE, Y.S., DUTTA, A. (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. **Genes Dev**, 1, 21.

LI, P., XIAO, S., WEI, N., ZHANG, Z., HUANG, R., GU, Y., et al., (2012). Fine mapping of a QTL for ear size on porcine chromosome 5 and identification of high mobility group AT-hook 2 (HMGA2) as a positional candidate gene. **Genetics, Selection, Evolution**. 44.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., et al., (2009) The Sequence Alignment/Map format and SAMtools. **Bioinformatics**. 25, 2078-2079.

LIVAK, K.J., SCHMITTGEN, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. **Methods**. 25, 402-408.

MAKVANDI-NEJAD, S., HOFFMAN, G. E., ALLEN, J.J., CHU, E., GU, E., CHANDLER, A.M., et al., (2012). Four loci explain 83% of size variation in the horse. **PLoS ONE**. 7, 1-6.

MCLAREN ,W., GIL, L., HUNT, S.E., RIAT, H.S., RITCHIE, G.R., THORMANN, A., et al., (2016). The Ensembl Variant Effect Predictor, **Genome Biol**. 6, 122.

O'CONNELL, J., GURDASANI, D., DELANEAU, O., PIRASTU, N., ULIVI, S., COCCA, M., et al., (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. **PLOS Genetics**. 10.

PEREIRA, A.G.T., UTSUNOMIYA,Y.T., MILANESI, M., TORRECILHA, R.B.P., CARMO, A.S., NEVES, H.H.R., et al., (2016). Pleiotropic genes affecting carcass traits in Bosindicus (Nellore) cattle are modulators of growth. **PLOS ONE**. 11.

PORTO-NETO, L.R., REVERTER, A., PRAYAGA, K. C., CHAN, E. K. F., DAVID, J., JOHNSTON, D. J., et al., (2014). The Genetic Architecture of Climatic Adaptation of Tropical Cattle. **PLOS ONE**, 9(11): e113284.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., et al., (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**. 81, 559-575.

QUINLAN, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. **Bioinformatics**. 47, 1- 34.

RABELO, R.E., SILVA, L.A.F., BRITO, L.A.B., MOURA, M.I, SILVA, O.C., CARVALHO, V.S., et al., (2008). Epidemiological aspects of surgical diseases of the genital tract in a population of 12,320 breeding bulls (1982-2007) in the state of Goiás, Brazil.**Ciência Animal Brasileira**. 9, 705-713.

ROBINSON, J.T., THORVALDSDÓTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E.S., GETZ, G. (2011) Integrative genomics viewer. **Nature Biotechnology**. 29, 24-26.

ROSENZWEIG, S.A., ATREYA, H.S. (2010). Defining the pathway to insulin-like growth factor system targeting in cancer. **Biochemical Pharmacology**. 80, 1115-1124.

SAATCHI, M.; SCHNABEL, R. D.; TAYLOR, J. F.; GARRICK, D. J. (2014). Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics**. 15, 442.

THORVALDSDÓTTIR, H.; ROBINSON, J. T.; MESIROV, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. **Briefings in Bioinformatics**. 14, 178-192.

UTSUNOMIYA, Y.T., MILANESI, M., UTSUNOMIYA, A.T.H., TORRECILHA, R.B.P., KIM, E.S., COSTA, M.S., et al., (2017). A PLAG1 mutation contributed to stature recovery in modern cattle. **Scientif reports**. 7.

UTSUNOMIYA, Y.T., MILANESI, M., UTSUNOMIYA, A.T.H., AJMONE-MARSAN, P., GARCIA, J. F. (2016).GHap: an R package for genome-wide haplotyping. **Bioinformatics**. 32, 286 1–2862.

UTSUNOMIYA, Y.T., CARMO, A.S., CARVALHEIRO, R., NEVES, H.H., MATOS, M.C., ZAVAREZ, L. B., et al., (2013). Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC Genetics**, 14, 52.

VON AHSEN, I., NIMZYK, R., KLEMKE, M., BULLERDIEK, J. (2008). A microRNA encoded in a highly conserved part of the mammalian *HMGA2* gene. **Cancer Genet Cytogenet**. 187, 43-44.

WANG, K., LI, M., HADLEY, D., LIU, R., GLESSNER, J., GRANT, S.F., et al., (2007).PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Res**. 17.

WEEDON, M.N., LETTRE, G., FREATHY, R.M., LINDGREN, C.M., VOIGHT, B.F., PERRY, J.R., et al., (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. **Nat Genet.** 39, 1245–1250.

YANG, T.L., GUO, Y., ZHANG, L.-S., TIAN, Q., YAN, H., GUO, Y.F., et. al., (2010). HMGA2 Is Confirmed To Be Associated with Human Adult Height. *Annals of Human Genetics.* 74, 11–16.

YANG, J., LEE, S.H., GODDARD, M.E., VISSCHER, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. **American Journal of Human Genetics**, 88, 76–82.

YANG, J., ZAITLEN, N.A., GODDARD, M.E., VISSCHER, P.M.; PRICE, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. **Nature Genetics.** 46 100-106.

ZAVAREZ, L.B., UTSUNOMIYA, Y.T., CARMO, A.S., NEVES, H.H.R., CARVALHEIRO, R., FERENCAKOVIC, M., et al., (2015). Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. **Frontiers in Genetics**, 5, 1-8.

ZIMIN, A.V., DELCHER, A.L., FLOREA, L., KELLEY, D.R., SCHATZ, M.C., PUIU, D., et al., (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome Biology**, 10, R42.

ZHOU, Y., UTSUNOMIYA, Y.T., XU, L., HAY, EL H.A., BICKHART, D.M., ALEXANDRE, P.A, et al., (2016). Genome-wide CNV analysis reveals variants associated with growth traits in *Bos indicus*. **BMC Genomics**, 17, 419.