



Association of Copy Number Variation at Intron 3 of *HMGA2* With Navel Length in *Bos indicus*

Tamiris 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} and Yuri Tani Utsunomiya^{2,3*}

¹ Department of Preventive Veterinary Medicine and Animal Reproduction, School of Agricultural and Veterinarian Sciences, São Paulo State University, Jaboticabal, Brazil, ² Collaborating Centre on Animal Genomics and Bioinformatics, International Atomic Energy Agency, Araçatuba, Brazil, ³ Department of Support, Production and Animal Health, School of Veterinary Medicine, São Paulo State University, Araçatuba, Brazil, ⁴ Department of Animal Science Food and Nutrition and Biodiversity and Ancient DNA Research Center – BioDNA, Università Cattolica del Sacro Cuore, Piacenza, Italy, ⁵ Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom, ⁶ Faculty of Medical Laboratory Sciences, University of Khartoum, Khartoum, Sudan, ⁷ Department of Animal Science, School of Agricultural and Veterinarian Sciences, São Paulo State University, Jaboticabal, Brazil, ⁸ GenSys Consultores Associados, Porto Alegre, Brazil, ⁹ Escola de Veterinária e Zootecnia, Universidade Federal de Goiás, Goiânia, Brazil, ¹⁰ LiveGene – CTLGH, International Livestock Research Institute, Addis Ababa, Ethiopia, ¹¹ Recombinetics, Inc., Saint Paul, MN, United States

OPEN ACCESS

Edited by:

Denis Milan,
Institut National de la Recherche
Agronomique (INRA), France

Reviewed by:

Christine Couldrey,
Livestock Improvement Corporation,
New Zealand
Florence Phocas,
INRA Centre Jouy-en-Josas, France

*Correspondence:

Yuri Tani Utsunomiya
yutunomiya@gmail.com

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 03 August 2018

Accepted: 23 November 2018

Published: 07 December 2018

Citation:

Aguiar TS, Torrecilha RBP, Milanesi M, Utsunomiya ATH, Trigo BB, Tijjani A, Musa HH, Lopes FL, Ajmone-Marsan P, Carneiro R, Neves HHdR, do Carmo AS, Hanotte O, Sonstegard TS, Garcia JF and Utsunomiya YT (2018) Association of Copy Number Variation at Intron 3 of *HMGA2* With Navel Length in *Bos indicus*. *Front. Genet.* 9:627. doi: 10.3389/fgene.2018.00627

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 Nelore cattle (*Bos indicus*) using data from 2,016 animals and 503,088 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 Nelore bulls 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 in cattle, these findings highlight the emerging role of variants impacting the insulin-like growth factor pathway to cattle breeding.

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

INTRODUCTION

Navel length at yearling (NY) 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 et al., 2016). This trait is especially important to cattle raised in Brazil, the second largest beef exporter of the world, where Nellore (*Bos indicus*) animals compose the majority of the herds and approximately 88% of the females in reproductive age are subjected to natural mating (Asbia, 2016). Consequently, visual scores of navel length are routinely recorded in national commercial breeding programs together with other growth traits such as body weight and carcass visual scores (Neves et al., 2014).

A recent genome-wide association study (GWAS) for body weight and carcass visual scores in Nellore cattle revealed pleiotropic loci sheltering genes that are known to modulate the insulin-like growth factor (IGF) or somatomedin pathway (Pereira et al., 2016). In particular, the locus containing the pleomorphic adenoma gene 1 (*PLAG1*), which is a transcription factor for the insulin-like growth factor gene 2 (*IGF2*), contributed significantly to the genetic variance of all traits analyzed. Indeed, the *PLAG1* chromosomal domain has been shown to affect a wide range of economically relevant traits in both *Bos taurus* (Karim et al., 2011; Bolormaa et al., 2014; Saatchi et al., 2014) and *B. indicus* (Fortes et al., 2013; Utsunomiya et al., 2013; Pereira et al., 2016) cattle. However, genome-wide scans for visual scores of navel length were not yet reported in *B. indicus* cattle, leaving opened the question of whether polymorphisms in *PLAG1* or any other modulator of the IGF pathway also contribute to the genetic variance of navel length in that species.

Here we report a haplotype-based GWAS for visual scores of navel length at yearling (NY) in Nellore cattle, which resulted in the discovery of a copy number variation (CNV) at intron 3 of *HMGA2*, another transcription factor for *IGF2* 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.

MATERIALS AND METHODS

Genotypes

Illumina BovineHD BeadChip assay (786,799 SNPs) genotypes of 953 Nellore bulls and 1,278 Nellore cows were available for analysis from previous studies (Carvalho et al., 2014; Zavarez et al., 2015). These animals were born between 1963 and 2008, parented over ten Nellore generations, and 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 resulted in a reduced set of 503,088 markers, which were retained for phasing with the Segmented HAPlotype Estimation and 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 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 length at yearling (NY) was recorded based on visual evaluation of animals by trained technicians. The navel length 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 was expected to account for any systematic effects associated to technicians. The distribution of NY phenotypes is presented in the **Supplementary Figure S1**. Prior to GWAS, EBVs were deregressed following Garrick et al. (2009) and filtered for a minimum accuracy of 0.70. This filter was used to reduce influence of heterogeneity of variance in the GWAS analysis and to ensure that the pseudo-phenotype of each animal was mainly based on own or progeny phenotypes, rather than information from parents or siblings. Therefore, the GWAS analysis was based on a subset of 2,016 animals.

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. Theoretical support to this approach was previously detailed by Da (2015). 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:

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

where 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 (i.e., polygenic scores), and e is the $n \times 1$ vector of random residual effects. It was assumed that $y \sim \text{MVN}(\mathbf{1}_n\mu + \mathbf{x}b, \mathbf{G}\sigma_u^2 + \mathbf{I}\sigma_e^2)$, where \mathbf{G} is the SNP-based genomic relationship matrix (GRM) excluding all SNPs pertaining to the same chromosome where the tested haplotype is located, 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. 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 (i.e., sliding windows of six markers moving one marker at a time), 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.

Association With Additional Traits

The most significant haplotype in the GWAS analysis was further subjected to a screening for pleiotropic effects by taking advantage of deregressed EBVs that were also available for nine additional traits, namely: (1) birth weight, (2) weight gain from birth to weaning, (3) weight gain from weaning to yearling, (3) scrotal circumference, (4) age at first calving, (5) gestation length, and visual scores for (6) conformation, (7) precocity, and (8) muscling at yearling. These traits are routinely recorded and used as part of the genetic evaluation system of the commercial breeding program, and the models used to obtain deregressed EBVs for these traits were previously described by Neves et al. (2014) and Pereira et al. (2016). For each trait, pseudo-phenotypes were regressed onto haplotype counts using linear regression, and effects were considered significant if the nominal p -value was smaller than 0.05. Following the same rationale as in the GWAS analysis, only animals presenting accuracy of deregressed EBV greater than 0.70 were analyzed. In order to evaluate potential interference in the regression analysis that could be caused by unaccounted heterogeneity of residual variance, we paralleled all analyses with weighted linear models using the weights proposed by Garrick et al. (2009).

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 *B. taurus* assembly (Zimin et al., 2009). The average sequencing coverage was $9.25\times$. 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 et al., 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 variant in other cattle breeds was investigated using re-sequencing data ($n = 13$) from five African zebu *B. indicus* breeds, 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) (Daetwyler et al., 2014). Per sample sequencing depth and number of animals per breed are provided in the **Supplementary File S1**. Comparative genomics and orthology analyses were performed using the UCSC (Casper et al., 2018) and Ensembl (Zerbino et al., 2018) genome browsers.

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' TCAAAGCCAAGTATTGCTG 3' (forward) and 5' TTTCTATGGTCCCAAGCG 3' (reverse) was designed considering the limits of the CNV region using the NCBI Primer-BLAST¹. 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. The qPCR analysis was then conducted using samples available from our DNA bank, and comprised 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 Length at Yearling in Nellore Cattle

A total of 2,048,168 six-marker haplotypes were tested for association with NY in Nellore cattle. The minimum, median and maximum number of observed haplotypes per six-marker window was 1, 4, and 15, respectively. Association results pointed to a single leading QTL region on chromosome 5 mapping between 47.9 and 48.2 Mbp (**Figure 1**). Four haplotypes from consecutive segments spanning positions 47,921,750–48,069,099 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 and exons 4 and 5 of *HMGA2*.

Additional Effects Associated With the *HMGA2* Haplotype

By using a panel of nine additional traits that are routinely recorded by the Nellore commercial breeding program, we investigated whether the *HMGA2* haplotype could also be associated with other economically important traits in Nellore cattle. After fitting regression models for all nine traits, we found weak but significant ($p < 0.05$) associations for visual scores of precocity and muscling at yearling (**Table 1**). The use of weights in the regression analyses had little impact on the results. Of note, the direction of the effects were opposite to that observed for

navel length at yearling, where the TCCTCCAAC haplotype was associated with higher scores for navel length and lower scores for precocity and muscling.

Nellore Sequence Data Reveals an Intronic CNV Correlated With the *HMGA2* Haplotype

Assuming that the TCCTCCAAC haplotype tagged an unobserved causal variant, we used haplotype counts to predict causal variant genotypes in a sample of 21 Nellore bulls with whole genome sequence data. From a total of 1,185 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 (CHR5:47,962,011, C > G), rs521422509 (CHR5:47,964,144, A > G), rs525140201 (CHR5:47,965,249, G > A), rs521794670 (CHR5:47,966,055, G > A), rs519732918 (CHR5:47,966,375, T > C), and rs516281664 (CHR5:47,967,920, C > T). While we could not exclude causality of any of these variants, none of them were novel or intragenic, making it difficult to select any specific variant for further scrutiny. However, analysis of coverage data revealed an additional variant comprising a \sim 6.2 kbp CNV spanning positions 48,074,233–48,080,443 (**Figure 2A**). This CNV was located at 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 \sim 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 \sim 6.2 kbp DNA segment. Inspection of alignments of the bovine *HMGA2*-CNVR sequence against the human reference genome assembly (GRCh38.p12) showed that the region spanned a *HMGA2* antisense RNA (accession numbers AC090673.2 and ENSG00000256083). The UCSC LiftOver tool further predicted that the antisense RNA would map to positions 48,070,427–48,085,449, which substantially overlapped with the detected CNV. The existence of AC090673.2 in the human genome was well supported by evidence from expressed sequence tags (ESTs) obtained from pooled human melanocyte, fetal heart and pregnant uterus samples (accession number AA773790.1). Therefore, the *HMGA2*-CNVR stood as the most plausible candidate variant underlying the NY associations.

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 2B**). Four out of five *B. indicus* breeds presented increased coverage compatible with the *HMGA2*-CNVR previously found in Nellore cattle, whereas all

¹<http://www.ncbi.nlm.nih.gov/tools/primer-blast>

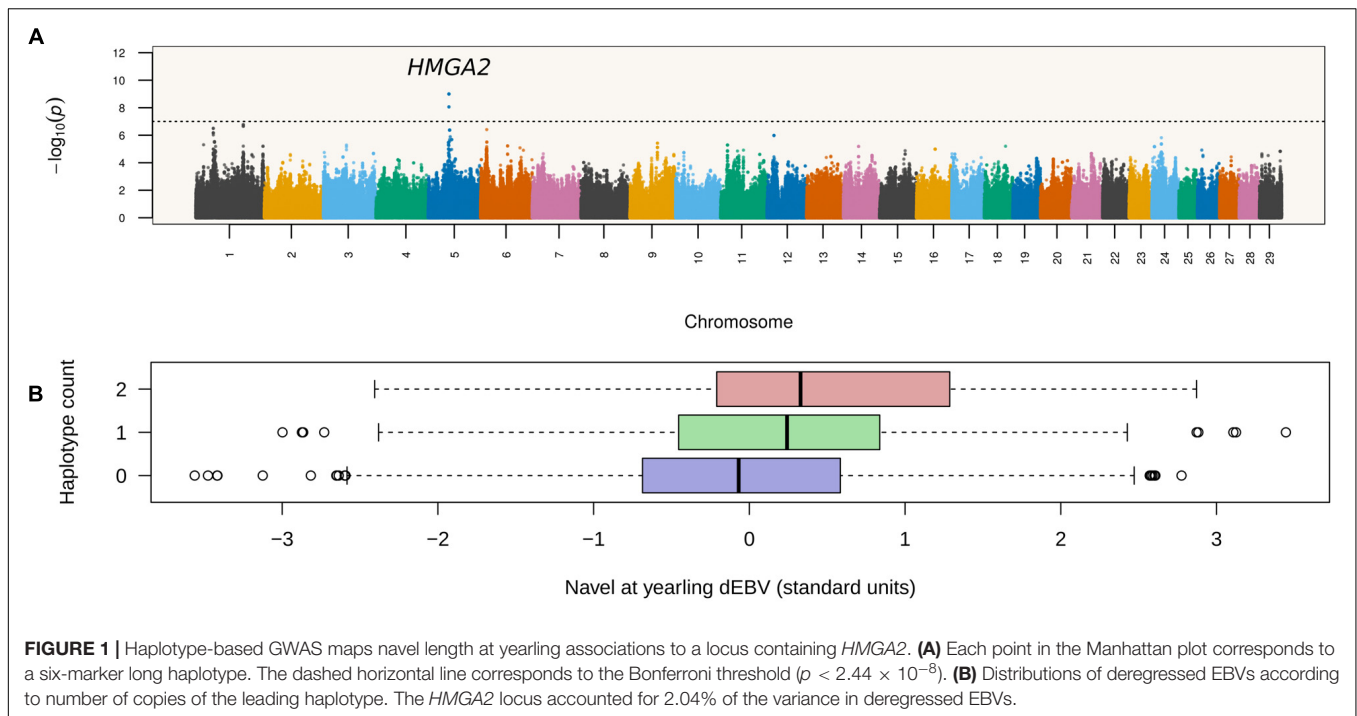


TABLE 1 | Effects of the *HMGA2*-CNVR tag haplotype on deregressed estimated breeding values of nine traits in Nellore cattle.

Trait	Sample size	Unweighted regression ^a			Weighted regression ^b		
		Coefficient	SE	p-value	Coefficient	SE	p-value
Birth weight	1,993	0.037	0.055	0.505	0.047	0.052	0.362
Weight gain from birth to weaning	2,108	0.325	0.262	0.215	0.309	0.264	0.241
Weight gain from weaning to yearling	2,052	-0.243	0.355	0.494	-0.221	0.356	0.536
Visual score for conformation	2,023	0.030	0.019	0.116	0.026	0.019	0.166
Visual score for precocity	2,022	-0.045	0.026	0.084	-0.055	0.026	0.036
Visual score for muscling	2,020	-0.060	0.025	0.017	-0.070	0.025	0.005
Scrotal circumference	1,627	-0.041	0.058	0.474	-0.050	0.058	0.396
Age at first calving	636	0.417	0.526	0.429	0.455	0.445	0.307
Gestation length	1,951	0.077	0.133	0.561	0.145	0.130	0.263

Significant effects ($p < 0.05$) are marked in bold.

^aSimple linear regression analysis; ^bWeighted regression using weights proposed by Garrick et al. (2009).

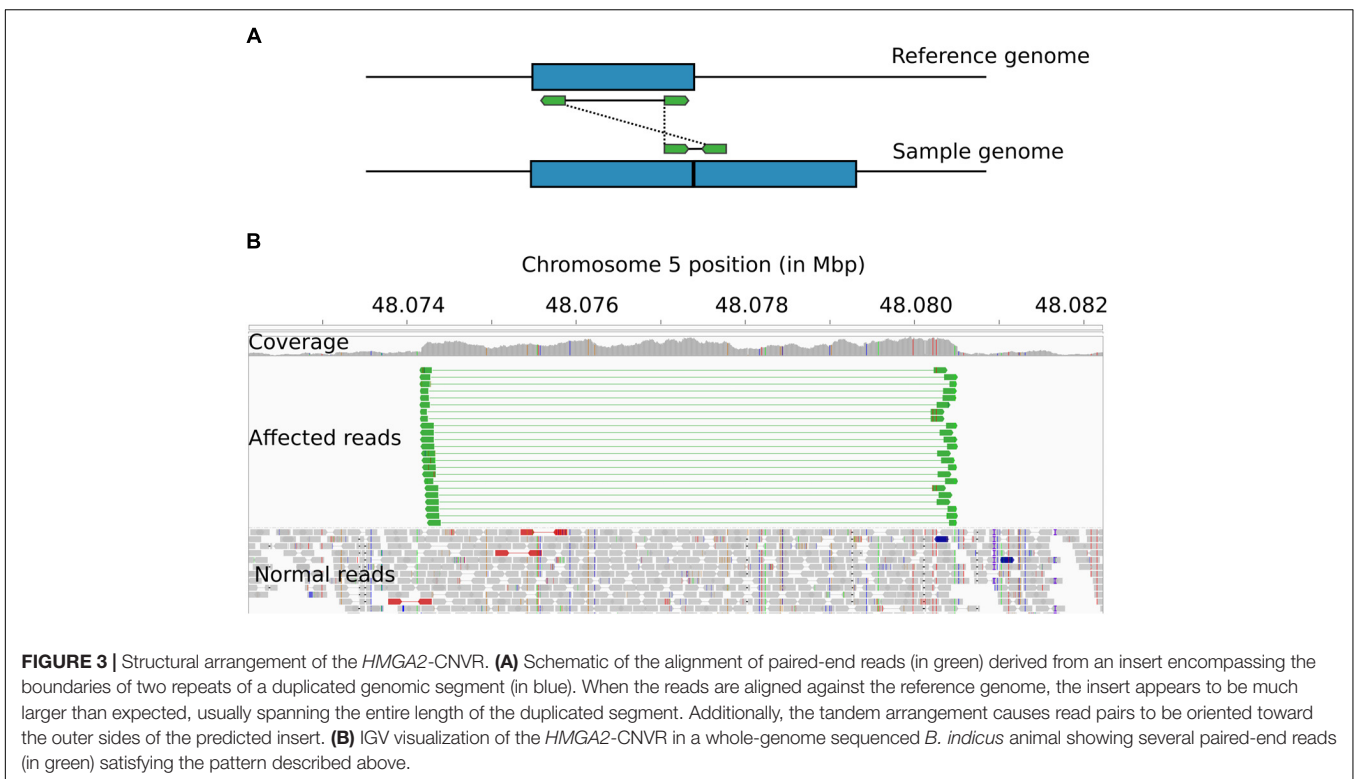
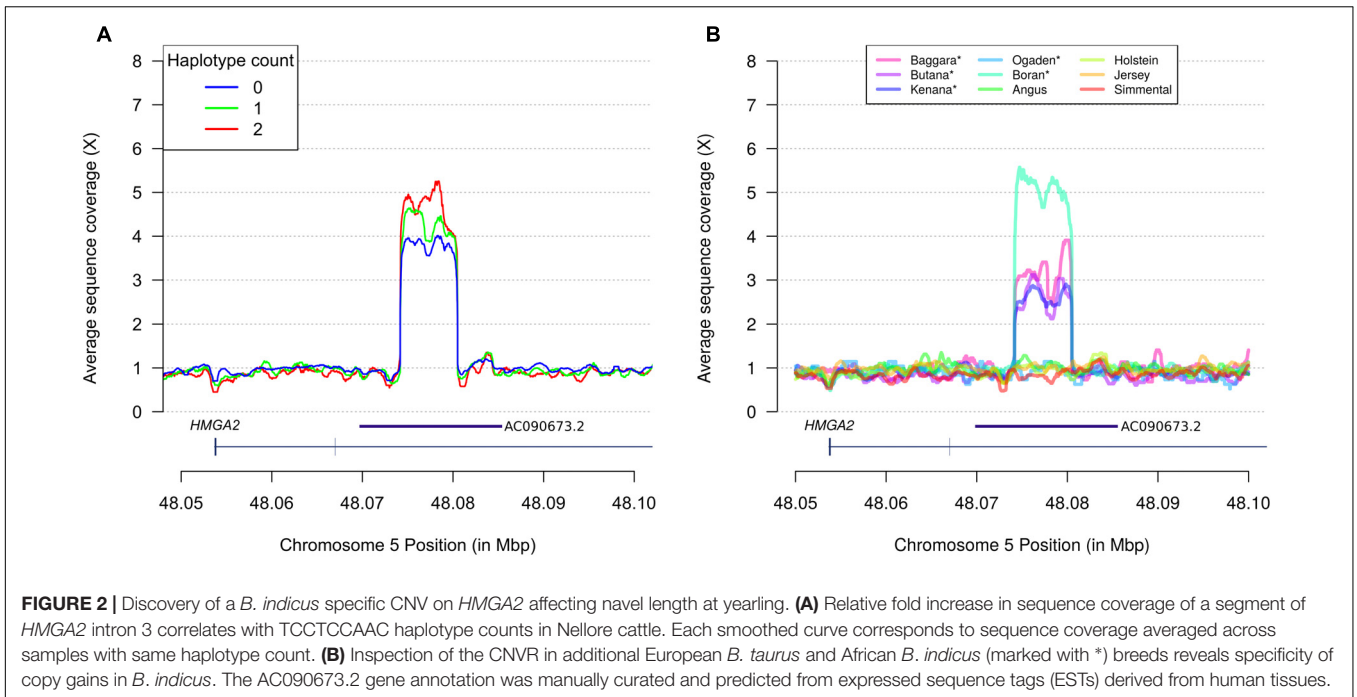
B. taurus breeds were copy neutral at the relevant positions on chromosome 5. These results strongly point to an indicine-specific CNV.

Read Pair Orientation Analysis Solves the *HMGA2*-CNVR Structural Arrangement

Visualization of read alignments in *B. indicus* samples with the IGV software showed inserts with larger than expected sizes that spanned the whole extension of the *HMGA2*-CNVR segment, with paired-end reads oriented to the outer sides of the inserts. This pattern of read pair orientation is indicative of duplications arranged in tandem with structurally identical repeats (Figure 3). Therefore, the *HMGA2*-CNVR structure is most likely explained by a tandem duplication.

qPCR Validates Haplotype-Based Predictions of CNV Genotypes

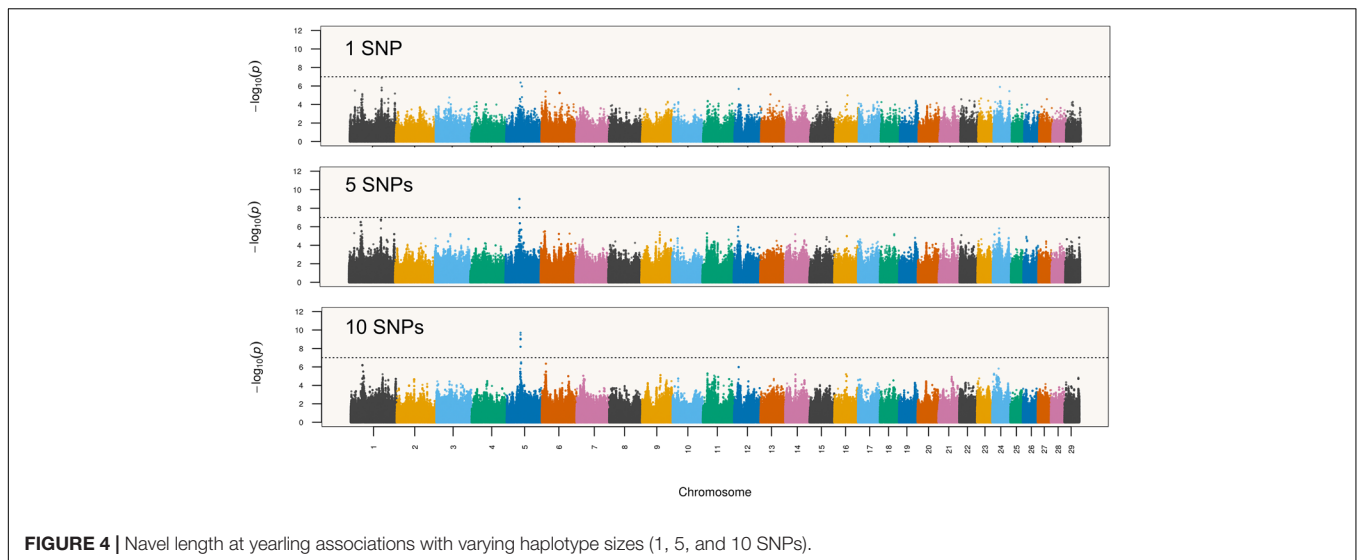
The TCCTCCAAC haplotype tag suggested that the frequency of additional copy gains at the *HMGA2*-CNVR in our Nellore sample was 16.5%. The numbers of animals carrying 0, 1, and 2 copies of the tag haplotype were 1,554, 619, and 58, respectively. Based on available samples in our DNA bank, we randomly 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 only three samples. We also used DNA from 4 Angus animals as copy neutral controls, assuming that *B. taurus* animals were copy neutral as observed in the sequence coverage analysis. The average $2^{-\Delta\Delta Ct}$ 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 Failed to Detect the *HMGA2*-CNVR

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 (1,902,119 tested haplotypes) and 10 markers (2,441,374 tested haplotypes) (**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 variant 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, United States) 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.

DISCUSSION

The 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 inhibition of apoptosis (Rosenzweig and Atreya, 2010; King and Wong, 2012). The main components 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 *B. taurus* (Karim et al., 2011; Bolormaa et al., 2014; Saatchi et al., 2014) and *B. indicus* (Fortes et al., 2013; Utsunomiya et al., 2013; Pereira et al., 2016) cattle, implicating a major contribution of variants impacting the IGF system to genetic variance of several traits in the bovine species.

The discovery of multiple quantitative trait loci (QTL) mapping to the bovine *PLAG1* chromosomal domain opens the question of whether other regulatory genes and variants affecting the expression of *IGF1* and *IGF2* might also lead to phenotypic differences of economical relevance in cattle. For instance, the major transcription factor affecting the expression of *IGF2* is encoded by the high mobility AT-hook 2 gene (*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., 2018) (**Figure 6**). Here, we reported the identification of a CNV located at intron 3 of *HMGA2* that was associated with navel length at yearling (NY) in Nellore (*B. indicus*) cattle. This CNV was detected through a combination of haplotype-based GWAS and whole genome sequence analysis. We also showed that this candidate variant was difficult to identify via conventional GWAS and state-of-the-art 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; Abi Habib et al., 2018). However, the importance of *HMGA2* as an oncogene, and its regulation by a series of

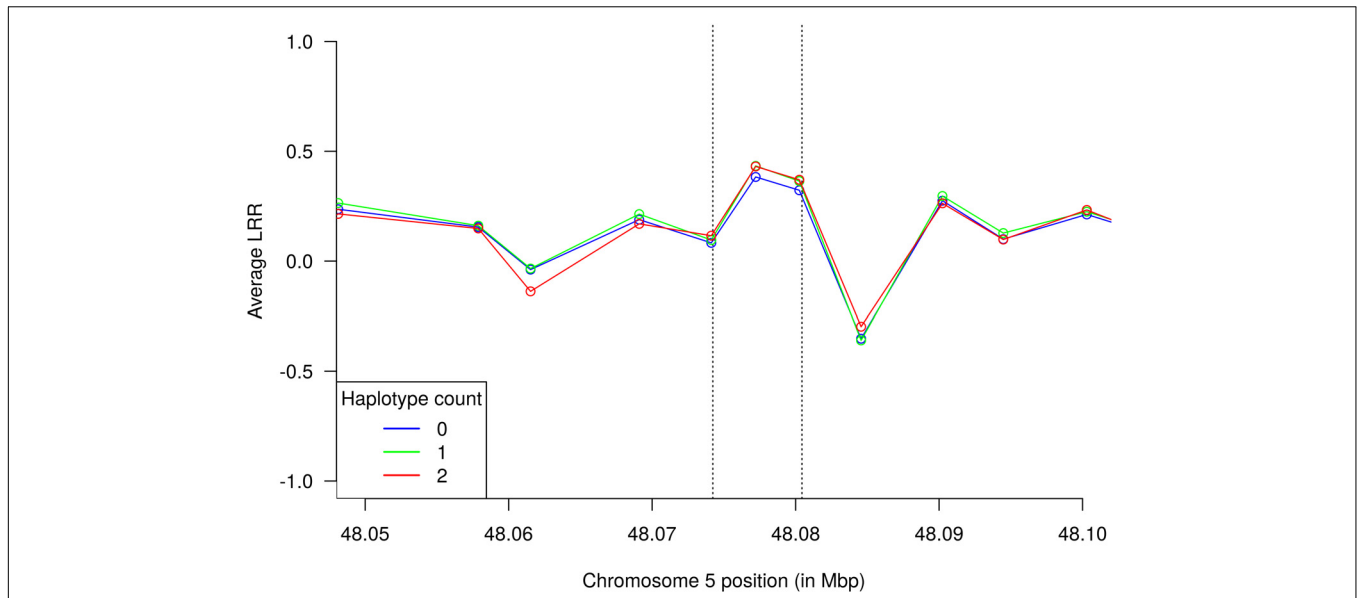


FIGURE 5 | Average probe intensity data from the BovineHD assay for the *HMGA2*-CNVR segment (vertical dashed lines). Each curve corresponds to average LRR values of samples with the same TCCTCCAAC haplotype count. Points in the curves indicate positions of SNP markers.

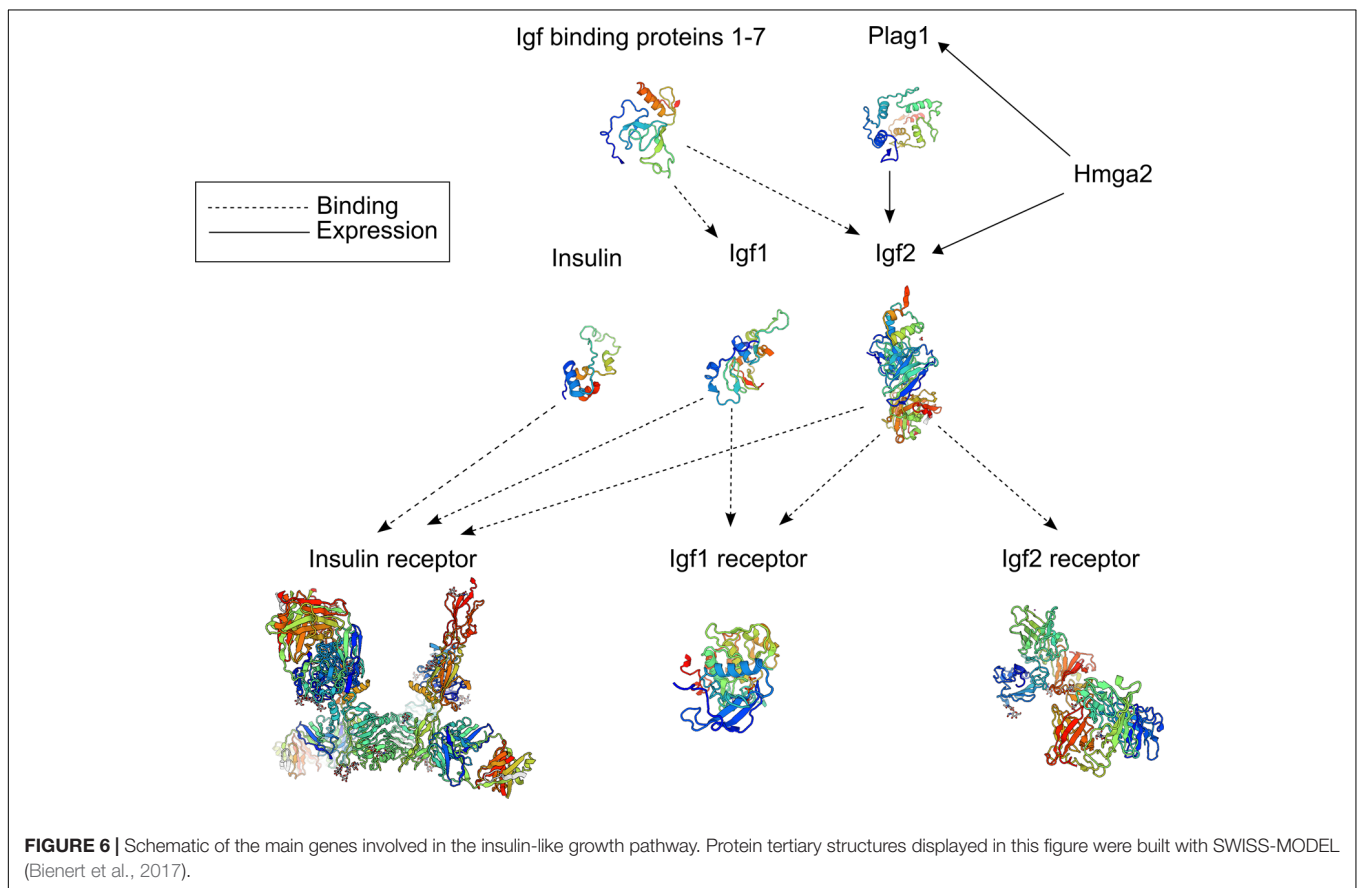


FIGURE 6 | Schematic of the main genes involved in the insulin-like growth pathway. Protein tertiary structures displayed in this figure were built with SWISS-MODEL (Bienert et al., 2017).

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 (Artzi et al., 2008; von Ahsen 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 antisense, long non-coding RNA (AC090673.2), 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 *B. indicus* 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), indicating that the *HMGA2*-CNVR might also segregate in these populations.

Apart from the observed effects of the *HMGA2*-CNVR on navel size, the candidate 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). In the present study alone, we found that other productive traits such as visual scores for precocity and muscling at yearling may also be affected by the candidate variant. Also, the detection of the *HMGA2*-CNVR in African *B. indicus* cattle suggests that the variant might be very old. 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 candidate variant is probably a wide spread ancient *B. indicus* mutation, which might have had an adaptive value in this subspecies. Given that the hump, the dewlap (i.e., saggy chest and neck skin) and the pendulous navel are phenotypic marks that morphologically differentiate *B. indicus* from *B. taurus* cattle, future analyses of the *HMGA2*-CNVR should address the question of whether this variant had a role in the evolutionary history of the *B. indicus* lineage.

REFERENCES

- 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. *Genet. Med.* 20, 250–258. doi: 10.1038/gim.2017.105
- Ajmone-Marsan, P., Garcia, J. F., Lenstra, J. A., and The Globaldiv Consortium. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* 19, 48–157. doi: 10.1002/evan.20267

DATA AVAILABILITY

Whole genome sequences of African and European cattle are available through the NCBI SRA Archive under BioProjects PRJNA312138 and PRJNA238491, respectively. The Nellore datasets are not publicly available because they comprise part of the genomic selection reference population from a commercial breeding program ran by an alliance of cattle breeders from Brazil. Data are, however, available for academic use from the authors upon request (requires a signed Material Transfer Agreement for exclusive research purpose).

AUTHOR CONTRIBUTIONS

YU and JG conceived and designed the study. AdC and TS coordinated genotyping and sequencing of Nellore animals. AT, HM, and OH provided and analyzed African cattle sequence data. RC and HN analyzed phenotypes and pedigree data and generated deregressed EBVs. AU performed genotype phasing. TA and YU performed association analyses. MM, YU, and PA-M designed the qPCR experiments. TA, BT, RT, and FL performed the qPCR assays. TA, RT, and MM performed Nellore and *B. taurus* sequence data analyses. TA and YU wrote the manuscript. All authors revised and agreed with the contents of the manuscript.

FUNDING

This research was supported by the São 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).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00627/full#supplementary-material>

FILE S1 | Sequencing depth and accession numbers of cattle whole genome sequences.

FIGURE S1 | Distribution of visual scores of navel length at yearling in 745,466 Nellore steers.

- Artzi, S., Kiezun, A., and Shomron, N. (2008). miRNAmir: A tool for homologous microRNA gene search. *BMC Bioinformatics* 9:39. doi: 10.1186/1471-2105-9-39
- Asbia. (2016). *Associação Brasileira de Inseminação Artificial. Relatório index do ano de 2014*. Available at: <http://www.asbia.org.br/novo/relatorios/> [accessed October 15, 2018]
- Ashdown, R. R. (2006). Functional, developmental and clinical anatomy of the bovine penis and prepuce. *CAB Rev.* 21, 29–37. doi: 10.1079/PAVSNNR20061021

- Barbato, M., Orozco-Terwengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome wide SNP data. *Front. Genet.* 6:109. doi: 10.3389/fgene.2015.00109
- Bergman, D., Halje, M., Nordin, M., and 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/gkx1132
- 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. doi: 10.1007/s11250-016-1143-1
- 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 Genet.* 10:4198. doi: 10.1371/journal.pgen.1004198
- 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. doi: 10.1038/s41588-018-0056-5
- 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.* 8:451. doi: 10.1371/journal.pbio.1000451
- Casper, J., Zweig, A. S., Villarreal, C., Tyner, C., Speir, M. L., Rosenbloom, K. R., et al. (2018). The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* 46, D762–D769. doi: 10.1093/nar/gkx1020
- Carvalho, 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. *Genet. Sel. Evol.* 46:69. doi: 10.1186/s12711-014-0069-1
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8
- Da, Y. (2015). Multi-allelic haplotype model based on genetic partition for genomic prediction, and variance component estimation using SNP markers. *BMC Genet.* 18:144. doi: 10.1186/s12863-015-0301-1
- 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. doi: 10.1038/ng.3034
- DeChiara, T. M., Robertson, E. J., and Efstratiadis. (1991). A Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859. doi: 10.1016/0092-8674(91)90513-X
- 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 Res.* 36:e126. doi: 10.1093/nar/gkn556
- 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 Nelore cattle. *BMC Genomics* 14:305. doi: 10.1186/1471-2164-14-305
- 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* 3, 372–378. doi: 10.1242/bio.20146759
- Fortes, M. R. S., Reverter, A., Kelly, M., McCulloch, R., and 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. doi: 10.1111/j.2047-2927.2013.00101.x
- Frischknecht, M., Jagannathan, V., Plattet, P., Neuditschko, M., Signer-Hasler, H., Bachmann, L., et al. (2015). A non-synonymous HMG2 variant decreases height in shetland ponies and other small horses. *PLoS One* 10:140749. doi: 10.1371/journal.pone.0140749
- Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55. doi: 10.1186/1297-9686-41-55
- 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. doi: 10.1126/science.1069878
- Hayward, J. J., Castelhan, M. G., Oliveira, K. C., Corey, E., Balkman, C., Baxter, T. L., et al. (2016). Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* 7:10460. doi: 10.1038/ncomms10460
- 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. *Dev. Growth Differ.* 46, 459–470. doi: 10.1111/j.1440-169x.2004.00762.x
- 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. *Nat. Genet.* 43, 405–413. doi: 10.1038/ng.814
- Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., et al. (2017). The genome landscape of indigenous African cattle. *Genome Biol.* 18:34. doi: 10.1186/s13059-017-1153-y
- King, E. R., and Wong, K. (2012). Insulin-like growth factor: current concepts and new developments in cancer therapy. *Recent Patents Anti-Cancer Drug Discov.* 7, 14–30. doi: 10.2174/157489212798357930
- Klemke, K., Müller, M. H., Wosniok, W., Markowski, D. N., Nimzyk, R., Maria Helmke, B. M., et al. (2014). Correlated expression of HMG2 and PLAG1 in thyroid tumors, uterine leiomyomas and experimental models. *PLoS One* 9:88126. doi: 10.1371/journal.pone.0088126
- Lee, Y. S., and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMG2 oncogene. *Genes Dev.* 1:21. doi: 10.1101/gad.1540407
- 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. doi: 10.1093/bioinformatics/btp352
- 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 (HMG2) as a positional candidate gene. *Genet. Sel. Evol.* 44:6. doi: 10.1186/1297-9686-44-6
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔC_T} Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- 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. doi: 10.1186/s13059-016-0974-4
- Neves, H. H., Carvalho, R., O'Brien, A. M., Utsunomiya, Y. T., do Carmo, A. S., Schenkel, F. S., et al. (2014). Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle. *Genet. Sel. Evol.* 27:17. doi: 10.1186/1297-9686-46-17
- 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 Genet.* 10:4234. doi: 10.1371/journal.pgen.1004234
- Pereira, A. G. T., Utsunomiya, Y. T., Milanese, M., Torrecilha, R. B. P., Carmo, A. S., Neves, H. H. R., et al. (2016). Pleiotropic genes affecting carcass traits in *Bos indicus* (Nelore) cattle are modulators of growth. *PLoS One* 11:158165. doi: 10.1371/journal.pone.0158165
- 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:e113284. doi: 10.1371/journal.pone.0113284
- 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. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Quinlan, A. R. (2014). BEDTools: the swiss-army tool for genome feature analysis. *Bioinformatics* 47, 1–34. doi: 10.1002/0471250953.bi1112s47
- 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 Anim. Brasil.* 9, 705–713.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., and Getz, G. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Rosenzweig, S. A., and Atreya, H. S. (2010). Defining the pathway to insulin-like growth factor system targeting in cancer. *Biochem. Pharmacol.* 80, 1115–1124. doi: 10.1016/j.bcp.2010.06.013
- Saatchi, M., Schnabel, R. D., Taylor, J. F., and Garrick, D. J. (2014). Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* 15:442. doi: 10.1186/1471-2164-15-442

- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017
- Utsunomiya, Y. T., Carmo, A. S., Carvalho, 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 Genet.* 14:52. doi: 10.1186/1471-2156-14-52
- Utsunomiya, Y. T., Milanesi, M., Utsunomiya, A. T. H., Ajmone-Marsan, P., and Garcia, J. F. (2016). GHap: an R package for genome-wide haplotyping. *Bioinformatics* 32, 1–2862. doi: 10.1093/bioinformatics/btw356
- 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. *Sci. Rep.* 7:17140. doi: 10.1038/s41598-017-17127-1
- von Ahnen, I., Nimzyk, R., Klemke, M., and Bullerdiek, J. (2008). A microRNA encoded in a highly conserved part of the mammalian HMGA2 gene. *Cancer Genet. Cytogenet.* 187, 43–44. doi: 10.1016/j.cancergencyto.2008.07.009
- 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, 1665–1674. doi: 10.1101/gr.6861907
- 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. doi: 10.1038/ng2121
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106. doi: 10.1038/ng.2876
- 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. *Ann. Hum. Genet.* 74, 11–16. doi: 10.1111/j.1469-1809.2009.00555.x
- Zavarez, L. B., Utsunomiya, Y. T., Carmo, A. S., Neves, H. H. R., Carvalho, R., Ferencakovic, M., et al. (2015). Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. *Front. Genet.* 5:5. doi: 10.3389/fgene.2015.00005
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46, D754–D761. doi: 10.1093/nar/gkx1098
- 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. doi: 10.1186/s12864-016-2461-4
- 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 Biol.* 10:R42. doi: 10.1186/gb-2009-10-4-r42

Conflict of Interest Statement: HN is employed by Gensys Consultores Associados and TS is employed by Recombinetics, Inc. TS and YU are Associate Editors at Frontiers in Genetics, Livestock Genomics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Mention of trade name proprietary product or specified equipment in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their respective institutions.

Copyright © 2018 Aguiar, Torrecilha, Milanesi, Utsunomiya, Trigo, Tijjani, Musa, Lopes, Ajmone-Marsan, Carvalho, Neves, do Carmo, Hanotte, Sonstegard, Garcia and Utsunomiya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.