Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.892

Genome-wide distribution of allele-specific expression in Nelore steers muscle

M. M. Souza[1], A. Zerlotini[2], S.C.M. Niciura[1], M.I.P. Rocha, W.J.S. Diniz[3], P.C. Tizioto[4], P.S.N. Oliveira[1], A.O. Lima[3], J. Afonso[3], A.S.M. Cesar[2], L.L. Coutinho[2] & L. C. A. Regitano[1]

[1] Embrapa Pecuária Sudeste, Washington Luiz Highway, Km 234, 13560-970, São Carlos, Brazil
mm.rissato@gmail.com *(Corresponding Author)*
[2] Embrapa Informática Agropecuária, AV. Dr. André Tosello, 209 – Cidade Universitária, Campinas, 13083-886, São Paulo, Brazil
[3] Federal University of São Carlos - UFSCar, Department of Genetic and Evolution, São Carlos, Washington Luiz Highway, Km 235, 13565-905, São Paulo
Brazil
[4] University of São Paulo - ESALQ, Department of Animal Science, Pádua Dias Ave., Piracicaba, 13418-900, São Paulo, Brazil

Summary

The difference between the allelic expression of a gene is known as allele-specific expression (ASE), and it is a common event in mammal's transcriptome. ASE is named genomic imprinting when guided by epigenetics mechanisms regarding the allele parental origin. Although these events are an important sources of phenotypic variation, knowledge thereof regarding cattle and effects on economically important traits are still incipient. Therefore, analysis and characterization of ASE profile in the muscle of adult Nelore animals were performed, and besides that, the ASE pattern and gene imprinting occurrence were checked in genes previously associated with meat tenderness. For this, RNA sequencing data of 146 Nelore steer muscles were used, and these animals were also genotyped for approximately 770,000 SNPs using Illumina BovineHD BeadChip. For each animal, a diploid genome was built based on their genotypes, and afterward, RNA-Seq reads were mapped to this genome. After counting the reads considering each SNP allele, a binomial test was applied to identify whether the allele contribution to total expression significantly diverged from 50%, followed by false discovery rate (FDR) test (0.05). Among 519 analyzed SNPs, 402 showed at least one sample with ASE, and they were spread among all autosomal chromosomes. SNPs showing ASE in more than five samples were tested regarding ASE extension, and most presented an incomplete pattern. Four ASE genes were previously identified in a Genome-Wide Association Study (GWAS) and five were found differentially expressed between extreme phenotype animals for meat tenderness. Therefore, the results will complement the bovine imprinted genes database, further they can provide new information to genetic effects prediction in animal breeding increasing the accuracy of meat tenderness selection.

Keywords: allelic, ASE, gene expression, sequencing, transcriptome

Introduction

According to Mendelian inheritance model, the equal contribution of the parental allele to gene expression is expected (Saupe, 2012). However, much evidence of unequal expression of the alleles of a gene has been found in mammals' genome (Tycko, 2010). Allele-specific expression (ASE) may be the consequence of genetics or epigenetic events, which could result in phenotypic variance among individuals (Muráni *et al.,* 2009; Tuch *et al.,* 2010) and bias in the predicted additive effect of molecular markers associated with production traits.

Polymorphisms in cis-regulatory sites are the principal cause of ASE. These *loci* affect only the DNA molecule in which they are inserted (Lagarrigue *et al.,* 2013). In addition, ASE also can be governed by epigenetic events, as genomic imprinting, in which a DNA methylation distinguishes the chromosomes inherited maternal and paternally resulting in a pattern of expression depending of the parental origin of the allele (Ruvinsky, 1999).

Even complex traits and diseases can be affected by genes showing allele-specific expression pattern, only one study was developed in the bovine whole-transcriptome (Chamberlain *et al.,* 2015). By RNA sequencing, Chamberlain *et al.* (2015) built the EAE profile of a dairy cow in 18 tissues. Although these authors could answer some questions about tissue-specific expression, other questions about repeatability and individual-specific patterns remain unresolved.

The large difference of available data and knowledge between human and bovine imprinting (http://igc.otago.ac.nz/) reinforces the need for the development of new studies aiming to describe the expression patterns of interest for animal production since economically important traits can be affected by variation in allelic expression.

Considering this scenario, the present study aimed to build the EAE profile of bovine *Longissimus dorsi*, to check the variation of EAE among individuals contributing to the elaboration of more accurate strategies of inclusion of genomic information in animal breeding programs.

Material and methods

The Institutional Animal Care and Use Committee Guidelines from Embrapa (Brazilian Agricultural Research Corporation) approved all the experimental procedures involving animals in this study.

Animals

One hundred forty-six Nelore steers, sired by 34 unrelated sires, were produced by Embrapa, Brazil. The steers were raised in feedlot system, receiving similar nutritional and sanitary management. After about 90 days of feeding the animals were slaughtered and *Longissimus dorsi* (LD) muscle samples were collected for RNA extraction.

DNA Genotyping and RNA sequencing

DNA was extracted from blood and genotyped by using Illumina BovineHD BeadChip (Illumina Inc, San Diego, CA, EUA), in two laboratories: USDA ARS Bovine Functional Genomics (Beltsville, MD, EUA) and ESALQ genomic center (Piracicaba, São Paulo, Brazil). Quality control was accessed by PLINK (Purcell *et al.,* 2007) and the linkage phase of the SNPs was obtained by Beagle 4.0 (Browning & Browning, 2007). All heterozygous

SNPs of each animal were identified to analyze the ASE.

Expression profiles of mRNA were measured with RNA-Seq in LD muscle samples (n = 146). RNA extraction, library preparation, and sequencing were realized at ESALQ genomic center (Piracicaba, São Paulo, Brazil). RNA library was prepared using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, EUA). Clustering and sequencing were done on Illumina HiSeq 2500® (Illumina, San Diego, CA, EUA).

Allele-Specific expression analysis (ASE)

We built an individual diploid genome *in silico* from individual haplotypes for each animal. For this, ALEA software (YOUNESY et al., 2014) was used using the bovine reference genome (Bos taurus UMD3.1, http://www.ensembl.org/Bos_taurus/Info/Index/) and individual genotypes obtained from the Illumina BovineHD BeadChip.

For each animal, reads were mapped to the individual diploid genome. We performed the reads mapping counting considering only the reads mapping specific to one allele of each SNP. Then, we selected SNPs for which the total reads counting was higher than 20. The binomial test was applied for each SNP of each animal to check whether the reads counting was statistical different between alleles. The null hypothesis was the half of the reads mapped in each SNP allele. Also, multiple test correction was performed by applying False Discovery Rate test (FDR, Benjamini-Hochberg ($P < 0,05$)).

Results and discussion

From 777,000 SNPs contained on BovineHD BeadChip, 447,437 SNPs passed on quality filter after genotyping. Around 3.6 billion of reads were produced in the sequencing, and 24.7 million of reads was the reads average total per sample.

Considering all SNPs heterozygous in all animals that were expressed in muscle, we performed 6,384 ASE tests. We found 519 SNPs showing at least one heterozygous animal analyzed. The results revealed that most SNPs tested here (77.45%; n = 402) presented ASE in at least one animal (n = 402) which are in agreement with the literature. Crowley *et al.* (2015) found 89% of tested genes in mouse showing ASE and Chamberlain *et al.* (2015) found from 74 to 89% in 18 bovine tissue. However, the results diverged when taking account only the analysis in leg muscle performed by Chamberlain et al. (2015). They found only 18% of analyzed SNPs showing ASE, but the proportion increased considerably when the sampling was raised to 20 sample of white blood cells and liver.

The discrepancy of ASE percentage among the results obtained by Chamberlain *et al.* (2015), when they tested one or 18 tissues, and one or 20 animals were also found in our results. We found animals showing 0.03% of tested heterozygous SNPs showing ASE meanwhile others showing 100% (n = 88 SNPs). So, the divergence of ASE among samples may be governed by specific environmental events or genotype (Chamberlain *et al.,* 2015). Therefore, as previously described by Lagarrigue *et al.* (2013), the comparison of the ASE results from one bovine muscle (Chamberlain *et al.,* 2015) with the present study in which 146 bovine muscle were analyzed, suggests the need of biological replicates in ASE studies.

The ASE observed here were well distributed throughout Nelore muscle transcriptome, with the average of 13.83 SNPs showing ASE per chromosome. Chromosome 3 showed the highest number of ASE observed (n = 31), while the opposite was found on chromosome 24 (n = 4).

ASE studies have been shown important for animal breeding programs. For example,

SNPs identified here showing ASE were identified to increase the accuracy of birds selection regarding the infection by Marek disease virus (Cheng *et al.,* 2015; Perumbakkam *et al.,* 2013). Also, ASE may contribute to gene imprinting identification, which, in turn, may contribute to increasing the animal breeding programs accuracy. For example, *IGF2R* is a well-known imprinted gene associated with growth traits (Berkowicz *et al.,* 2011). Others imprinted genes, *PEG3* and *GNAS*, were related to milk production and fertility, respectively (Magee *et al.,* 2014; Sikora *et al.,* 2011).

Despite a large number of imprinted validated genes in human and mouse, it is important to study specific genes of livestock animals since the conservation of the allelic expression pattern between primates and rodents has been lower than expected (Khatib, 2007; Monk *et al.,* 2006), and the same may be possible to occur between human and production animal.

Conclusion

The results presented here show that ASE is present and widely distributed in the bovine transcriptome. The first step was done to provide the ASE profile as a basis for significant advances in the future that will enhance information on regulatory elements.

Literature Cited

Berkowicz, E.W., Magee, D. a, Sikora, K.M., Berry, D.P., Howard, D.J., Mullen, M.P., Evans, R.D., Spillane, C., MacHugh, D.E., 2011. Single nucleotide polymorphisms at the imprinted bovine insulin-like growth factor 2 (IGF2) locus are associated with dairy performance in Irish Holstein-Friesian cattle. J. Dairy Res. 78, 1–8.

Browning, S.R., Browning, B.L., 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering 81, 1084–1097.

Chamberlain, A.J., Vander Jagt, C.J., Hayes, B.J., Khansefid, M., Marett, L.C., Millen, C.A., Nguyen, T.T.T., Goddard, M.E., 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics 16, 993.

Cheng, H.H., Perumbakkam, S., Pyrkosz, A.B., Dunn, J.R., Legarra, A., Muir, W.M., 2015. Fine mapping of QTL and genomic prediction using allele-specific expression SNPs demonstrates that the complex trait of genetic resistance to Marek's disease is predominantly determined by transcriptional regulation. BMC Genomics 16, 1–9.

Crowley, J.J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I.K., Kim, Y., Wang, J.R., Morgan, A.P., Calaway, J.D., Aylor, D.L., Yun, Z., Bell, T.A., Buus, R.J., Calaway, M.E., Didion, J.P., Gooch, T.J., Hansen, S.D., Robinson, N.N., Shaw, G.D., Spence, J.S., Quackenbush, C.R., Barrick, C.J., Nonneman, R.J., Kim, K., Xenakis, J., Xie, Y., Valdar, W., Lenarcic, A.B., Wang, W., Welsh, C.E., Fu, C.-P., Zhang, Z., Holt, J., Guo, Z., Threadgill, D.W., Tarantino, L.M., Miller, D.R., Zou, F., McMillan, L., Sullivan, P.F., Pardo-Manuel de Villena, F., 2015. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat. Genet. 47, 353–60.

Khatib, H., 2007. Is it genomic imprinting or preferential expression? BioEssays.

Lagarrigue, S., Martin, L., Hormozdiari, F., Roux, P.F., Pan, C., van Nas, A., Demeure, O., Cantor, R., Ghazalpour, A., Eskin, E., Lusis, A.J., 2013. Analysis of allele-specific expression in mouse liver by RNA-seq: A comparison with Cis-eQTL identified using genetic linkage. Genetics 195, 1157–1166.

Magee, D.A., Spillane, C., Berkowicz, E.W., Sikora, K.M., Machugh, D.E., 2014. Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. Anim. Genet. 45, 25–39.

Monk, D., Arnaud, P., Apostolidou, S., Hills, F. a, Kelsey, G., Stanier, P., Feil, R., Moore, G.E., 2006. Limited evolutionary conservation of imprinting in the human placenta. Proc. Natl. Acad. Sci. U. S. A. 103, 6623–6628.

Muráni, E., Ponsuksili, S., Srikanchai, T., Maak, S., Wimmers, K., 2009. Expression of the porcine adrenergic receptor beta 2 gene in longissimus dorsi muscle is affected by cis-regulatory DNA variation. Anim. Genet. 40, 80–89.

Perumbakkam, S., Muir, W.M., Black-Pyrkosz, A., Okimoto, R., Cheng, H.H., 2013. Comparison and contrast of genes and biological pathways responding to Marek's disease virus infection using allele-specific expression and differential expression in broiler and layer chickens. BMC Genomics 14, 64.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–75.

Ruvinsky, A., 1999. Basics of gametic imprinting. J Anim Sci 77 Suppl 2, 228–237.

Saupe, S.J., 2012. A fungal gene reinforces Mendel's laws by counteracting genetic cheating. Proc. Natl. Acad. Sci. U. S. A. 109, 11900–1.

Sikora, K.M., Magee, D. a, Berkowicz, E.W., Berry, D.P., Howard, D.J., Mullen, M.P., Evans, R.D., Machugh, D.E., Spillane, C., 2011. DNA sequence polymorphisms within the bovine guanine nucleotide-binding protein Gs subunit alpha (Gsα)-encoding (GNAS) genomic imprinting domain are associated with performance traits. BMC Genet. 12, 4.

Tuch, B.B., Laborde, R.R., Xu, X., Gu, J., Chung, C.B., Monighetti, C.K., Stanley, S.J., Olsen, K.D., Kasperbauer, J.L., Moore, E.J., Broomer, A.J., Tan, R., Brzoska, P.M., Muller, M.W., Siddiqui, A.S., Asmann, Y.W., Sun, Y., Kuersten, S., Barker, M.A., De La Vega, F.M., Smith, D.I., 2010. Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. PLoS One 5, e9317.

Tycko, B., 2010. Allele-specific DNA methylation: Beyond imprinting. Hum. Mol. Genet. 19.

Younesy, H., Moller, T., Heravi-Moussavi, A., Cheng, J.B., Costello, J.F., Lorincz, M.C., Karimi, M.M., Jones, S.J., 2014. ALEA: a toolbox for allele-specific epigenomics analysis. Bioinformatics.