Genome-wide differential DNA methylation in tropically adapted Creole cattle and their Iberian ancestors

Natalia Sevane[1]*, Rodrigo Martínez[2], Michael W. Bruford[1,3]

[1]School of Biosciences, Cardiff University, Cardiff, UK

[2]Corporación Colombiana De Investigación Agropecuaria (Corpoica), Centro de investigaciones Tibaitatá, Bogotá, Colombia

[3]Sustainable Places Research Institute, Cardiff University, Cardiff, UK

* Correspondence (NS): nsevane@ucm.com; Sevane-FernandezN@cardiff.ac.uk; Tel: +44 29 2087 5073

Summary

Enhancing climate resilience and sustainable production for animals in harsh environments are important goals for the livestock industry given the predicted impacts of climate change. Rapid adaptation to extreme climatic conditions has already been imposed on livestock species, including those exported after Columbus' arrival in the Americas. We compared the methylomes of two Creole cattle breeds living in tropical environments with their putative Spanish ancestors to understand the epigenetic mechanisms underlying rapid adaptation of a domestic species to a new and more physiologically challenging environment. Reduced representation bisulfite sequencing (RRBS) was used to assess differences in methylation in Creole and Spanish samples and revealed 334 differentially methylated regions (DMRs) using high stringency parameters (*p-value* < 0.01, ≥ 4 CpGs within a distance of 200 bp, mean methylation difference > 25%), annotated to 263 unique features. Gene ontology analysis revealed candidates involved in tropical adaptation processes, including genes differentially hyper- or hypomethylated above 80% in Creole samples displaying biological functions related to immune response (*IRF6*, *PRGDR, FAM19A5, PRLYRP1*), nervous system (*GBX2, NKX2-8, RPGR*), energy management (*BTD*), heat resistance (*CYB561*) and skin and coat attributes (*LGR6*). Our results entail that major environmental changes imposed on Creole cattle has had an impact on their methylomes measurable today, which affects genes implicated in important

pathways for adaptation. Although further work is needed, this first characterization of methylation patterns driven by profound environmental change provides a valuable pointer for the identification of biomarkers of resilience for improved cattle performance and welfare under predicted climatic change models.

Keywords *Bos taurus*, Criollo, livestock, epigenomics, RRBS

Running Head Epigenomics of tropical adaptation in Creole cattle

Introduction

Assessment of climate change impacts predicts a progressive upward trend in average temperatures over the coming century, with climatic fluctuations that may lead to a simplification of vegetation, a decrease in forage production and quality, and changes in organismal life cycles (Ciscar *et al.* 2014). For animal health, the distribution and extent of parasitic and infectious disease may increase, as natural control via low winter temperatures will be reduced. This increased disease risk and the adverse effects of extreme humidity on health may also affect temperate regions, where rainfall is predicted to increase. Moreover, stress generated by adaptation to changing conditions coupled with temperature increase may compromise immune responses to pathogens and external challenges, and lead to reduction in food intake, growth, milk yield and reproductive efficiency (Hahn 1999), jeopardizing animal welfare. Thus, improving climate resilience and enhancing sustainable production are important goals for the livestock industry. However, classical breeding programs may not provide efficient medium to long-term strategies equipped to counter the expected pace of climate change. Therefore, relying on short-term responses coupled with the ability to convey heritable phenotypic plasticity to future generations (Weyrich *et al.* 2016) could provide a better alternative for facing this imminent challenge.

Events in human history have included episodes where rapid adaptation to extreme climatic conditions have been imposed on a limited number of domestic animals. One example of such an event is Columbus' arrival in the Americas. Livestock species were brought from the Iberian

Peninsula to the Americas on Columbus' second journey in 1493 (Rodero *et al.* 1992) and spread throughout the continent, adapting to a wide range of alien environmental conditions and giving rise to 'Creole' animal populations (Rouse 1997). The total number of Iberian cattle brought to the Americas is estimated to have been less than 1,000 (Rodero *et al.* 1992). After nearly 300 years of Creole cattle expansion, several other European breeds were introduced and crossed with local populations (Willham 1982), as well as with Indian Zebu cattle, especially in tropical areas (Santiago 1978). Creole cattle were subsequently displaced into marginal, demanding environmental areas where they still occur. Examples include the Costeño con Cuernos and San Martinero breeds of Colombia, which descend from Spanish cattle and also have minor influences from Continental and/or Zebu breeds (Martínez *et al.* 2012; Ginja *et al.* 2013). These breeds are therefore the product of several centuries of adaptation to new, local and challenging environments. The Costeño con Cuernos was developed in Caribbean Colombia and tolerates high temperatures and humidity, being found from the swamp areas of Córdoba and Magdalena to the dry savannah of Sucre and Bolívar (Pinzón 1984). The San Martinero was developed in the Colombian Orinoquia region in the 17th century and is almost uniquely adapted to tropical rainforests (Holdrige & Hunter 1961).

A central goal of evolutionary biology, and an increasingly relevant one to agriculture, is to elucidate the genetic architecture of adaptation. The past decade has yielded an increasing number of examples where regulatory changes have been shown to contribute to species-specific adaptations and to reproductive isolation (Blekhman *et al.* 2008). There is mounting evidence that heritable variation in relevant traits can be generated through a suite of epigenetic mechanisms, even in the absence of genetic variation, which eventually might promote permanent changes in DNA sequence (Varriale 2014; Fagny *et al.* 2015). Among epigenetic mechanisms, DNA methylation via 5-methylcytosine is a key modification in vertebrate genomes that imparts an additional layer of heritable regulatory information upon DNA and is essential for viability in a myriad of biological processes (Lister & Ecker 2009). Epigenomic studies in cattle include muscle and placental tissues analysed with non-base-resolution methods (Su *et al.* 2014; Huang *et al.* 2014) and two recent studies using base-resolution techniques, a low coverage whole genome bisulfite sequencing (WGBS) analysis of

bovine placenta (Schroeder *et al.* 2015) and reduced representation bisulfite sequencing (RRBS) of ten bovine tissues, including blood (Zhou *et al.* 2016). However, these only described DNA methylome landscapes, not exploring environmental influences on phenotypic variation. Thus, despite increasing knowledge about the genes involved in bovine adaptation to tropical climate (Gautier *et al.* 2009; Chan *et al.* 2010; Porto-Neto *et al.* 2014; Makina *et al.* 2015; Wang *et al.* 2016; Pitt *et al.* 2018), we lack understanding of relevant epigenetic function (see Varriale 2014 for a review).

This study therefore aimed to address the role of epigenetic regulation on tropical adaptation in cattle by comparing the methylomes of modern tropical Creole bovine breeds with modern day samples from breeds including their putative Iberian ancestors. Although the number of samples analysed is relatively small, as in many similar studies (e.g. Korkmaz & Kerr 2017; Semik *et al.* 2017), we included five different breeds to establish epigenomic differentiation among groups, accounting for breed similarities related to their geographical location, i.e. the Iberian Peninsula (three breeds) and Colombia (two breeds), and used high stringency parameters to detect significant differentially methylated regions (DMRs). We generated a genome-wide map of DNA methylation at a single nucleotide resolution in cattle that provides, apart from the inherent advance in knowledge on the bovine epigenome, insights into the biology and evolution of a species under profound climate change and a base for future climate-related research in cattle.

Material and methods

Samples and DNA extraction

Five New and Old World cattle breeds were analysed in this study. The sample comprised Colombian Creole cattle Costeño con Cuernos (n = 2) and San Martinero (n = 1) breeds and Iberian cattle representing the main ancestors of these Creole populations including the Asturiana de los Valles (n = 1), Lidia (n = 1) and Retinta (n = 1) breeds. Samples were collected from adult males between 7 and 11 years old. Animals were reared in their native environment under extensive conditions with access to characteristic local vegetation available, growing under the particular climatic and dietary conditions that gave rise to the different breed

adaptations (Table 1). DNA was extracted from blood samples using the UltraClean BloodSpin DNA Isolation Kit (MO BIO Laboratories, Inc) for the Creole samples and the QIAamp DNA Blood Mini Kit (Qiagen) for the Spanish samples. The concentration and quality of genomic DNA was evaluated using the Qubit dsDNA HS Assay Kit (Life Technologies).

Reduced representation bisulfite sequencing (RRBS)

Genomic DNA (0.5-1.0 μg) from each sample was restricted with *Msp*I enzyme (New England Biosciences), cleaned using DNA Clean and concentrator-25 columns (Zymo Research), and eluted in 60 μl for library preparation. The sticky ends produced by *Msp*I digestion were filled with CG nucleotides and Illumina sequencing adapters. The TruSeq Nano DNA LT Library Prep Kit (Illumina) was used for 3'adenylation and adapter ligation. The end-repaired samples were purified using 2.5X AMpure XP Beads (Beckman Coulter) and eluted in 20 μl resuspension buffer. After adapter ligation, samples were again purified using 1.0X AMpure XP Beads and eluted in 40 μl resuspension buffer. Size-selection of DNA fragments (~175-225 bp) was performed using a 2% Agarose gel (Invitrogen), and the selected fragments were purified twice using 1.0X AMpure XP Beads and finally eluted in 22 μl resuspension buffer. Bisulfite conversion of non-methylated cytosines was performed on 20 μl size-selected fragments using the EZ DNA Methylation-Lightning Kit (Zymo Research). PCR (20 cycles) was performed to enrich the sequencing library by using a TruSeq Nano DNA LT Library Prep Kit (Illumina). The Pfu Turbo Cx Hotstart DNA polymerase (Agilent Technologies) and 10 mM dNTP mix (Life Technologies) were used for PCR reactions. After enrichment, the library was purified twice using 1X AMpure XP Beads (Beckman Coulter) and quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). The average library size was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). The libraries were then pooled in equimolar ratios of 2 nM, and 6.0 pM of the pool was clustered and sample tracked using the cBot (Illumina) and sequenced following a 2 x 150 bp protocol for 300 cycles using the HiSeq 2500 system (Illumina).

RRBS data analysis and genome-wide DNA methylation levels

Quality assessment and control was performed using the Trim Galore software (Babraham bioinformatics, UK). For adapter trimming the minimum required adapter overlap was 1 bp. To remove potential methylation-biased bases from the *Msp*I digestion end-repair reaction, RRBS reads were trimmed a further 2 bp when adapter contamination was detected and by 2 bp at the start when read started with CAA or CGA. Trimming was performed on all reads using a minimum Phred quality score of 20. Sequences were mapped with single end mapping to the bovine genome assembly UMD3.1.1 using Bismark (Krueger & Andrews 2011). Following optimisation, a seed length of 20 bp was chosen and only one mismatch was allowed. The minimum alignment score function was set at L,0,-0.6. Only the reads that were aligned to a unique region in the genome were used for further analysis.

For CpG level comparison, percent methylation of individual CpGs was calculated using MethylKit package in R (Akalin *et al.* 2012) and the coverage files from Bismark aligner. To prevent PCR bias and increase the power of the statistical tests we discarded bases with high (above 99.9th percentile of coverage in each sample) and low (below 10X coverage, CpG_{10}) read coverage. Each sequenced and filtered CpG_{10} site was assigned a percentage methylation score. The CpG_{10} bisulfite conversion rate was calculated as the number of thymines (non-methylated cytosines) divided by coverage for each non-CpG cytosine as implemented in MethylKit. Coverage and correlation plots were generated also by MethylKit. The pattern of methylation around different components of the cattle genome, including gene bodies (defined as the region from transcription start site -TSS- to transcription termination site -TTS-), TSSs, TTSs, and CpG islands (CpGI), was also investigated using the Seqmonk software (Babraham bioinformatics, UK), from 20 kb upstream to 20 kb downstream. CpG_{10} were annotated with the closest/overlapping TSS (±100 kb) (Miele & Dekker 2008; Sanyal *et al.* 2012) and CpGI using identgenloc program from the DMAP package (Stockwell *et al.* 2014). Promoters were defined as -0-2 kb of TSS, and CpGI shores and shelves as ±0-2 kb and ±2-4 kb flanking regions of CpGI, respectively.

Differentially methylated region (DMR) analysis

DMRs were established among Creole and Spanish groups to account for breed similarities related to their geographical location, i.e. the Iberian Peninsula (three breeds, three samples = three biological replicates within the same group) and Colombia (two breeds, three samples = three biological replicates within the same group). To compare spatially contiguous stretches of methylated cytosines across the Creole and the Spanish genomes, DMRs were determined using the R package dispersion shrinkage for sequencing data (DSS) (Feng *et al.* 2014), which outperforms other methods when sample size per group is small owing to the adoption of Wald test with shrinkage for determining differentially methylated cytosines (DMC) (Zhang *et al.* 2016). We identified DMRs using the coverage files from Bismark and the callDMR function with a *p*-value threshold of 0.01, delta=0.1 and otherwise default parameters. To be considered significant, a DMR was required to contain at least 3 CpG sites (default parameter, although the smallest significant DMR included 4 CpGs) within a distance of 200 bp, and with an absolute mean methylation difference greater than 25% when comparing Creole and Spanish samples (Akalin *et al.* 2012). As CpG_{10}, DMRs were annotated with the closest/overlapping TSS (±100 kb) and CpGI using the identgenloc program from the DMAP package.

Gene ontology (GO) analysis

Annotated DMRs were subjected to GO enrichment using the PANTHER v.10 web resource (Mi *et al.* 2016). This GO classification system was used to assign putative function to each gene by way of biological process, molecular function and cellular components. The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang *et al.* 2009) was used to determine processes of major biological significance through the Functional Annotation Cluster (FAC) tool based on the GO annotation function. High stringency ease score parameters were selected to obtain confident enrichment scores. KEGG pathway analyses were performed using both DAVID and the WebGeStalt overrepresentation enrichment analysis (ORA) (Wang et al. 2013) to map clusters of genes involved in common pathways and processes.

Validation of RRBS data with HiSeq bisulfite sequencing PCR (HiSeq-BSP)

We performed validation of RRBS data with HiSeq-BSP for three DMRs annotated to immune, cancer and nervous system genes, displaying both hyper- and hypomethylation patterns (Table S1). The initial concentration of genomic DNA was measured using the Qubit dsDNA HS Assay Kit (Life Technologies). The samples were then diluted accordingly to achieve the recommended DNA input of 500 ng at a concentration of 25 ng/µL for bisulfite treatment. The samples were bisulfite-treated using EZ DNA Methylation-Lightning Kit (Zymo Research). The treated DNA was PCR-amplified using specific primers for BSP designed and validated by Zymo Research (Table S1). The amplified product for these three assays were pooled together for each sample and sequencing libraries were made by using TruSeq Nano DNA LT Library Prep Kit (Illumina). Following the library preparation, the final concentration of the library was measured using the Qubit dsDNA HS Assay Kit (Life Technologies). The libraries were diluted to 12 pM and were sequenced by using the 600 Cycles v3 Reagent Kit (Illumina) on the MiSeq (Illumina) on a 150-base paired-end run. Sequence reads were trimmed, aligned and analysed as described above.

Results

Assessment of RRBS data and genome-wide DNA methylation levels

Fragmentation with the restriction enzyme *Msp*I of blood-extracted DNA from three Creole and three Spanish samples resulted in high quality sequencing RRBS libraries enriched for high CG regions. Illumina HiSeq 2500 sequencing generated between 15 and 38 million reads per sample (accession number GSE101796) and a total of 136 million reads (Table S2). Quality control analysis using Trim Galore and MethylKit indicated that the 150 bp sequences displayed the expected nucleotide composition based on *Msp*I digestion and bisulfite conversion (98% average sodium bisulfite conversion efficiency). On average, 98.2% of reads passed the filtering process (Table S2). The mean percentage of mapped reads was 85%, with 33-61% of reads mapping to multiple locations of the genome and 29-43% mapping uniquely (Table S2). Sequences that did not map, or did not map uniquely, were excluded from the analysis. After alignment, we filtered the CpG dinucleotides based on a coverage of 10 or more reads (CpG_{10}). The number of CpG_{10} per sample ranged from 0.4 to 1.6 million, and the mean coverage from 33 to 106 (Table S3). Of these sites, 20,234 were present in all six samples (Table

S4). We observed high positive correlations between all the samples analysed (mean Pearson's correlation coefficient = 0.8), although clear variation was present between them (Figure S1). The distribution of sequence read coverage of CpG_{10} per sample is shown in Figure S2, and highlights that despite the observation that the filtered CpG_{10} displayed high mean coverage, the libraries did not suffer from bias due to excessive amplification of a subset of fragments, as reflected in the absence of peaks on the right-hand side of each histogram. The RRBS protocol has been shown to enrich for CpGIs and, as CpGIs have been universally reported to be regions of gene regulation via methylcytosine and are generally demethylated, the percent methylation of CpGs in RRBS libraries is expected to be lower than the average methylation of the genome (~80%). Accordingly, global CpG_{10} methylation ranged from 51 to 57% across samples (Table S3). The distribution of methylation at each CpG_{10} site revealed a bimodal pattern, with heavy methylation (>95%) of 39 to 53% CpG_{10} and completely unmethylated bases (<5%) ranging between 35% and 47% (Figure S3). The median methylation was high (84%) (Table S3), reflecting the heavy hypermethylation of 48% of CpG_{10} sites. However, hypomethylated CpG_{10} sites were also evident, including 42% of the analysed CpG sites (Figure S3). The percentage of CHG and CHH methylation was low in cattle blood, ranging from 0.9% to 1.6% (Table S3). RRBS reads were detected in most chromosomal regions (chromosomes 1–29 and X) in each group, although some gaps existed (Figure 1). This even read distribution indicated that cattle blood methylomes can be detected by RRBS technology with good representation, thereby ensuring accurate examination of variation in DNA methylation. The distribution of CpG_{10} related to CpGIs (6,235, 31%) revealed that RRBS data is highly enriched in CpGI cores (85%), while only a small amount is in CpGI shores (11%) and shelves (4%) (Table S4). The distribution of CpG_{10} in relation to genes (4,389, 22%) showed that almost 85% mapped to gene bodies (92% located in introns and 8% in exons) and a much smaller percentage mapped to gene promoters (15%), with the main amount located in intergenic regions (15,845, 78%) (Table S4). DNA methylation levels sharply decreased in the 2 kb region upstream of TSSs and dropped to the lowest point before TSSs (Figure 2A), corresponding with the distribution of gene promoters, usually prone to transcription, whereas levels dramatically increased in the 3' direction, peaking 5' to the TTS (Figure 2A), related with the methylation of gene bodies

contributing to chromatin structure stability and the regulation of gene expression (Bird 2002). The level dropped slightly and plateaued after TTS (Figure 2A). As expected, the level of methylation in CpGIs was lower than outside CpGIs (Figure 2B).

Differential methylation between Creole and Spanish cattle samples

Comparison between spatially contiguous stretches of DMCs from Creole and Spanish samples revealed 334 DMRs (*p-value* < 0.01, ≥ 4 CpGs within a distance of 200 bp, mean methylation difference > 25%, Table S5). Annotation of these DMRs showed that 275 sites (82%), corresponding to 263 unique features, were overlapping a gene or within a distance of ±100 kb from the closest TSS. Approximately 37% of DMRs overlapped a gene, while ~4% were in regions 2 kb upstream of TSS or promoters. Intragenic DMR were equally divided between introns (52%) and exons (48%). Around 36% of DMRs were located in CpGIs, mainly in CpGI cores (81%), whereas only 12% and 7% overlapped CpGIs shores and shelves, respectively. Interestingly, a high proportion of DMRs (71%) displayed hypermethylation in Creole samples. Table 2 shows the DMRs overlapping a gene or CpGI hyper- and hypomethylated above 80% in Creole samples.

Gene ontology (GO) analysis

Among the 263 differentially methylated unique annotated features, functional data for 213 genes was obtained with PANTHER, including the GO classes molecular function (the primary activities of gene products at the molecular level), biological process (sets of molecular events or operations with a defined beginning and end) and cellular component (Figure S4). The annotated DMRs were then analysed using DAVID and WebGeStalt tools. DAVID FAC analysis produced 16 enriched functional clusters under high stringency conditions for 115 DAVID IDs (Table S6). Among these enriched functional clusters, homeobox, epidermal growth factor (two clusters) and immunoglobulin (two clusters) were identified. We analysed the distribution of annotated DMRs along the cattle chromosomes, confirming one enriched genomic region in chromosome 21 that comprised 12 genes related to the cellular component membrane (Table S6). KEGG pathway analysis retrieved a total of 14 pathways (Table 3): ten from WebGeStalt,

288 including immune related processes such as leukocyte, T cell and lymphocyte differentiation
289 and activation, and circulatory system development or cell proliferation (Figure 3); and four
290 from DAVID tool –acute myeloid leukemia, insulin signalling pathway, Rap1 signaling pathway,
291 microRNAs in cancer (Figure S5).
292
293 *Validation of RRBS data with HiSeq bisulfite sequencing PCR (HiSeq-BSP)*
294 We used HiSeq-BSP to assess the methylation patterns of three gene annotated DMRs,
295 including regions displaying high and low differential methylation levels between Creole and
296 Spanish samples and implicated in immune (*SERPINB1*), cancer (*SHOX2*) and nervous system
297 (*NRXN2*) processes. The HiSeq-BSP methylation results were significant for the three amplified
298 regions (*p-value* < 0.01, ≥ 4 CpGs, mean methylation difference ≥ 10%) and concordant with the
299 methylation profiles obtained with the RRBS analysis (Table S7).
300
301 **Discussion**
302 Studies on adaptation are key to disentangling the evolutionary potential of organisms in
303 response to biotic and abiotic stress and other environmental challenges, which could
304 potentially be highly relevant in the context of global climate change. Tropical environments
305 are characterized by high temperature and humidity, episodes of feed and water scarcity and
306 virulent tropical diseases and parasite infections. Creole cattle demonstrate greater resistance
307 to such conditions, surviving, breeding and producing efficiently in the tropics (Hernández-
308 Cerón *et al.* 2004; Martínez *et al.* 2008). Two tropically adapted Creole breeds and their likely
309 Spanish ancestors were analysed to establish epigenomic differences among groups accounting
310 for breed similarities related to their geographical location. The Costeño con Cuernos and San
311 Martinero breeds have been developed under physiologically challenging tropical conditions.
312 The Iberian breed Retinta is distributed throughout central and southern Iberia, which is
313 characterized by a xeric climate. The Asturiana de los Valles breed reflects the northern Iberian
314 gene-pool and is exposed to a milder climate, mostly cold and damp. The Lidia breed (Spanish
315 fighting bull) has not been selected for productivity traits and thus may be the most
316 representative modern descendent of Iberian cattle herds back in the 15th century. We

317 detected 334 highly significant DMRs between the groups. The methylation profiles obtained
318 were consistent with previous studies (e.g. Zhou *et al.* 2016). High stringency parameters to
319 detect DMRs (*p-value* < 0.01, ≥ 4 CpGs within a distance of 200 pb, mean methylation
320 difference > 25%) when compared with other studies (e.g. Gao *et al.* 2014; Day *et al.* 2016;
321 Shankar *et al.* 2015; Baerwald *et al.* 2016), were taken as statistically significant to overcome
322 the relatively small number of biological replicates characteristic of many epigenomic
323 experiments (e.g. Miele *et al.* 2008; Zhou *et al.* 2016; Semik *et al.* 2017).
324 In concordance with previous studies on bovine adaptation to tropical climates, including both
325 taurine and indicine (Gautier *et al.* 2009; Chan *et al.* 2010; Porto-Neto *et al.* 2014; Makina *et al.*
326 2015; Wang *et al.* 2016; Pitt *et al.* 2018), we found a number of differentially methylated genes
327 between Creole and Spanish groups implicated in several biological processes key for survival in
328 harsh environments, such as immunity, nervous system processes, energy management, heat
329 resistance and skin and coat attributes (Table S5).
330 Tropical cattle carry lower burdens of ticks, have enhanced disease resistance and superior
331 innate immunity, which is reflected in a higher number of genes under selection related to the
332 immune system in studies on adaptation to tropical conditions (e.g. Amorim *et al.* 2015; Liu *et*
333 *al.* 2018; Pitt *et al.* 2018). Some of the genes hypermethylated above 80% in Creole samples are
334 implicated in immune processes (Table 2): i) *IRF6,* involved in inflammatory responses,
335 macrophage activation and dysregulation of metabolic and immunologic homeostasis (Li *et al.*
336 2017); ii) *PRGDR,* which plays an important role in the immune response found in allergic
337 diseases, apart from facilitating smooth muscle relaxation and vasodilatation, inhibiting platelet
338 aggregation and contributing to the regulation of pain perception and sleep (Pettipher *et al.*
339 2007); and iii) *FAM19A5,* a brain-specific chemokine or neurokine that acts as regulator of
340 immune and nervous cells (Tom Tang *et al.* 2004). On the contrary, a DMR was found
341 hypomethylated above 80% in Creole samples in a CpGI core located inside an exon of the
342 *PRLYRP1* gene. The protein encoded by this gene has been reported to interact with microbes
343 to maintain intestinal homeostasis (Seabury *et al.* 2010) and has been associated with
344 resistance to *Mycobacterium avium ssp. paratuberculosis* (Pant *et al.* 2011), both in cattle. This
345 gene is also associated with several health, reproduction and body conformation traits in

Holstein cows (Cole *et al.* 2011). The ability to cope with parasitic and infectious diseases in the adaptation to new environments also seems relevant at the multi-genic level given the high enrichment of pathways such as regulation of leukocyte differentiation, T cell activation, leukocyte cell-cell adhesion or lymphocyte activation (Figure 3, Table 3), and the presence of two enriched functional clusters related to immunoglobulins (Table S6). These findings are also in agreement with the work of Fagny *et al.* (2015) that describes the existence of epigenetic variability on immune processes implicated in the adaptation to changes in habitat and lifestyle in humans.

Nervous system processes, including changes in behaviour, circadian clock, olfactory and eye function or chemosensory perception, are key for animals to adapt to new light, food, reproduction or predatory conditions. Genes with roles in nervous system processes hypermethylated above 80% in Creole samples include (Table 2): i) *GBX2*, modulator of thalamus cells development (Mallika *et al.* 2015); ii) *NKX2-8*, a regional homeobox gene with functions in neuronal development (Safra *et al.* 2013) as well as in tumor suppression; and iii) *FAM19A5* (see above). Another gene showing the same methylation pattern and implicated in eye function is *RPGR*. The protein encoded by this gene localizes to the outer of rod photoreceptors and is crucial for their viability, its deficiency causing X-linked retinitis pigmentosa (Lyraki *et al.* 2016).

The efficient management of energy storage and mobilization during wet and dry seasons, respectively, provides a greater ability to tolerate poor feed in harsh environments (Amorim *et al.* 2015). Methylation differences in the insulin signalling pathway (Table 3) or genes such as *DAGLA* (diacylglycerol lipase alpha), *FADS2* (fatty acid desaturase 2) or *LMF1* (lipase maturation factor 1) (Table S5), may determine variations in energy metabolism. A gene hypermethylated above 80% in Creole samples is *BTD*, whose protein catalyses the recycling of biotin from biocytin (Wolf 2012) (Table 2). Biotin is a member of the B Vitamin group and is an essential nutrient in the formation of keratin, as well as for gluconeogenesis, lipogenesis and protein synthesis. Biotin treatments have been reported to have beneficial effects on milk production, hoof health and reproduction traits (Wilde 2006; Lean & Rabiee 2011).

374 Genes involved in cardiovascular physiology can facilitate heat resistance in tropical climates.
375 We found a mean hypermethylation level of above 80% in Creole samples for the *PTGDR* gene,
376 which facilitates smooth muscle relaxation and vasodilatation (Pettipher *et al.* 2007), and the
377 *CYB561* gene, influencing cardiovascular responses to sympathetic activation (Fung *et al.* 2008),
378 as well as the high enrichment of the circulatory system development pathway (Table 2, Table
379 3, Figure 3). Skin and coat attributes are also important for adaptation to harsh conditions, with
380 a direct influence on the thermo-resistance to tropical conditions. The *IRF6* gene, also
381 implicated in immune homeostasis, promotes epithelial cell proliferation and differentiation
382 (Richardson *et al.* 2006) and was hypermethylated above 80% in the Creole group (Table 2). On
383 the contrary, *LGR6* gene, which establishes sebaceous glands and interfollicular epidermis
384 postnatally (Snippert *et al.* 2010), showed hypomethylation above 80% in Creole samples. In
385 addition, two enriched functional clusters related to epidermal growth factor were identified
386 with the DAVID FAC analysis (Table S6). Four microRNAs, regulation of which at the 3'
387 untranslated region plays important roles in the modulation of gene expression (Su *et al.* 2011),
388 were also differentially methylated between Creole and Spanish cattle groups.
389 Some cancers, especially in young animals, might be a by-product of novel adaptation and have
390 their origins in recent evolutionary changes in morphology and life-history, driving evolution of
391 many features of cellular behaviour and regulation (Leroi *et al.* 2003). Concordantly, rapid bouts
392 of evolution, such as artificial selection in domestic species, have been shown to make animals
393 prone to different cancers (Leroi *et al.* 2003). Epigenetic changes, especially DNA methylation,
394 alter signal-transduction pathways during the early stages of tumor development. Tumor cells,
395 opposed to normal cells, show local hypermethylation of some CpGI combined with global
396 genome demethylation (Bernstein *et al.* 2007). Taking into account that RRBS enriches for GC-
397 rich regions such as CpGIs (Laird 2010), the high proportion of hypermethylated DMRs in Creole
398 samples (71%), along with the high number of genes differentially hyper- or hypomethylated
399 above 80% in these samples and related to oncogenic processes (in particular, *NKX2-8*, *LATS2*,
400 *BRAT1*, *BLM*, *TP53I11* also known as *PIG11*, *TM4SF5*, *TRIM25*, *LGR6*) (Table 2), as well as the
401 high enrichment of several pathways implicated in cancer (acute myeloid leukemia, Rap1
402 signaling pathway, microRNAs in cancer) (Table 3, Figure S5), might reflect an on-going

403 adaptation process to tropical conditions on the descendants from the cattle brought from
404 Iberia to Colombia.
405 Three regions were chosen to verify RRBS methylation levels including both hyper and hypo-
406 methylated DMRs and genes related to the main biological processes immunity (*SERPINB1*),
407 cancer (*SHOX2*) and nervous system (*NRXN2*). Levels of methylation were higher overall in the
408 RRBS dataset than obtained by HiSeq-BSP, but the observed direction and tendency of changes
409 were consistent for all the regions under analysis (Table S7). The difference in magnitude
410 between RRBS and HiSeq-BSP results may have been due to the lower bisulfite conversion rate
411 obtained for the HiSeq-BSP protocol (89%) compared with RRBS (98%) and/or PCR bias.
412 Although DNA methylation is universally associated with gene expression silencing (Bird 2002),
413 the complex gene and pathway connections coupled with the long-range interactions of
414 regulatory elements that cannot simply be predicted by genomic proximity (Miele & Dekker
415 2008; Sanyal et al. 2012), hinders the extrapolation of epigenomic and genomic factors, along
416 with environmental influences, to phenotypic transitions. Moreover, RRBS only covers a small
417 fraction of the genome and cellular heterogeneity is a major challenge when comparing DNA
418 methylation across samples. Blood samples consist of a mixture of immune cells in varying
419 proportions with unique methylation profiles that may have hindered the ability to detect
420 DMRs (Reinius *et al.* 2012). The presence of C/T SNPs at CpGs may be also a confounding
421 variable, especially when methylation levels are compared among individuals from genetically
422 nonhomogeneous populations (Daca-Roszak et al. 2015). Thus, these results should be
423 considered as a preliminary survey, highlighting the need for additional epigenomic studies on a
424 wider sample set under more standardized conditions, using more extensive techniques like
425 WGBS and strengthening the connection between epigenomic and phenotypic variability by
426 integrating also genomic and gene expression data sets.
427
428 In conclusion, we characterize differential methylation patterns between tropically adapted
429 bovine breeds and their main ancestors for the first time, and show that challenging climate
430 and environmental factors imposed on a reduced number of animals had an impact on their
431 methylome pattern still measurable today, affecting genes implicated in important signalling

pathways for adaptation and pointing towards the epigenetic fine-tuning on the regulation of gene activity. The comparison between both sample groups identified DMRs annotated to genes directly or indirectly involved in tropical adaptation processes, such as immunity, nervous system processes, energy management, heat resistance and skin and coat attributes. The ability of epigenetic changes to provide an initial rapid and flexible response to environmental challenges, makes epigenetic studies a promising field to uncover alternative mechanisms driven evolution of adaptive phenotypes, eventually generating permanent genetic changes. The DMRs detected in this study, along with the tissue analysed, blood, that is easily accessible and reflects the immune status of individuals, provide a valuable starting point for the identification of epigenetic biomarkers of resilience for improved cattle performance and welfare under predicted climatic change models.

Data availability

The data sets supporting the results of this article were deposited in the Gene Expression Omnibus (GEO) with accession GSE101796.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

We thank Prof. J. Cañón and Prof. S. Dunner from Universidad Complutense de Madrid (UCM) for their contribution of the Spanish samples included in the study. We also thank M. Barbato and P. Orozco-terWengel from Cardiff University and M. Montes from UCM Genetic Service for their collaboration in data analysis. NS is a recipient of a Marie Skłodowska-Curie Individual Fellowship and this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No DLV-655100. MWB is funded by BBSRC through the FACCE-JPI ERA-NET Climate Smart Agriculture project CLIMGEN (BB/M019276/1).

461 **Refereces**

462 Akalin A., Kormaksson M., Li S., Garrett-Bakelmann F.E., Figeuroa M.E., Melnick A. & Mason CE. (2012)
463 methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles.
464 Genome Biology 13, R87.

465 Amorim C.E., Daub J.T., Salzano F.M., Foll M. & Excoffier L. (2015) Detection of convergent genome-wide
466 signals of adaptation to tropical forests in humans. PLoS One 10, e0121557.

467 Baerwald M.R., Meek M.H., Stephens M.R., Nagarajan R.P., Goodbla A.M., Tomalty K.M.H., Thorgaard
468 G.H., May B. & Nichols K.M. (2016) Migration-related phenotypic divergence is associated with
469 epigenetic modifications in rainbow trout. Molecular Ecology 25, 1785-1800.

470 Bernstein B.E., Meissner A. & Lander E.S. (2007) The mammalian epigenome. Cell 128, 669-681.

471 Bird A. (2002) DNA methylation patterns and epigenetic memory. Genes & Development 16, 6-21.

472 Blekhman R., Oshlack A., Chabot A.E., Smyth G.K. & Gilad Y. (2008) Gene regulation in primates evolves
473 under tissue-specific selection pressures. PLoS Genetics 4, e1000271.

474 Chan E.K., Nagaraj S.H. & Reverter A. (2010) The evolution of tropical adaptation: comparing taurine and
475 zebu cattle. Animal Genetics 41, 467-477.

476 Ciscar J.C., Feyen L., Soria A,. et al. (2014) Climate Impacts in Europe. The JRC PESETA II Project. JRC
477 Scientific and Policy Reports; 2014. EUR 26586EN.

478 Cole J.B., Wiggans G.R., Ma L., Sonstegard T.S., Lawlor T.J. Jr., Crooker B.A., Van Tassell C.P., Yang J.,
479 Wang S., Matukumalli L.K. & Da Y. (2011) Genome-wide association analysis of thirty one
480 production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows.
481 BMC Genomics 12, 408.

482 Daca-Roszak P., Pfeifer A., Żebracka-Gala J., Rusinek D., Szybińska A., Jarząb B., Witt M. & Ziętkiewicz E.
483 (2015) Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450
484 BeadChip Array: implications for comparative population studies. BMC Genomics 16, 1003.

485 Day S.E., Coletta R.L., Kim J.Y., Campbell L.E., Benjamin T.R., Roust L.R., De Filippis E.A., Dinu V., Shaibi
486 G.Q., Mandarino L.J. & Coletta D.K. (2016) Next-generation sequencing methylation profiling of
487 subjects with obesity identifies novel gene changes. Clinical Epigenetics 8,77.

488 Fagny M., Patin E., MacIsaac J.L., et al. (2015) The epigenomic landscape of African rainforest hunter-
489 gatherers and farmers. Nature Communications 6, 10047.

490 Feng H., Conneely K.N. & Wu H. (2014) A Bayesian hierarchical model to detect differentially methylated
491 loci from single nucleotide resolution sequencing data. Nucleic Acids Research 42, e69.

492 Fung M.M., Nguyen C., Mehtani P., Salem R.M., Perez B., Thomas B., Das M., Schork N.J., Mahata S.K.,
493 Ziegler M.G. & O'Connor D.T. (2008) Genetic variation within adrenergic pathways determines in
494 vivo effects of presynaptic stimulation in humans. Circulation 117, 517-525.
495 Gao F., Zhang J., Jiang P., Gong D., Wang J.W., Xia Y., Østergaard M.V., Wang J. & Sangild P.T. (2014)
496 Marked methylation changes in intestinal genes during the perinatal period of preterm neonates.
497 BMC Genomics 15, 716.
498 Gautier M., Flori L., Riebler A., Jaffrézic F., Laloé D., Gut I., Moazami-Goudarzi K. & Foulley J.L. (2009) A
499 whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics
500 10, 550.
501 Ginja C., Gama L.T., Cortes O., Delgado J.V., Dunner S., García D., Landi V., Martín-Burriel I., Martínez-
502 Martínez A., Penedo M.C., Rodellar C., Zaragoza P., Cañon J. & BioBovis Consortium. (2013) Analysis
503 of conservation priorities of Iberoamerican cattle based on autosomal microsatellite markers.
504 Genetics Selection Evolution 30, 45-35.
505 Hahn G.L. (1999) Dynamic responses of cattle to thermal heat loads. Journal of Animal Science 77, 10-
506 20.
507 Hernández-Cerón J., Chase C.C. & Hansen P.J. (2004) Differences in sensitivity to heat-shock between
508 preimplantation embryos from heat-tolerant (Brahman and Romosinuano) and heat-sensitive
509 (Angus) breeds. Journal of Dairy Science 87, 53-58.
510 Holdrige L.R. & Hunter J.R. (1961) Clave de las Asociaciones Climáticas del mundo y guía para el uso de la
511 tierra en los trópicos. Suplemento de la Revista Académica Colombiana de Ciencias 11, 14.
512 Huang D.W., Sherman B.T. & Lempicki R.A. (2009) Systematic and integrative analysis of large gene lists
513 using DAVID Bioinformatics Resources. Nature Protocols 4, 44-57.
514 Huang Y.Z., Sun J.J., Zhang L.Z., Li C.J., Womack J.E., Li Z.J., Lan X.Y., Lei C.Z., Zhang C.L., Zhao X. & Chen
515 H. (2014) Genome-wide DNA methylation profiles and their relationships with mRNA and the
516 microRNA transcriptome in bovine muscle tissue (Bos taurine). Scientific Reports 4, 6546.
517 Korkmaz F.T. & Kerr D.E. (2017) Genome-wide methylation analysis reveals differentially methylated loci
518 that are associated with an age-dependent increase in bovine fibroblast response to LPS. BMC
519 Genomics 18, 405.
520 Krueger F. & Andrews S.R. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
521 applications. Bioinformatics 27, 1571-1572.
522 Laird P.W. (2010) Principles and challenges of genomewide DNA methylation analysis. Nature Reviews
523 Genetics 11, 191-203.

524 Lean I.J. & Rabiee A.R. (2011) Effect of feeding biotin on milk production and hoof health in lactating
525 dairy cows: a quantitative assessment. Journal of Dairy Science 94, 1465-1476.
526 Leroi A.M., Koufopanou V. & Burt A. (2003) Cancer selection. Nature Reviews Cancer 3, 226-231.
527 Li C., Ying W., Huang Z., Brehm T., Morin A., Vella A.T. & Zhou B. (2017) IRF6 regulates alternative
528 activation by suppressing PPARγ in male murine macrophages. Endocrinology 158, 2837-2847.
529 Lister R. & Ecker J.R. (2009) Finding the fifth base: Genome-wide sequencing of cytosine methylation.
530 Genome Research 19, 959-966.
531 Liu Y-H., Wang L. & Xu T. (2018) Whole-Genome Sequencing of African Dogs Provides Insights into
532 Adaptations against Tropical Parasites. Molecular Biology and Evolution 35, 287-298.
533 Lyraki R., Megaw R. & Hurd T. (2016) Disease mechanisms of X-linked retinitis pigmentosa due to RP2
534 and RPGR mutations. Biochemical Society Transactions 44, 1235-1244.
535 Makina S.O., Muchadeyi F.C., van Marle-Köster E., Taylor J.F., Makgahlela M.L. & Maiwashe A. (2015)
536 Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genetics Selection
537 Evolution 47, 92.
538 Mallika C., Guo Q. & Li J.Y. (2015) Gbx2 is essential for maintaining thalamic neuron identity and
539 repressing habenular characters in the developing thalamus. Developmental Biology 407, 26-39.
540 Martínez A.M., Gama L.T., Cañón J., et al. (2012) Genetic Footprints of Iberian Cattle in America 500
541 Years after the Arrival of Columbus. PLoS One 7, e49066.
542 Martínez R.A., García D., Gallego J.L., Onofre G., Pérez J. & Cañón J. (2008) Genetic variability in
543 Colombian Creole cattle populations estimated by pedigree information. Journal of Animal Science
544 86, 545-552.
545 Mi H., Poudel S., Muruganujan A., Casagrande J.T. (2016) Thomas PD. PANTHER version 10: expanded
546 protein families and functions, and analysis tools. Nucleic Acids Research 44, D336-342.
547 Miele A. & Dekker J. (2008) Long-range chromosomal interactions and gene regulation. Molecular
548 Biosystems 4, 1046-1057.
549 Pant S.D., Verschoor C.P., Schenkel F.S., You Q., Kelton D.F. & Karrow N.A. (2011) Bovine PGLYRP1
550 polymorphisms and their association with resistance to Mycobacterium avium ssp. paratuberculosis.
551 Animal Genetics 42, 354-360.
552 Pettipher R., Hansel T.T. & Armer R. (2007) Antagonism of the prostaglandin D2 receptors DP1 and
553 CRTH2 as an approach to treat allergic diseases. Nature Reviews Drug Discovery 6, 313-325.
554 Pinzón M.E. (1984) Historia de la ganadería bovina en Colombia. Suplemento ganadero. Banco
555 Ganadero. Santafé de Bogotá Colombia 4, (1):208.

556 Pitt D., Bruford M.W., Barbato M., Orozco-terWengel P., Martínez R. & Sevane N. (2018) Demography
557 and rapid local adaptation shape Creole cattle genome diversity in the tropics. Evolutionary
558 Applications, https://doi.org/10.1111/eva.12641
559 Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F.,
560 Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A. & Barendse
561 W. (2014) The genetic architecture of climatic adaptation of tropical cattle. PLoS One 9, e113284.
562 Reinius L.E., Acevedo N., Joerink M., Pershagen G., Dahlén S.E., Greco D., Söderhäll C., Scheynius A. &
563 Kere J. (2012) Differential DNA methylation in purified human blood cells: implications for cell
564 lineage and studies on disease susceptibility. PLoS ONE 7, e41361.
565 Richardson R.J., Dixon J., Malhotra S., Hardman M.J., Knowles L., Boot-Handford R.P., Shore P.,
566 Whitmarsh A. & Dixon M.J. (2006) Irf6 is a key determinant of the keratinocyte proliferation-
567 differentiation switch. Nature Genetics 38, 1329-1334.
568 Rodero E., Rodero A. & Delgado J.V. (1992) Primitive Andalusian livestock and their implications in the
569 discovery of America. Archivos de Zootecnia 41, 383-400.
570 Rouse J.E. (1997) The Criollo: Spanish cattle in the Americas. Norman (OK): University of Oklahoma
571 Press, Oklahoma.
572 Safra N., Bassuk A.G., Ferguson P.J., Aguilar M., Coulson R.L., Thomas N., Hitchens P.L., Dickinson P.J.,
573 Vernau K.M., Wolf Z.T. & Bannasch D.L. (2013) Genome-wide association mapping in dogs enables
574 identification of the homeobox gene, NKX2-8, as a genetic component of neural tube defects in
575 humans. PLoS Genetics 9, e1003646.
576 Santiago A.A. (1978) Evolution of Zebu cattle in Brazil. The Zebu Journal 1, 6.
577 Sanyal A., Lajoie B.R., Jain G. & Dekker J. (2012) The long-range interaction landscape of gene promoters.
578 Nature 489, 109-413.
579 Schroeder D.I., Jayashankar K., Douglas K.C., Thirkill T.L., York D. & Dickinson P.J. (2015) Early
580 developmental and evolutionary origins of gene body DNA methylation patterns in mammalian
581 placentas. PLoS Genetics 11, e1005442.
582 Seabury C.M., Seabury P.M., Decker J.E., Schnabel R.D., Taylor J.F. & Womack J.E. (2010) Diversity and
583 evolution of 11 innate immune genes in Bos taurus taurus and Bos taurus indicus cattle. Proceedings
584 of the National Academy of Sciences USA 107, 151-156.
585 Semik E., Ząbek T., Gurgul A., Fornal A., Szmatoła T., Pawlina K., Wnuk M., Klukowska-Rötzler J., Koch C.,
586 Mählmann K. & Bugno-Poniewierska M. (2017) Comparative analysis of DNA methylation patterns of
587 equine sarcoid and healthy skin samples. Veterinary and Comparative Oncology 16, 37-46.

588 Shankar K., Kang P., Zhong Y., Borengasser S.J., Wingfield C., Saben J., Gomez-Acevedo H. & Thakali K.M.
589 (2015) Transcriptomic and epigenomic landscapes during cell fusion in BeWo trophoblast cells.
590 Placenta 36, 1342-1351.
591 Snippert H.J., Haegebarth A., Kasper M., Jaks V., van Es J.H., Barker N., van de Wetering M., van den
592 Born M., Begthel H., Vries R.G., Stange D.E., Toftgård R. & Clevers H. (2010) Lgr6 marks stem cells in
593 the hair follicle that generate all cell lineages of the skin. Science 327, 1385-1389.
594 Stockwell P.A., Chatterjee A., Rodger E.J. & Morison I.M. (2014) DMAP: differential methylation analysis
595 package for RRBS and WGBS data. Bioinformatics 30, 1814-022.
596 Su J., Wang Y., Xing X., Liu J. & Zhang Y. (2014) Genome-wide analysis of DNA methylation in bovine
597 placentas. BMC Genomics 15, 12.
598 Su Z., Xia J. & Zhao Z. (2011) Functional complementation between transcriptional methylation
599 regulation and post-transcriptional microRNA regulation in the human genome. BMC Genomics 12,
600 S15.
601 Tom Tang Y., Emtage P., Funk W.D., Hu T., Arterburn M., Park E.E. & Rupp F. (2004) TAFA: a novel
602 secreted family with conserved cysteine residues and restricted expression in the brain. Genomics
603 83, 727-734.
604 Varriale A. (2014) DNA methylation, epigenetics, and evolution in vertebrates: facts and challenges.
605 International Journal of Evolutionary Biology, 475981.
606 Wang J., Duncan D., Shi Z. & Zhang B. (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt):
607 update 2013. Nucleic Acids Research 41, W77-W83.
608 Wang M.D., Dzama K., Rees D.J. & Muchadeyi F.C. (2016) Tropically adapted cattle of Africa:
609 perspectives on potential role of copy number variations. Animal Genetics 47, 154-164.
610 Weyrich A., Lenz D., Jeschek M., Chung T.H., Rübensam K., Göritz F., Jewgenow K. & Fickel J. (2016)
611 Paternal intergenerational epigenetic response to heat exposure in male Wild guinea pigs.
612 Molecular Ecology 25, 1729-1740.
613 Wilde D. (2006) Influence of macro and micro minerals in the peri-parturient period on fertility in dairy
614 cattle. Animal Reproduction Science 96, 240-249.
615 Willham R.L. (1982) Genetic improvement of beef cattle in the United States: cattle, people and their
616 interaction. Journal of Animal Science 54, 659-666.
617 Wolf B. (2012) Biotinidase deficiency: if you have to have an inherited metabolic disease, this is the one
618 to have. Genetics in Medicine 14, 565-575.

619 Zhang Y., Baheti S. & Sun Z. (2016) Statistical method evaluation for differentially methylated CpGs in
620 base resolution next-generation DNA sequencing data. Briefings in Bioinformatics pii, bbw133.
621 Zhou Y., Xu L., Bickhart D.M., Abdel Hay E.H., Schroeder S.G., Connor E.E., Alexander L.J., Sonstegard T.S.,
622 Van Tassell C.P., Chen H., Liu G.E. (2016) Reduced representation bisulphite sequencing of ten
623 bovine somatic tissues reveals DNA methylation patterns and their impacts on gene expression.
624 BMC Genomics 17, 779.
625

626 **Table 1.** Geographic and climatic conditions of Creole and Spanish breeds.

Breed	Location	MASL[1]	MAT[2] (°C)	MARH[3] (%)	MAR[4] (mm)
Costeño con Cuernos	Department of Córdoba (Sinú river valley, Colombia)	300	30	80	2,500
San Martinero	Department of Meta (Colombia)	700	21	70	1,800
Asturiana de los Valles	Mieres (Asturias, Spain)	380	11	80	1,000
Lidia (Casta Navarra)	Igúzquiza (Navarra, Spain)	450	12	67	600
Retinta	Tierra de Barros (Badajoz, Spain)	400	17	66	450

627 [1] Metres above sea level
628 [2] Mean annual temperature
629 [3] Mean annual relative humidity
630 [4] Mean annual rainfall
631

Table 2. Differentially methylated regions (DMRs) overlapping a gene or CpGI and showing hyper- and hypomethylated levels above 80% in Creole samples.

Symbol	Gene name	Gene overlap	CpGI relation	Meth diff[1]
		Hypermethylated in Creole samples		
GBX2	gastrulation brain homeobox 2 (E1BJ47)	-	CpGI core	-0.84
LATS2	large tumor suppressor kinase 2	on intron	-	-0.83
BRAT1	BRCA1 associated ATM activator 1	-	CpGI shelf	-0.83
BTD	Biotinidase (F1MJM4)	intron exon boundary	-	-0.82
BLM	Bloom syndrome RecQ like helicase	on intron	-	-0.82
NKX2-8	NK2 homeobox 8 (E1BAC5)	on exon	CpGI core	-0.81
IRF6	interferon regulatory factor 6	on intron	CpGI core	-0.81
PTGDR	prostaglandin D2 receptor (PD2R)	on exon	CpGI core	-0.81
TP53I11	tumor protein p53 inducible protein 11 (PIG11)	on intron	-	-0.81
TM4SF5	transmembrane 4 L six family member 5 (T4S5)	intron exon boundary	-	-0.81
ATP13A3	ATPase 13A3 (E1BG26)	-	CpGI core	-0.81
TRIM25	tripartite motif containing 25 (A6QLA8)	on intron	CpGI shelf	-0.80
CYB561	cytochrome b-561 (CY561)	-	CpGI core	-0.80
FAM19A5	family with sequence similarity 19 member A5, C-C motif chemokine like (F19A5)	on exon	-	-0.80
RPGR	retinitis pigmentosa GTPase regulator	-	CpGI core	-0.80
SNX13	sorting nexin 13	on intron	-	-0.80
		Hypomethylated in Creole samples		
TNRC18	trinucleotide repeat containing 18	on intron	-	0.80
PAPLN	papilin, proteoglycan like sulfated glycoprotein	exon intron boundary	-	0.80
LGR6	leucine rich repeat containing G protein-coupled receptor 6 (LOC100336662)	on intron	-	0.81
PGLYRP1	peptidoglycan recognition protein 1 (PGRP1)	on exon	CpGI core	0.84

[1]Methylation differences averaged from all CpG sites within the defined region. Negative differential methylation values indicate hypermethylation in Creole samples; positive differential methylation values indicate hypomethylation in Creole samples.

638 **Table 3.** KEGG pathway enrichment analysis of differentially methylated genes between Creole
639 and Spanish cattle samples using WebGeStalt and DAVID tools.

GO term	Description	Gene count	Enrichment score	Genes
GO:1902105	Regulation of leukocyte differentiation	7	5,717	NRARP,LIF,PGLYRP1,PRKCZ,RUNX1,NKAP,CD83
GO:0050863	Regulation of T cell activation	7	5,449	NRARP,CD5,PRKCZ,MAD1L1,LMO1,NKAP,CD83
GO:1903037	Regulation of leukocyte cell-cell adhesion	7	5,128	NRARP,CD5,PRKCZ,MAD1L1,LMO1,NKAP,CD83
GO:0051249	Regulation of lymphocyte activation	8	4,689	NRARP,CD5,PGLYRP1,PRKCZ,MAD1L1,LMO1,NKAP, CD83
GO:0022407	Regulation of cell-cell adhesion	8	4,634	NRARP,CD5,PRKCZ,ALOX12,MAD1L1,LMO1,NKAP,C D83
GO:0050865	Regulation of cell activation	10	4,837	NRARP,CD5,PGLYRP1,PRKCZ,ALOX12,PDGFA,MAD1 L1,LMO1,NKAP,CD83
GO:0072359	Circulatory system development	14	3,013	NRARP,LIF,ALOX12,PDGFA,GBX2,MAP2K2,SMAD6, CYP1B1,EOMES,BAK1,SHOX2,FLRT2,ADM2,BCOR
GO:2000026	Regulation of multicellular organismal development	19	2,334	NRARP,LIF,PGLYRP1,PRKCZ,ALOX12,PDGFA,MAP2K 2,CYP1B1,EOMES,SHOX2,RUNX1,RFX4,BHLHE23,FL RT2,NKAP,CD83,ADM2,PHOX2B,BCOR
GO:0008283	Cell proliferation	21	2,341	NRARP,LIF,NPR3,ALOX12,PDGFA,GBX2,MAP2K2,SM AD6,CYP1B1,P3H2,BAK1,MAD1L1,DAGLA,SHOX2,L MO1,LTBP3,NKX2-8,MAB21L2,NKAP,IRF6,PHOX2B
GO:0048513	Animal organ development	28	1,915	NRARP,LIF,PGLYRP1,PRKCZ,ALOX12,PDGFA,GBX2, MAP2K2,SMAD6,CYP1B1,EOMES,BAK1,MAD1L1,SH OX2,IFITM5,HOXB1,RUNX1,RFX4,NKX2-8,BHLHE23,MAB21L2,FLRT2,NKAP,TMEM14C,IRF6, CD83,PHOX2B,BCOR
bta05221	Acute myeloid leukemia	3	6,448	MAP2K2, KIT, RUNX1
bta04910	Insulin signalling pathway	4	3,540	PRKCZ, SOCS2, MAP2K2, PRKAR1B
bta04015	Rap1 signalling pathway	5	2,812	PRKCZ, MAPK12, PDGFA, MAP2K2, KIT
bta05206	MicroRNAs in cancer	7	3,397	DNMT3A, CYP24A1, BAK1, CYP1B1, PDGFA, MAP2K2, PAK4

640
641
642

Figures

Figure 1. Chromosomal distribution of reads in the Creole (A) and Spanish (B) grouped samples. The distribution of reads is shown in a gradient from blue (low) to red (high).

Figure 2. DNA methylation levels in relation to gene bodies (A) and CpG islands (CpGI) (B). Gene bodies were defined as the region from the transcription start site (TSS) to transcription termination site (TTS).

Figure 3. KEGG pathway overrepresentation enrichment analysis (ORA) of differentially methylated genes between Creole and Spanish cattle samples performed with WebGeStalt.

Supporting information

Table S1. Bisulfite PCR primer sequences used for HiSeq-BSP validation of RRBS data on Creole and Spanish cattle samples.

Table S2. Creole (SM1, CCC1, CCC2) and Spanish (LD1, RAV, RET) cattle RRBS data summary.

Table S3. Number, coverage and methylation distribution of CpG_{10} in Creole (SM1, CCC1, CCC2) and Spanish (LD1, RAV, RET) cattle RRBS methylomes.

Table S4. CpG_{10} present in all six samples; annotated with the closest/overlapping transcription start sites (TSS) (±100 kb), including promoters (-2kb), introns and exons; and with the closest/overlapping CpG island (CpGI), including CpGI shores (±0-2 kb) and shelves (±2-4 kb).

Figure S1. Scatter plot and correlation of CpG_{10} methylation between Creole and Spanish cattle samples. Numbers above the diagonal denote pair-wise Pearson's correlation scores. The histograms on the diagonal are the methylation distribution of CpG_{10} sites for each sample. Below the diagonal, the scatter plots of percentage methylation values for each pair in Creole (SM1, CCC1, CCC2) and Spanish (LD1, RAV, RET) RRBS libraries are shown.

Figure S2. CpG_{10} site coverage histogram of (A) Creole (SM1, CCC1, CCC2) and (B) Spanish (LD1, RAV, RET) cattle RRBS libraries.

Figure S3. CpG_{10} methylation distribution in (A) Creole (SM1, CCC1, CCC2) and (B) Spanish (LD1, RAV, RET) cattle RRBS libraries.

Table S5. Differentially methylated regions (DMRs) showing a p-value < 0.01, ≥ 4 CpGs within a distance of 200 bp and mean methylation difference > 25% between Creole and Spanish sample

672 groups; annotation with the closest/overlapping transcription start sites (TSS) (±100 kb),
673 including promoters (-2kb), introns and exons; annotation with the closest/overlapping CpG
674 island (CpGI), including CpGI shores (±0-2 kb) and shelves (±2-4 kb). Positive differential
675 methylation values indicate hypomethylation in Creole samples; negative differential
676 methylation values indicate hypermethylation in Creole samples.
677 **Figure S4**. Functional annotation of genes differentially methylated between Creole and
678 Spanish cattle samples using PANTHER. A) Molecular function; B) Biological process; C) Cellular
679 component.
680 **Table S6**. Detailed functional annotation of the differentially methylated regions (DMRs)
681 showing a *p-value* < 0.01, ≥ 4 CpGs within a distance of 200 bp and a mean methylation
682 difference > 25% between Creole and Spanish cattle samples using DAVID Functional
683 Annotation Cluster (FAC) analysis under high stringency ease scores. Distribution of enriched
684 genomic regions along the cattle chromosomes obtained with DAVID.
685 **Figure S5.** KEGG signalling pathways of differentially methylated genes between Creole and
686 Spanish cattle samples obtained with DAVID tool.
687 **Table S7.** Comparison between HiSeq-BSP and RRBS DMR results.
688