

ISSN: 1559-2294 (Print) 1559-2308 (Online) Journal homepage: <https://www.tandfonline.com/loi/kepi20>

ISSN: 1559-2294 (Print) 1559-2308 (Online) Journal homepage: <https://www.tandfonline.com/loi/kepi20>

## Comparative analyses of sperm DNA methylomes among human, mouse and cattle provide insights into epigenomic evolution and complex traits

Lingzhao Fang, Yang Zhou, Shuli Liu, Jicai Jiang, Derek M. Bickhart, Daniel J. Null, Bingjie Li, Steven G. Schroeder, Benjamin D. Rosen, John B. Cole, Curtis P. Van Tassell, Li Ma & George E. Liu

To cite this article: Lingzhao Fang, Yang Zhou, Shuli Liu, Jicai Jiang, Derek M. Bickhart, Daniel J. Null, Bingjie Li, Steven G. Schroeder, Benjamin D. Rosen, John B. Cole, Curtis P. Van Tassell, Li Ma & George E. Liu (2019) Comparative analyses of sperm DNA methylomes among human, mouse and cattle provide insights into epigenomic evolution and complex traits, *Epigenetics*, 14:3, 260-276, DOI: [10.1080/15592294.2019.1582217](https://doi.org/10.1080/15592294.2019.1582217)

To link to this article: <https://doi.org/10.1080/15592294.2019.1582217>



This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.



View supplementary material [↗](#)



Accepted author version posted online: 27 Feb 2019.  
Published online: 24 Mar 2019.



Submit your article to this journal [↗](#)



Article views: 1280



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

RESEARCH PAPER



## Comparative analyses of sperm DNA methylomes among human, mouse and cattle provide insights into epigenomic evolution and complex traits

Lingzhao Fang<sup>a,b,\*</sup>, Yang Zhou<sup>c\*</sup>, Shuli Liu<sup>a,d\*</sup>, Jicai Jiang<sup>b</sup>, Derek M. Bickhart <sup>e</sup>, Daniel J. Null<sup>a</sup>, Bingjie Li<sup>a</sup>, Steven G. Schroeder <sup>a</sup>, Benjamin D. Rosen <sup>a</sup>, John B. Cole <sup>a</sup>, Curtis P. Van Tassell <sup>a</sup>, Li Ma<sup>b</sup>, and George E. Liu <sup>a</sup>

<sup>a</sup>Animal Genomics and Improvement Laboratory, BARC, Agricultural Research Service, USDA, Beltsville, MD, USA; <sup>b</sup>Department of Animal and Avian Sciences, University of Maryland, College Park, MD, USA; <sup>c</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Education Ministry of China, Huazhong Agricultural University, Wuhan, Hubei, China; <sup>d</sup>Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture & National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China; <sup>e</sup>Dairy Forage Research Center, Agricultural Research Service, USDA, Madison, WI, USA

### ABSTRACT

Sperm DNA methylation is crucial for fertility and viability of offspring but epigenome evolution in mammals is largely understudied. By comparing sperm DNA methylomes and large-scale genome-wide association study (GWAS) signals between human and cattle, we aimed to examine the DNA methylome evolution and its associations with complex phenotypes in mammals. Our analysis revealed that genes with conserved non-methylated promoters (e.g., *ANKS1A* and *WNT7A*) among human and cattle were involved in common system and embryo development, and enriched for GWAS signals of body conformation traits in both species, while genes with conserved hypermethylated promoters (e.g., *TCAP* and *CD80*) were engaged in immune responses and highlighted by immune-related traits. On the other hand, genes with human-specific hypomethylated promoters (e.g., *FOXP2* and *HYDIN*) were engaged in neuron system development and enriched for GWAS signals of brain-related traits, while genes with cattle-specific hypomethylated promoters (e.g., *LDHB* and *DGAT2*) mainly participated in lipid storage and metabolism. We validated our findings using sperm-retained nucleosome, preimplantation transcriptome, and adult tissue transcriptome data, as well as sequence evolutionary features, including motif binding sites, mutation rates, recombination rates and evolution signatures. In conclusion, our results demonstrate important roles of epigenome evolution in shaping the genetic architecture underlying complex phenotypes, hence enhance signal prioritization in GWAS and provide valuable information for human neurological disorders and livestock genetic improvement.

### ARTICLE HISTORY

Received 19 October 2018  
Revised 5 January 2019  
Accepted 3 February 2019

### KEYWORDS

Sperm DNA methylation; epigenome evolution; large-scale GWAS; cattle complex traits; human complex traits

## Introduction

Proper DNA methylation in sperm is essential for the embryogenesis and normal development through gene expression regulation [1-5]. Aberrant sperm DNA methylation is often associated with impaired male fertility status, embryo quality and offspring disorders susceptibility [6-9]. Compared to other epigenetic regulatory elements like histone modifications, DNA methylation is a more stable regulator of gene expression [10], which controls a long-term transcriptional activity of corresponding genes [11]. More importantly, although it undergoes nearly complete reprogramming in mammals during germ cell and preimplantation development, DNA methylome can be inherited over generations with a heritability of

approximate 20% (i.e., transgenerational inheritance) [10,12-15]. For instance, Wang et al., (2014) reported that the methylation levels of approximate 6.8% of GpG sites maintained stable across sperm and embryonic developmental stages in mouse, with the majority of which were either hypermethylated (> 80%) or hypomethylated (< 20%) [13]. More recently, Li et al., (2018) confirmed that hypomethylated regions around gene promoters were highly preserve over developmental stages and across species, and were regulated by Ploypomb through ten-eleven translocation proteins [16].

Furthermore, comparative analysis of sperm DNA methylation has been a promising way to determine genomic regions under epigenome evolution [17-19],

**CONTACT** Li Ma  [lima@umd.edu](mailto:lima@umd.edu)  Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA; George E. Liu  [George.Liu@ars.usda.gov](mailto:George.Liu@ars.usda.gov)  Animal Genomics and Improvement Laboratory, USDA-ARS, Building 306, Room 111, BARC-East, Beltsville, MD 20705, USA

\*These authors contributed equally to this work

 Supplementary data for this article can be accessed [here](#).

This work was authored as part of the Contributor's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

which is partially responsible for environmental adaptation [20]. For instance, by comparing sperm methylation across seven mammalian species, Qu et al. (2018) have revealed that hypomethylated regions (HMRs) around promoters are under substantial evolutionary changes, and exhibit strong lineage-specific aspects [18]. The DNA methylation also closely interplays with the underlying nucleotide sequence due to the increased transition rates of methylated CpG to TpG [17,20,21], thereby influencing the evolution of genome and contributing to the fixation of beneficial environmental adaptations into genome [22,23]. Yi (2017) has reviewed that the spontaneous epimutations in the germline, which are caused by the environmental perturbation, may influence the sequence evolution due to their transgenerational inheritance model, and in turn the sequence evolution may affect methylome evolution through three ways, including mutation of CpG sites, biased gene conversion and transcription factor (TF) binding site turnover [23]. Together, we hypothesized that studying DNA methylome evolution in sperm could help explore the evolutionarily molecular mechanism underlying lineage-specific complex traits, thus contributing to a better understanding of environmental adaptation.

Cattle (*Bos taurus*) and human diverged from a common ancestor ~90 million years ago [24], and cattle is a valuable source for understanding the biology and evolution of mammals with the large-scale and accurate phenotypes, which were computed based on millions of offspring. We have conducted genome-wide association study (GWAS) for 35 complex traits with the sample size of 27,214 U.S. Holstein bulls with high reliable phenotypes and imputed sequence variants ( $n = \sim 3$  million), including body type, fertility and production traits (Jiang et al., 2018, submitted; <https://www.biorxiv.org/content/early/2018/09/26/428227>). We sequenced 10 Holstein cattle sperm using the whole-genome bisulfite sequencing (WGBS) technology, aiming to determine genomic regions under epigenetic evolution through comparing with the human data [25]. We studied the associations of epigenetic evolution with complex phenotypes by integrating these epigenetic features with GWAS signals for 35 and 60 complex traits and diseases in cattle and human, respectively. We validated our findings through large-scale integration analyses of sperm-retained nucleosome, histone modifications, preimplantation transcriptome, tissue

proteome data, and several sequence evolutionary features (e.g., motif binding sites, mutation rates, recombination rates and evolution signatures). Our study illustrated that the epigenome evolution contributes to the shape of the genetic architecture underlying complex traits and diseases.

## Results

### General characteristics of sperm methylation

Our WGBS data had an average mapping rate of 71.47% with an overall methylation level of approximate 75% for all CpG loci in the cattle genome (Supplementary Table S1). In general, we found the majority of genomic elements (e.g., genic regions and repeat elements like LINE and SINE) were highly methylated ( $> 80\%$  on average) in the cattle sperm (Supplementary Fig. S1A), and similar patterns held in the human sperm methylation [17,25]. Of note were the methylation levels of promoters (i.e.,  $\pm 1000$ bp of transcriptional start site – TSS) and CG islands (CGI), which followed a clear bimodal pattern (i.e., peaks at  $< 20\%$  and  $> 80\%$ , respectively) with an overall average of less than 40% (Supplementary Fig. S1B). We further observed that the global methylation levels among all ten cattle samples were highly correlated (Pearson's correlation  $r > 0.91$ ) (Supplementary Fig. S1C), consistent in each genomic element, particularly in promoters, CGI, low complexity sequence and simple repeats (Supplementary Fig. S1D). This was in line with the human sperm methylation levels, which exhibited high between-individual correlations of 0.89, 0.94 and 0.78 for genome-wide, promoters and repeats, respectively [17]. Our results demonstrated that the sperm DNA methylation is globally highly methylated and correlated among individuals.

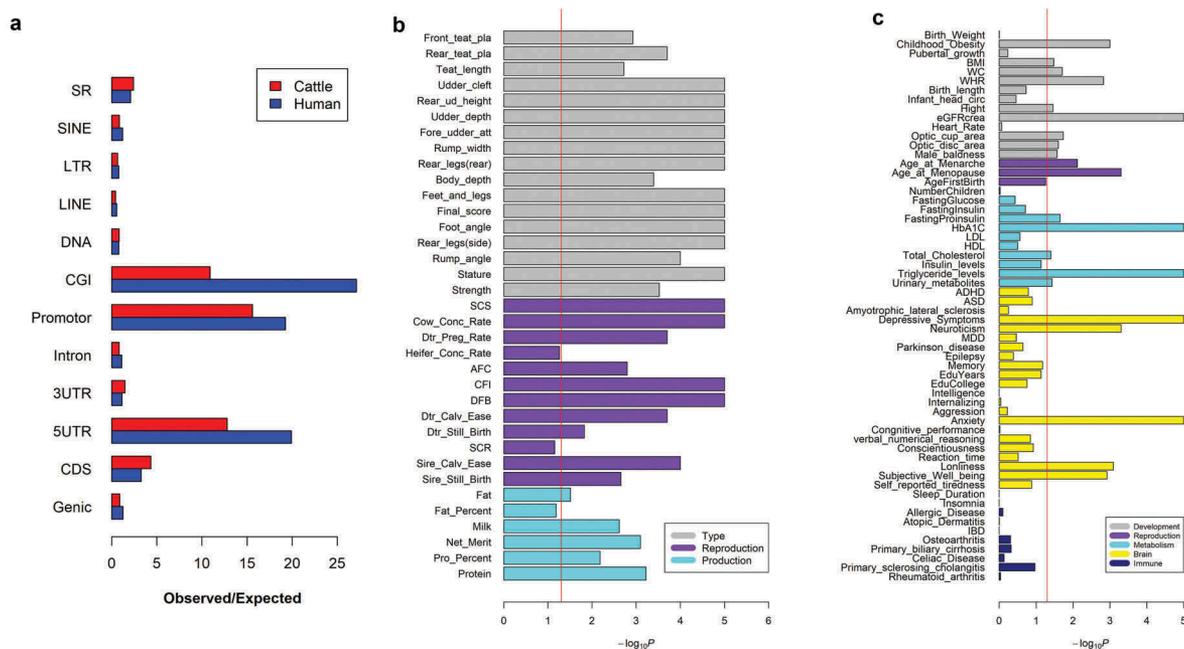
### Sperm HMRs were enriched for GWAS signals of developmental traits

Because of the well-known roles of HMRs in the gene activation and regulation [26,27], we subsequently explored the sperm HMRs in both human and cattle, which may contribute to multiple biological processes in the embryonic development and organ morphogenesis, hence potentially influencing a range of complex phenotypes. In total, we detected

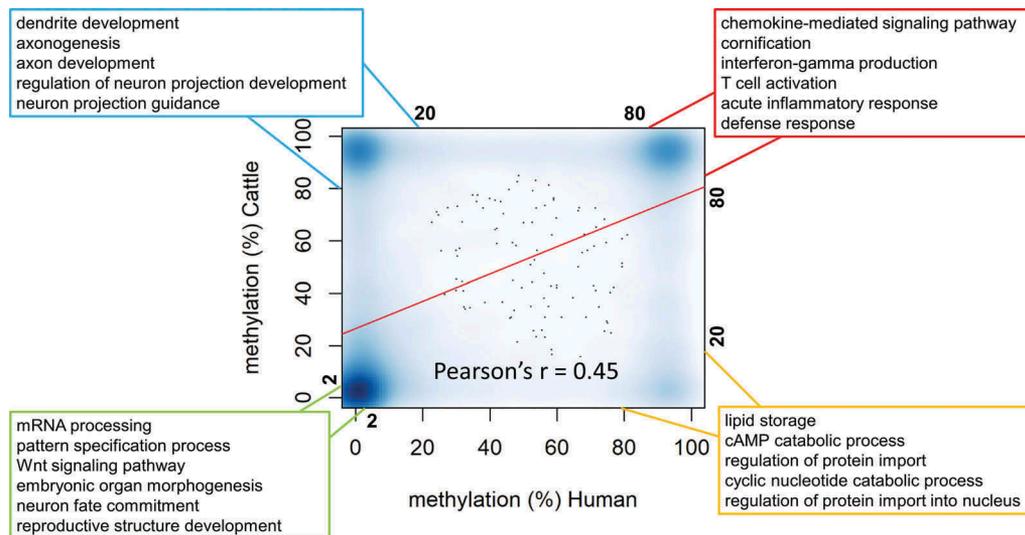
72,618 (covering ~0.79% of cattle genome) and 138,329 (covering ~2.5% of human genome) HMRs in cattle and human, respectively. The HMRs overlapped with all genomic elements, but highly intersected promoters and CGI, which were consistent in cattle and human (Figure 1(a)). We found that cattle sperm HMRs were significantly ( $P < 0.05$ ; genome-wide marker-set test) enriched for GWAS signals of 32 out of the 35 complex traits in cattle. Of special note were the enrichments (i.e.,  $-\log_{10}P$ ) of body type and reproduction traits, which were significantly ( $P < 0.05$ ; Wilcoxon-test) higher than those of milk production traits (Figure 1(b)). We confirmed that human sperm HMRs were also significantly enriched for GWAS signals across many complex traits, including 9 out of 14 body developmental traits, 2 out of 4 reproductive traits, 5 out of 10 metabolic traits, and 5 out of 24 brain-related traits, but none out of 8 immune-related traits (Figure 1(c)). Our results demonstrated that sperm HMRs were likely to harbor regulatory causative mutations across many complex traits, particularly in developmental traits (e.g., human height and cattle stature), which strongly supported their crucial roles in the normal development.

### Sperm methylation evolution of orthologous genes in mammals

Since promoters were believed to mediate gene expression via methylation levels and to guide evolution across millions of years [28,29], we investigated the conservation and divergence of methylation in promoters of 14,459 orthologous gene-pairs between human and cattle. Consistent with that cattle and human shared ~80% sequence sequence homology [24,30], we found methylation levels of promoters in orthologous genes were also correlated (Pearson's  $r = 0.45$ ) (Figure 2), implying that a decent fraction of epigenome remained conserved across millions of years of evolution in mammals. To confirm this, we further examined the mouse sperm methylome [31], and found that correlations of promoters of orthologous genes were 0.53 ( $n = 14,392$ ) and 0.44 ( $n = 14,459$ ) for human vs. mouse and cattle vs. mouse, respectively (Supplementary Fig. S2). Based on the bimodal distribution of promoter methylation (Fig. S1B), we here identified four categories of genes based on their promoter's methylation levels in human and cattle (Figure 2): (1) 2,761 genes with non-methylated (< 2%) promoters in both species (nMeth-genes); (2) 1,904 genes with



**Figure 1.** The enrichment of sperm hypomethylated regions (HMRs) for genomic elements and GWAS signals. (a) The enrichment of HMRs across genomic elements. (b) The GWAS signal enrichment of cattle sperm HMRs for 35 complex traits in dairy cattle. (c) The GWAS signal enrichment of human sperm HMRs for 60 complex traits in human. The red line corresponds to  $P = 0.05$  on the basis of genome-wide marker-set test.



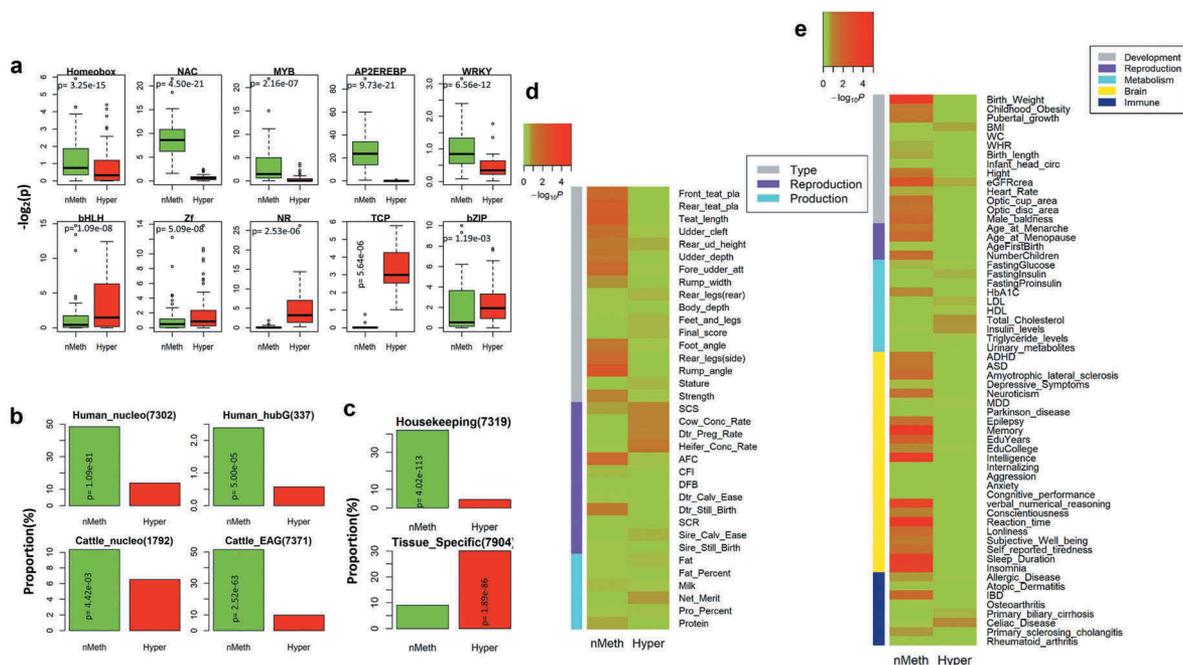
**Figure 2.** The relationship of methylation in promoters of 14,459 orthologous gene-pairs between human and cattle. The promoters of genes were defined as 1000bp up and down-stream transcriptional start sites (TSS). The biological processes terms in each box were the top significantly (adjusted- $p$  value < 0.05) enriched representative terms for the corresponding gene lists using Gene Ontology database (details in Table S2). The green box was for 2,761 genes with non-methylated (< 2%) promoters in both species (nMeth-genes); The red box was for 1,904 genes with hypermethylated (> 80%) promoters in both species (hyper-genes); The blue box was for 2,228 genes with hypermethylated promoters in cattle but hypomethylated (< 20%) promoters in human (CRHO); The orange box was for 410 genes with hypomethylated promoters in cattle but hypermethylated promoters in human (COHR).

hypermethylated (> 80%) promoters in both species (hyper-genes); (3) 2,228 genes with hypermethylated promoters in cattle but hypomethylated (< 20%) promoters in human (CRHO); (4) 410 genes with hypomethylated promoters in cattle but hypermethylated promoters in human (COHR). Function annotation using the Gene Ontology (GO) database for both cattle and human revealed their distinct biological functions of evolutionary significance: nMeth-genes (e.g., *ANKS1A* and *WNT7A*) significantly (Bonferroni-adjusted  $P < 0.01$ ) participated in biological processes of embryonic developmental importance, such as primarily mRNA processing, WNT signaling pathway and embryonic development; hyper-genes (e.g., *TCAP* and *CD80*) played significant roles in the immune system, such as T cell activation and lymphocyte proliferation; CRHO genes (e.g., *FOXP2* and *HYDIN*) mainly functioned in neurodevelopment, such as axon and dendrite development; while COHR genes (e.g., *LDHB* and *DGAT2*) participated in lipid storage and metabolism (Figure 2; details for all significant GO terms see Supplementary Table S2). Additionally, as expected we confirmed that promoters of nMeth-genes and hyper-genes consistently exhibited hypomethylation and hypermethylation in the mouse sperm, respectively, revealing their

conservation in methylation across mammals. On the other hand, CRHO and COHR showed no difference in methylation in the mouse sperm, probably because they were human and cattle-specific hypomethylated genes, respectively. Of interest, the methylation levels of promoters of both CRHO and COHR tended to be lower than those of the remaining genes in the mouse sperm, indicating their particular functions in the mouse require further investigations (Supplementary Fig. S3).

### Genes with conserved none and highly methylated promoters

Our motif analyses revealed that promoters of nMeth-genes were significantly (FDR < 0.01) enriched for 99 and 86 out of all 921 tested motifs in human and cattle, respectively, while promoters of hyper-genes were significantly enriched for only 7 motifs in human and cattle (Supplementary Table S3). These results illustrated that promoters of nMeth-genes were hotspots of transcription factors, suggesting that developmental genes required a complex regulatory mechanism that involved a large amount of regulatory elements and factors to ensure their proper functions. We then grouped all the tested motifs based on their



**Figure 3.** Comparison of genes with conserved non-methylated (< 2%) promoters (nMeth-genes) and genes with conserved hypermethylated (> 80%) promoters (hyper-genes) in human and cattle. (a) Difference in motif enrichment between nMeth-genes and hyper-genes in human, similar results held in cattle (Table S5). (b) Overlaps of nMeth-genes and hyper-genes with genes associated with sperm-retained nucleosome and genes activated preimplantation in human and cattle, and  $P$  values calculated by Fisher exact test. hubG is the hub genes that were activated preimplantation, while EAG was the embryonic activated genes preimplantation. (c) Overlaps of nMeth-genes and hyper-genes with house-keeping and tissue-specific expressed genes, and  $P$  values calculated by Fisher exact test. (d) GWAS signal enrichment of nMeth-genes and hyper-genes for 35 complex traits in cattle, and  $P$  values were obtained by genic marker-set test. (e) GWAS signal enrichment of nMeth-genes and hyper-genes for 60 complex traits in human, and  $P$  values were obtained by genic marker-set test.

DNA binding domains, and observed that promoters of nMeth-genes had a significantly higher enrichment than those of hyper-genes across many embryonic developmental motifs, including Homeobox, NAC, MYB, AP2EREBP and WRKY motif families. In contrast, promoters of hyper-genes prefer to enrich for motifs associated with immunity and metabolism, including bHLH, zf, NR, TCP and bZIP motif families (Figure 3(a)). To further support that nMeth-genes not hyper-genes function at the early stages of embryonic development in mammals, we examined datasets of sperm-retained nucleosomes and embryonic activation transcriptomes in both human and cattle [32–36], which corresponded to transcriptionally active genes before implantation. We confirmed that nMeth-genes not hyper-genes were significantly ( $P < 0.01$ ; Fisher exact test) overlapped with nucleosome-associated genes (i.e., genes in the closest vicinity to nucleosome peaks) and embryonic activated genes, which was consistent in human and cattle (Figure 3(b)). In

addition, nMeth-genes also significantly intersected genes that were associated with the attendant histone modifications (i.e., H3K4me2 and H3K27me3) in human sperm (Supplementary Fig. S4) [32], which were believed to be markedly enriched at loci of early embryonic developmental importance [32,37]. We further revealed that nMeth-genes were house-keeping genes, while hyper-genes were tissue-specific genes through examining the dataset of human protein atlas that measured the gene expression levels in all major tissues and organs in the human body [38] (Figure 3(c)). Our genic marker-set test analyses demonstrated that nMeth-genes were significantly ( $P < 0.05$ ) enriched for GWAS signals in body type traits in cattle, while hyper-genes were significantly and selectively enriched for several immune-related traits in cattle, including somatic cell sore (SCS), cow conception rate, daughter pregnancy rate and heifer conception rate [39,40] (Figure 3(d)). Similar patterns held in human, as nMeth-genes were significantly enriched for many

body developmental and brain-related traits, while hyper-genes were more likely to be associated with metabolism and immune-related traits (Figure 3(e)). Together, all these observations indicated that non-methylated active promoters in sperm were essential for maintaining the plasticity of developmental genes to guide the embryonic development, while hyper-methylated inactive promoters in sperm might be important for silencing immune genes to ensure the successful implantation [41].

We defined a gene, whose promoter and genic regions bear suggestive significant SNPs ( $P < 1e-5$ ) for at least two complex traits, as a pleiotropic gene. We observed 175 out of 960 and 317 out of 2,261 pleiotropic genes were nMeth-genes in cattle and human, respectively, as more often than expected ( $P = 8.24e-5$  in cattle;  $P = 3.65e-3$  in human; Fisher exact test). This indicated that nMeth-genes were likely to be pleiotropic genes, thereby influencing multiple complex traits and diseases. We found 30 nMeth-genes were common pleiotropic genes shared in both human and cattle, and many of those were associated with similar traits between human and cattle (Table 1). For example, *ANKS1A* gene, which is a well-known epidermal growth factor receptor [42], had non-methylated promoters in both human and cattle. It was associated with multiple body developmental traits in both human and cattle (Figure 4(a)). In contrast, we found two hyper-genes, *CCRL2* and *TCAP*, were associated with immune-related traits in both cattle and human. *CCRL2*, a chemokine receptor, was associated with SCS (chr22:53567811;  $P = 8.90e-06$ ) in cattle, and with rheumatoid arthritis (rs6762266;  $P = 4.30e-06$ ) in human. Here we showed *TCAP* as an example in Figure 4(b), which was engaged in many immune pathways [43]. *TCAP* was associated with SCS in cattle, and with primary biliary cirrhosis, inflammatory bowel disease, urinary metabolites, rheumatoid arthritis, and allergic disease in human.

#### **Genes with cattle and human-specific low methylated promoters**

Our motif analyses showed that promoters of CRHO were significantly enriched for 27 motifs in human, including many developmental motifs belonging to AP2EREBP and MYB families, but they were

significantly enriched for only 7 motifs in cattle, with the majority of those belonged to the bHLH family (Supplementary Table S3). In contrast, promoters of COHR were significantly enriched for two metabolism-related motifs, *LXRE* and *MEF2C*, in human, but for two developmental motifs, *ERF115* and *CRF10*, in cattle (Supplementary Table S3), suggesting these genes may start to function at the early embryonic development exclusively in cattle. Genic marker-set test analysis revealed CRHO had a significantly higher enrichment (i.e.,  $-\log_{10}P$ ) than COHR for brain-related traits in human, while COHR had a significantly higher enrichment than CRHO for milk production traits in cattle (Figure 5(a)).

We detected 914 brain-associated genes, whose genic and promoter regions harbor suggestive significant SNPs of brain-related traits, among those 196 were CRHO genes, as more often than expected ( $P = 1.63e-22$ ; Fisher exact test) (Supplementary Table S4). This also indicated that CRHO genes were associated with brain-related traits. For example, *FOXP2*, a well-known language and speech related gene [44], had a hypomethylated promoter with a CGI in human, but a hypermethylated promoter without any CGI in cattle (Figure 5(b,c)). It was associated with four brain-related traits in human, including attention deficit hyperactivity disorder, insomnia, sleep duration, and verbal numerical reasoning (Figure 5(b)). Similarly, we detected 517 production-associated genes in cattle, among which 17 were COHR genes ( $P = 6.65e-02$ ; Fisher exact test) (Supplementary Table S5). For example, *LDHB*, which was significantly up-regulated during lactation in dairy cows [45], had a hypomethylated promoter with a CGI in cattle, but a hypermethylated promoter without any CGI in human (Figure 5(d,e)). It was associated with three milk production traits in cattle, including milk yield, protein yield and fat percentage (Figure 5(d)). All the results here demonstrated that comparative epigenome analyses could contribute to the detection of causative genes for lineage-specific traits, thereby providing novel insights into the genetic and evolutionary mechanisms underlying, which is partially related to environmental adaptation.

#### **Sequence evolution accompanying sperm methylation evolution**

Since methylcytosines were hotspots of mutations that were the ultimate source of natural selection

**Table 1.** The summary for 30 pleiotropic genes (harboring SNP with  $P < 1e-05$  for at least two traits) with non-methylated promoters (nMeth-genes) in both human and cattle.

Gene	Human			Cattle		
	Chr	Position	Trait	Chr	Position	Trait
<i>ACAD11</i>	3	132,558,138–132,660,723	<b>LDL</b> (rs2270801; 4.78e-06); <b>Total_Cholesterol</b> (rs2270801; 1.89e-06)	1	137,993,897–138,111,878	<b>Final_score</b> (1:138,018,400; 5.65e-06); <b>Sire_Still_Birth</b> (1:138,018,400; 5.59e-06)
<i>ADCY5</i>	3	123,282,296–123,449,758	<b>FastingGlucose</b> (rs11708067; 6.79e-09); <b>Birth_Weight</b> (rs11719201; 6.40e-27)	1	68,263,141–68,420,954	<b>Cow_Conc_Rate</b> (1:68,404,625; 3.08e-06); <b>Dtr_Preg_Rate</b> (1:68,404,625; 1.18e-07)
<i>AFF1</i>	4	86,935,002–87,141,054	<b>HDL</b> (rs442177; 2.74e-07); <b>Optic_cup_area</b> (rs192337969; 6.81e-07); <b>HEIGHT</b> (rs1408; 4.90e-06); <b>Optic_disc_area</b> (rs192337969; 6.81e-07)	6	103,688,011–103,825,580	<b>AFC_DYD</b> (6:103,782,949; 3.52e-07); <b>Heifer_Conc_Rate</b> (6:103,782,949; 5.28e-06)
<i>ANKS1A</i>	6	34,889,265–35,091,413	<b>Pubertal_growth</b> (rs9469890; 9.41e-08); <b>Rheumatoid_arthritis</b> (6:34,889,423; 7.60e-06); <b>FastingInsulin</b> (rs2820237; 5.03e-07); <b>Total_Cholesterol</b> (rs3822921; 2.61e-06); <b>HDL</b> (rs7742443; 6.27e-06); <b>BMI</b> (rs6457796; 1.15e-09); <b>HEIGHT</b> (rs13210323; 1.10e-27); <b>WC</b> (rs9380470; 5.80e-08)	23	8,838,316–9,004,998	<b>Sire_Still_Birth</b> (23:8,906,458; 9.18e-10); <b>Stature</b> (23:8,988,566; 5.36e-09); <b>Udder_depth</b> (23:8,988,566; 1.70e-10)
<i>ARID5B</i>	10	61,901,300–62,096,944	<b>ADHD</b> (rs713240; 4.56e-07); <b>Rheumatoid_arthritis</b> (rs71508903; 2.30e-20)	28	18,003,736–18,191,994	<b>Milk</b> (28:18,186,635; 2.20e-07); <b>Protein</b> (28:18,028,657; 7.26e-06)
<i>BCAS3</i>	17	60,677,453–61,392,838	<b>eGFRcrea</b> (rs11657044; 7.90e-22); <b>BMI</b> (rs8075273; 1.98e-07); <b>HEIGHT</b> (rs2286530; 3.10e-13); <b>WHR</b> (rs7223966; 1.40e-07)	19	12,242,530–12,588,255	<b>Heifer_Conc_Rate</b> (19:12,459,490; 1.18e-09); <b>Sire_Calv_Ease</b> (19:12,432,636; 5.06e-07); <b>Sire_Still_Birth</b> (19:12,432,710; 3.56e-07)
<i>CCDC85A</i>	2	56,184,123–56,386,173	<b>Insomnia</b> (rs7423933; 9.05e-06); <b>Age_at_Menarche</b> (rs6747380; 2.10e-17)	11	38,706,659–38,928,773	<b>Fore_udder_att</b> (11:38,764,147; 1.32e-07); <b>SCR</b> (11:38,711,041; 3.71e-05); <b>Udder_depth</b> (11: 38,764,147; 5.85e-07)
<i>CCSER1</i>	4	90,127,535–91,601,913	<b>Optic_cup_area</b> (rs150782662; 2.41e-06); <b>verbal_numerical_reasoning</b> (rs372420824; 8.78e-06); <b>Optic_disc_area</b> (rs150782662; 2.41e-06)	6	35,100,701–35,938,394	<b>Fat_Percent</b> (6:35,683,142; 6.96e-13); <b>Milk</b> (6:35,683,142; 2.46e-08); <b>Pro_Percent</b> (6:35,683,142; 2.79e-21); <b>Sire_Still_Birth</b> (6:35,352,421; 1.15e-08)
<i>CHEK2</i>	22	28,687,743–28,742,422	<b>Age_at_Menopause</b> (rs6005843; 2.00e-06); <b>Male_baldness</b> (rs2236141; 5.18e-06); <b>Optic_cup_area</b> (rs5762752; 1.45e-08)	17	70,269,592–70,305,247	<b>Dtr_Calv_Ease</b> (17:70,269,288; 2.25e-11); <b>Sire_Calv_Ease</b> (17:70,281,680; 4.99e-10)
<i>DPYD</i>	1	97,077,743–97,921,049	<b>Age_at_Menarche</b> (rs11165924; 2.00e-06); <b>Intelligence</b> (rs78164635; 9.09e-08); <b>WC</b> (rs11165922; 7.50e-06)	3	45,563,732–46,487,165	<b>Dtr_Still_Birth</b> (3:45,649,946; 6.68e-06); <b>Sire_Still_Birth</b> (3:46,458,100; 2.24e-08)
<i>FCHSD2</i>	11	72,836,745–73,142,261	<b>FastingProinsulin</b> (rs11605166; 4.65e-31); <b>IBD</b> (11:72,680,931; 1.07e-06)	15	53,161,555–53,404,313	<b>Milk</b> (15:53,253,308; 1.61e-06); <b>Pro_Percent</b> (15:53,396,577; 2.53e-11)
<i>FOLH1</i>	11	49,146,635–49,208,670	<b>HDL</b> (rs648728; 1.99e-09); <b>HEIGHT</b> (rs202700; 1.60e-11)	29	5,904,662–5,977,332	<b>AFC</b> (29:5,917,586; 3.62e-07); <b>Sire_Still_Birth</b> (29:5,958,709; 5.87e-06)
<i>GNAQ</i>	9	77,716,087–78,031,458	<b>Intelligence</b> (rs11371901; 3.66e-06); <b>verbal_numerical_reasoning</b> (rs11371901; 8.12e-07)	8	53,970,972–54,280,697	<b>Rump_angle</b> (8:54,066,760; 6.31e-06); <b>SCR</b> (8: 54,065,711; 1.48e-06)

(Continued)

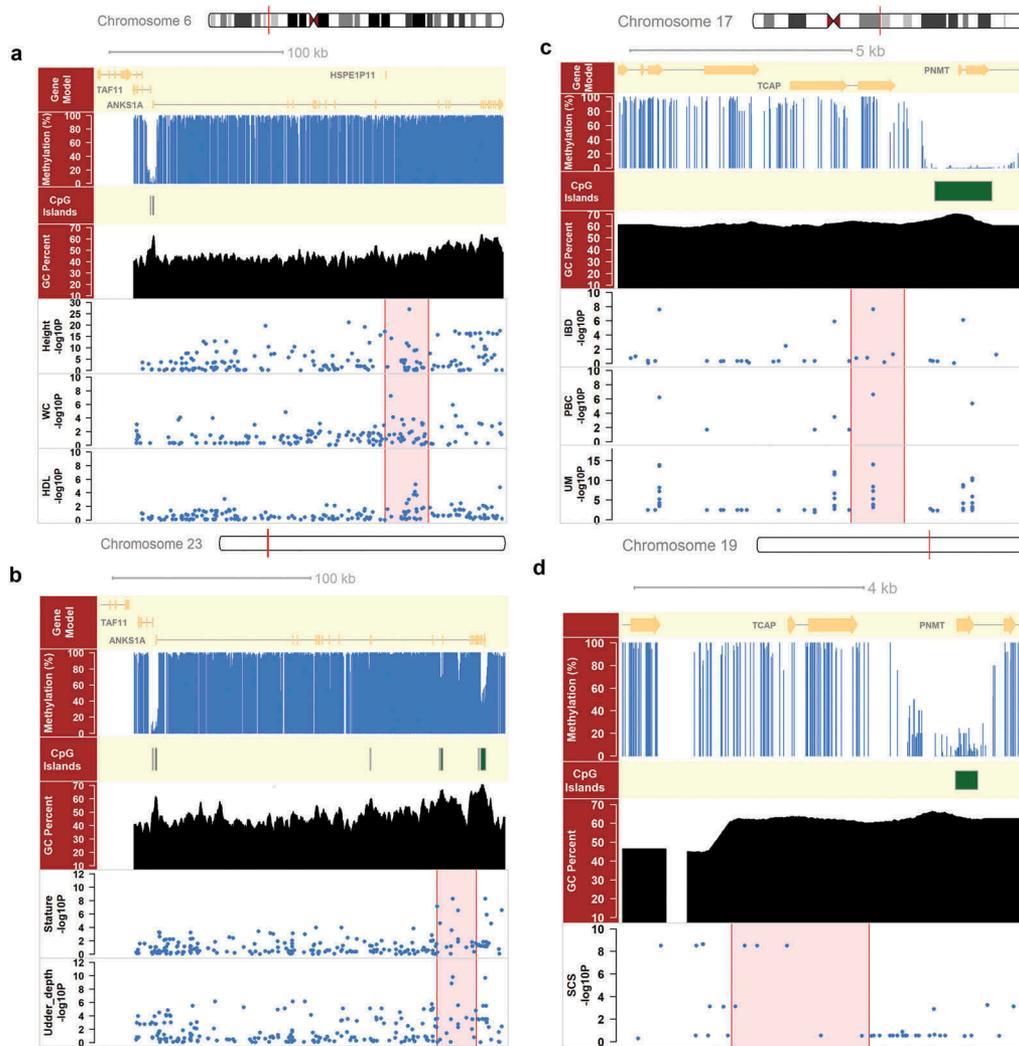
Table 1. (Continued).

Gene	Human			Cattle		
	Chr	Position	Trait	Chr	Position	Trait
<i>GPC5</i>	13	91,398,607–92,873,682	<b>Self_reported_tiredness</b> (13:92,092,884; 1.91e-06); <b>HEIGHT</b> (rs2147190; 2.00e-07)	12	66,292,489–67,324,791	<b>Dtr_Calv_Ease</b> (12:66,700,564; 2.66e-07); <b>Dtr_Still_Birth</b> (12:66,702,395; 1.30e-06); <b>Sire_Still_Birth</b> (12:67,181,879; 2.62e-07)
<i>HERC1</i>	15	63,608,618–63,833,942	<b>Male_baldness</b> (rs533456110; 7.45e-06); <b>Celiac_Disease</b> (rs7168440; 8.86e-06)	10	46,241,889–46,455,782	<b>Fat_Percent</b> (10:46,450,064; 4.15e-06); <b>Heifer_Conc_Rate</b> (10:46,311,738; 4.83e-09); <b>Pro_Percent</b> (10:46,307,205; 5.69e-12); <b>Sire_Still_Birth</b> (10:46,431,104; 8.65e-06); <b>Fat</b> (1:82,005,751; 2.35e-06); <b>Sire_Still_Birth</b> (1:82,024,433; 4.44e-09)
<i>IGF2BP2</i>	3	185,643,739–185,825,056	<b>BMI</b> (rs1687230; 3.93e-06); <b>HEIGHT</b> (rs17826758; 8.10e-19)	1	81,992,398–82,161,528	<b>Body_depth</b> (5:83,624,334; 2.93e-07); <b>Sire_Still_Birth</b> (5:83,431,446; 1.62e-07)
<i>ITPR2</i>	12	26,336,515–26,833,198	<b>Birth_Weight</b> (rs12823128; 3.20e-08); <b>IBD</b> (12: 26,588,799; 5.59e-06)	5	83,398,954–83,986,176	<b>Milk</b> (28:19,452,458; 5.40e-06); <b>Protein</b> (28:19,452,458; 6.17e-07); <b>SCR</b> (28:19,426,830; 5.51e-06)
<i>JMJD1C</i>	10	63,167,221–63,521,850	<b>EduCollege</b> (rs7896518; 5.84e-06); <b>Intelligence</b> (rs907; 5.43e-08); <b>Allergic_Disease</b> (rs184034096; 5.42e-06); <b>HDL</b> (rs10761731; 2.52e-07); <b>WHR</b> (rs10509189; 5.90e-07)	28	19,393,375–19,663,523	<b>Birth_Weight</b> (rs488079; 1.70e-07); <b>IBD</b> (1:155,888,666; 6.10e-06)
<i>KIAA0907</i>	1	155,913,043–155,934,442	<b>Birth_Weight</b> (rs488079; 1.70e-07); <b>IBD</b> (1:155,888,666; 6.10e-06)	3	14,889,681–14,906,957	<b>Body_depth</b> (11:75,510,879; 9.70e-06); <b>Final_score</b> (11:75,466,899; 1.50e-08); <b>Sire_Still_Birth</b> (11:75,466,899; 3.32e-06)
<i>KLHL29</i>	2	23,385,217–23,708,611	<b>Birth_Weight</b> (rs72796106; 6.70e-11); <b>HEIGHT</b> (rs3795938; 1.00e-15)	11	75,290,644–75,623,195	<b>SCR</b> (24:34,823,294 ; 4.69e-06); <b>Sire_Calv_Ease</b> (24:34,839,875; 2.65e-06)
<i>MIB1</i>	18	21,704,957–21,870,957	<b>BMI</b> (rs1788785; 1.98e-08); <b>HEIGHT</b> (rs3017036; 4.20e-09)	24	34,812,337–34,901,981	<b>Fat_Percent</b> (6:38,773,632; 8.91e-09); <b>Milk</b> (6:38,773,632; 8.90e-06); <b>Pro_Percent</b> (6:38,767,794; 7.09e-28); <b>Sire_Still_Birth</b> (6:38,782,871; 6.85e-07)
<i>NCAPG</i>	4	17,810,902–17,844,862	<b>Birth_Weight</b> (rs2074974; 5.00e-10); <b>HEIGHT</b> (rs11931594; 7.70e-60)	6	38,765,969–38,812,051	<b>Dtr_Still_Birth</b> (6:51,620,860; 7.48e-07); <b>Sire_Still_Birth</b> (6:51,682,838; 2.25e-08)
<i>PCDH7</i>	4	30,720,415–31,146,805	<b>ADHD</b> (rs145102344; 6.43e-08); <b>Epilepsy</b> (rs1044352; 3.40e-08)	6	51,536,863–52,007,783	<b>Fore_udder_att</b> (11:38,184,059; 3.75e-07); <b>Udder_depth</b> (11:38,184,059; 2.40e-07)
<i>PNPT1</i>	2	55,634,265–55,693,910	<b>HEIGHT</b> (rs706550; 5.90e-09); <b>WC</b> (rs4146921; 2.20e-10)	11	38,161,134–38,200,964	<b>Dtr_Still_Birth</b> (9:67,048,185; 2.99e-06); <b>Front_teat_pla</b> (9: 67,007,258; 1.75e-06); <b>Sire_Still_Birth</b> (9: 67,194,986; 6.09e-07)
<i>PTPRK</i>	6	127,968,779–128,520,674	<b>Age_at_Menarche</b> (rs6912749; 3.30e-06); <b>Intelligence</b> (rs11962619; 5.43e-06); <b>Allergic_Disease</b> (rs35469349; 2.29e-07)	9	66,968,686–67,594,209	<b>Fore_udder_att</b> (18: 43,639,816; 4.40e-06); <b>Prod_Life</b> (18: 43,618,348; 7.03e-06); <b>SCR</b> (18: 43,616,769; 5.47e-06); <b>SCS</b> (18: 43,591,719; 1.75e-08); <b>Udder_depth</b> (18: 43,639,816; 3.24e-06)
<i>RHPN2</i>	19	32,978,593–33,064,888	<b>Optic_cup_area</b> (rs11880141; 3.41e-07); <b>Optic_disc_area</b> (rs11880141; 3.41e-07)	18	43,574,291–43,644,737	<b>Fat_Percent</b> (6: 36,410,910; 2.88e-07); <b>Pro_Percent</b> (6: 36,410,910; 1.25e-10)
<i>SNCA</i>	4	89,724,099–89,838,315	<b>Neuroticism</b> (rs1372520; 2.10e-06); <b>Parkinson_disease</b> (rs356165; 9.28e-21)	6	36,285,494–36,432,426	

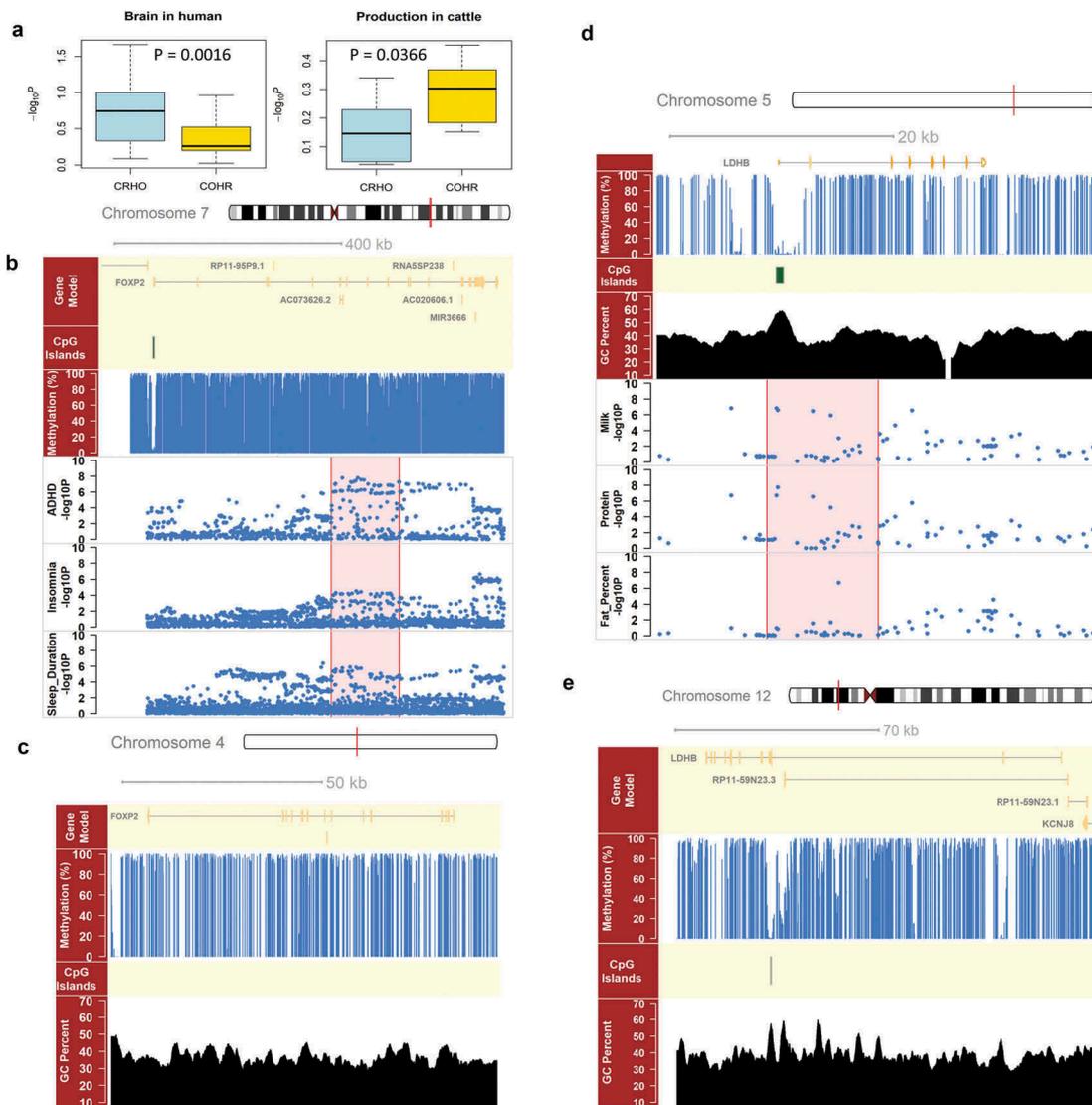
(Continued)

Table 1. (Continued).

Gene	Human			Cattle		
	Chr	Position	Trait	Chr	Position	Trait
<i>WNT7A</i>	3	13,816,258–13,880,121	<b>eGFRcrea</b> (rs6795744; 9.60e-09); <b>Rheumatoid_arthritis</b> (3: 13,870,727; 6.10e-06)	22	58,809,372–58,873,170	<b>Dtr_Calv_Ease</b> (22: 58,841,770; 1.11e-08); <b>Sire_Still_Birth</b> (22: 58,841,770; 1.96e-10); <b>Teat_length</b> (22: 58,816,379; 2.28e-06)
<i>ZBTB7B</i>	1	155,002,630–155,018,522	<b>Allergic_Disease</b> (rs1870940; 8.66e-06); <b>Birth_Weight</b> (rs3753639; 1.30e-12)	3	15,634,915–15,648,928	<b>Fat_Percent</b> (3: 15,638,565; 9.61e-14); <b>Milk</b> (3: 15,638,565; 9.48e-12); <b>Pro_Percent</b> (3: 15,638,565; 4.10e-76)



**Figure 4.** Comparison of *ANKS1A* and *TCAP* genes between human and cattle. (a) *ANKS1A* had a non-methylated promoter with a CpG island (CGI) on chromosome 6 in human, and bore suggestive significant SNPs ( $P < 1e-5$ ) for three human complex traits, including height, waist circumference (WC) and high-density lipoprotein (HDL). (b) *ANKS1A* had a non-methylated promoter with a CGI on chromosome 23 in cattle, and bore suggestive significant SNPs for two cattle complex traits, i.e., stature and udder depth. (c) *TCAP* had a hypermethylated promoter without any CGI on chromosome 17 in human, and bore suggestive significant SNPs for three immune-related traits, including inflammatory bowel disease (IBD), primary biliary cirrhosis (PBC) and urinary metabolites (UM). (d) *TCAP* had a hypermethylated promoter without CGI on chromosome 19 in cattle, and bore suggestive significant SNPs for somatic cell sore (SCS) in cattle.



**Figure 5.** GWAS single enrichment of species-specific genes and comparison of *FOXP2* and *LDHB* between human and cattle. (a) Comparison of GWAS single enrichment for species-specific genes in species-specific traits, i.e., brain-related traits in human and milk production traits in cattle. CRHO were genes with hypermethylated promoters in cattle but hypomethylated ones in human, whereas COHR were genes with hypomethylated promoters in cattle but hypermethylated ones in human. (b) *FOXP2* had a non-methylated promoter with a CpG island (CGI) on chromosome 7 in human, and bore suggestive significant SNPs ( $P < 1e-5$ ) for three brain-related traits, including sleep duration, insomnia and attention deficit hyperactivity disorder (ADHD). (c) *FOXP2* had a hypermethylated promoter without any CGI on chromosome 4 in cattle. (d) *LDHB* had a non-methylated promoter with a CGI on chromosome 5 in cattle, and bore suggestive significant SNPs ( $P < 1e-5$ ) for three cattle production traits, including milk yield, protein yield and fat percentage. (e) *LDHB* had a hypermethylated promoter without any CGI on chromosome 12 in human.

and evolution [46–48], we explored the associations of sequence evolution with sperm methylation evolution. We first estimated the SNP density of C/A, C/G, and C/T over all the CpG sites, which represented CpG mutation rates in the recent populations [49]. As expected we found that promoters of nMeth-genes had much lower mutation rates than those of hyper-genes, consistent in human and cattle. Whereas, compared to COHR, promoters of

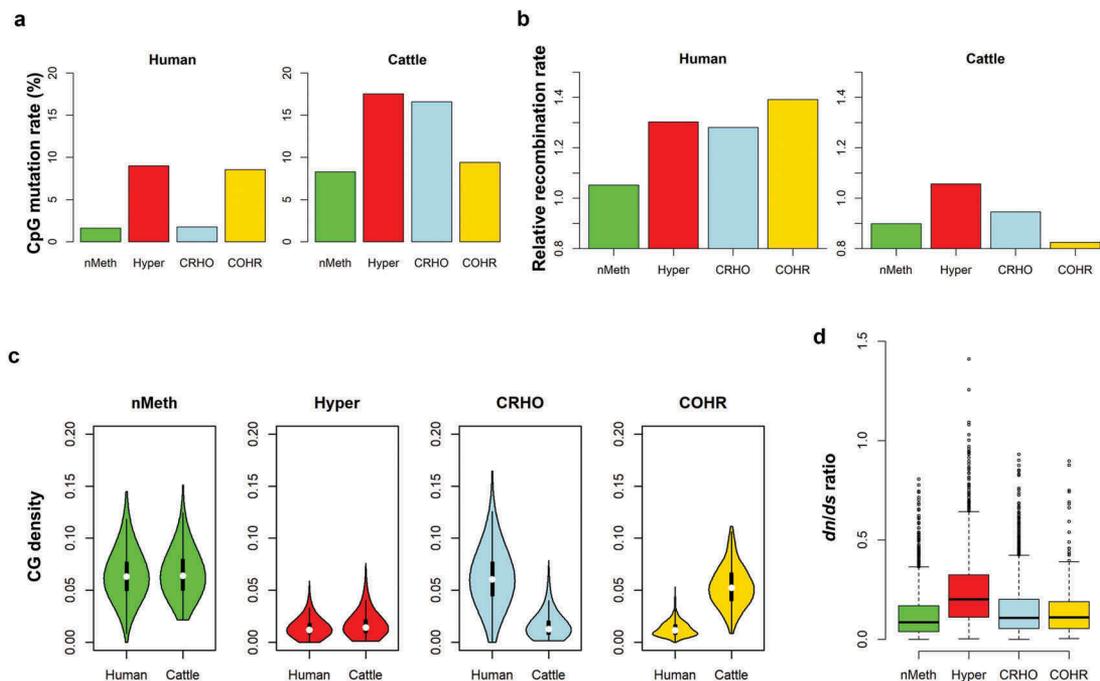
CRHO had lower mutation rates in human, but higher in cattle (Figure 6(a)). We then examined recombination maps of human [50] and cattle [51], where we used the recombination rate of the SNP that was within or in closest proximity to a promoter to represent the corresponding gene. Consistent with mutation rates, we found that nMeth-genes had lower recombination rates in average than hyper-genes in both human and cattle,

while species-specific genes exhibited opposite trends between human and cattle (Figure 6(b)). We further observed that non-methylated promoters had higher CpG densities than hypermethylated ones, and conserved promoters had similar CpG densities between human and cattle, whereas species-specific promoters had significantly diverged CpG densities between human and cattle (Figure 6(c)). All these implied that species-specific promoters experienced distinct CpG depletion pressures after the divergence of human and cattle. Furthermore, we demonstrated that protein-coding regions of those four gene-sets experienced distinct selective constraints through examining the paired  $dn/ds$  ratio between human and cattle within the last ~90 million years. The nMeth-genes had the lowest  $dn/ds$  values, suggesting that they were under strong purifying selection, while hyper-genes had the highest  $dn/ds$  values, indicating that they evolved fast. On the other hand, species-specific genes methylation exhibited intermediate  $dn/ds$  values (Figure 6(d)). All these evidences here implied that genome and

epigenome, particularly in genic and promoter regions, evolved together to affect complex phenotypes in mammals.

## Discussion

To our knowledge, this study for the first time investigated the associations of epigenetic evolution with complex phenotypes in mammals with such large-scale GWAS results of cattle and human. We revealed multiple epigenetic features of biological and evolutionary importance in sperm, with apparent evolutionary impacts in the underlying DNA sequence, which were enriched for GWAS signals of complex traits and diseases. Our findings also, for the first time, showed that genes/genome regions under epigenome evolution in sperm were associated with lineage-specific phenotypic variation, potentially contributing to the environmental adaptation. For instance, genes with human-specific hypomethylated promoters were associated with neuro-system development



**Figure 6.** Sequence evolutionary features accompany the sperm methylome. (a) Comparison of CpG mutation rates in promoters of the four gene-sets in human and cattle. nMeth represented genes with conserved non-methylated promoters in both species; hyper represented genes with conserved hypermethylated promoters in both species; CRHO were genes with hypermethylated promoters in cattle but hypomethylated ones in human; whereas COHR were genes with hypomethylated promoters in cattle but hypermethylated ones in human. (b) Comparison of recombination rates. We used the recombination rate of the SNP that was within or in closest proximity to a promoter to represent the corresponding gene. The relative recombination rate was calculated as the proportion of the average recombination rate of a gene-set over the average genome-wide recombination rate. (c) Comparison of CG density. (d) Comparison of  $dn/ds$  ratios in protein coding regions.

and brain-related diseases in human, whereas genes with cattle-specific hypomethylated promoters were relevant with lipid and protein metabolism. The results here provided new biological and evolutionary insights into the genetic mechanism underlying complex traits and diseases in mammals.

We found that genes with conserved non-methylated promoters in sperm highly intersected the sperm-retained nucleosomes, histones and pre-implantation activated genes, as well significantly enriched for GWAS signals of body conformation traits in both human and cattle. This was consistent with previous findings that the nucleosome and histones retained in sperm were enriched at developmental genes, whose promoters were generally hypomethylated in sperm, such as microRNA clusters and HOX gene clusters [32], although the majority of nucleosomes are replaced by protamine during sperm maturing in human. Jiang et al. (2013) pinpointed that DNA methylation in sperm not oocyte was inherited by embryos in Zebrafish [52], and Wang et al., (2014) also reported that the methylation levels of a fraction (~6.8%) of GpG sits in mouse sperm retained across embryonic developmental stages [13]. Additionally, our findings showed that sperm HMRs were significantly enriched for GWAS signals of multiple brain-related traits in human, such as neuroticism and depressive symptoms. This was consistent with previous reports that sperm DNA methylation alterations in older fathers might contribute to the increased incidence of neuropsychiatric and other diseases in their offspring [6,8]. Together, the sperm DNA methylome not only facilitates mature gamete function, but also guides the early embryogenesis and influences the later life [6,8,32,53,54].

Sperm DNA methylome affects the evolution of mammalian genomes [18]. Our results showed that genes with conserved hypermethylated promoters in sperm had higher mutation and recombination rates, and higher  $dn/ds$  ratios but lower CpG density than genes with conserved hypomethylated promoters, consistent with previous findings that the transition rates of methylated CpG to TpG mutation rates was ~10 fold higher than other dinucleotides [48]. We also demonstrated that species-specific hypomethylated genes had higher CpG densities than their orthologous genes in other species,

which was in agreement with a previous study that proposed even small differences in methylation may result in substantial loss of CpGs over a relatively short evolutionary period (i.e., divergence time between chimpanzee and human) [17]. One central evolutionary regime proposed that the CpG richness is driven by the low DNA methylation and consequentially by the low CpG deamination rates [55]. Together, it generally agrees with the hypothesis that the high-CG density in the genome may arise as the result of protection from methylation-induced mutations over long evolutionary periods. We currently observed that genes with hypomethylated promoters in sperm exhibited low recombination rates, which was in line with a previous observation that early developmental genes tended to have low recombination rates in their regulatory domains [56]. Additionally, we found that protein coding regions of nMeth-genes (i.e., essential genes) had much lower  $dn/ds$  values than that of hyper-genes, consistent with previous evidence that essential genes whose mutations were under strong purifying selection and thereby evolved slowly, whereas non-essential genes were under relaxed purifying selection, and thus evolved faster [57,58]. These are all in line with the history records that both human and cattle experienced many infectious diseases, and thus the immune genes of cattle and human might be evolved fast to deal with the fast evolution of pathogens [59].

Furthermore, our current study showed that promoters with species-specific methylation were enriched for different types of TF binding motifs, consistent with that evolutionary alterations in TF binding motifs could influence the shaping of methylome among species [20]. Previous comparisons of sperm methylome between human and chimpanzee revealed that genes associated with human-specific HMRs were selectively and functionally related to neuronal functions [17]. This is in line with our findings that genes with species-specific hypomethylated promoters between human and cattle are enriched for GWAS signals of species-specific traits. It is intriguing to hypothesize that alterations in the epigenetic state or genome regulatory state may allow phenotypic flexibility in a short run along one lineage, and the selective pressure in genome then drives the subsequent sequence changes in a long run if a trait shows advantages

in adaptation [17]. A current study also supported this by suggesting that during the evolution of heteromorphic chromosomes in the white-throated sparrow, the rapid and substantial regulatory evolution (i.e., transcriptome evolution) prevails before the large-scale genetic degeneration [60]. Of note, in the current paper we have focused on the promoters of genes, as the intergenic regulatory elements (e.g., enhancers) were not available in cattle yet. Also the intergenic regulatory elements are less conserved than promoters. It would be of interest to investigate the epigenetic evolution in the intergenic regulatory regions when more functional genomic data available in cattle, such as histone modification and chromatin interaction data.

## Materials and methods

### *Sperm collection and sequence library preparation*

No animal experiments were performed in this study, and ethics committee approval was therefore not required. References are provided where animal data were used.

We sampled 10 semen straws from 8 fertile, health and representative U.S. Holstein bulls. These semen samples were collected from bulls by an artificial insemination company using a standardized procedure with artificial vaginas. Each ejaculate normally had 4–5 billion sperm cells with a high mobility in a volume of 5 ml. The ejaculate was then diluted with a medium (extender), which included yolk from hens' eggs or heat-treated whole milk and glycerol. The distributed semen units/straws (0.5 ml, typically containing 10 to 40 million each), were transported and stored in liquid nitrogen tanks. After thawing and washing away the extender, we did visual examination of selected sperm samples under a microscope and found somatic cells usually were less than 1% and over 90% of sperm cells are morphologically normal. We isolated genomic DNA using QIAamp DNA Mini Kit protocol (QIAGEN, Valencia, CA, USA), and evaluated the quality of isolated DNA using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). We constructed the libraries using qualified genomic DNA as described previously in details [36], then sequenced using HiSeq X Ten (Illumina, San Diego,

CA, USA) with a 150bp paired-end technology. Details of the sample collection and WGBS procedure for human and mouse sperm methylome data can be found in [25] and in [31], respectively. We reanalyzed human and mouse sperm methylomes data using the same following procedure as cattle sperm methylomes.

### *Raw data profiling and methylation calling*

We employed FastQC v 0.11.2 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Trim Galore v 0.4.0 ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) to check the sequence data quality and to clean the data, respectively [36]. Generally, the adapters were removed, and the reads with low quality ( $Q < 20$ ) and shorter than 20 bp were filtered. We aligned the cleaned data to the reference genome UMD 3.1 for cattle sperm data, hg19 for human, and mm10 for mouse, respectively, using bowtie2 [61]. We then applied Bismark software to extract methylcytosine information [62]. All the details have been described previously [36]. The promoter regions were defined as the 1000bp down and upstream of transcriptional start sites (TSS) in the current study.

### *Hypomethylated region (HMR) detection*

We used a strictly method to detect the HMR as described previously [36]. Briefly, we scanned the whole genome using a sliding window approach with a window-length of 200bp and a step-size of 50bp. Bases on the binomial distributions of methylation across the genomic features (Fig. S1B), we considered a contiguous region with no less than 80% hypomethylated CpGs (i.e., the average regional methylation  $\leq 20\%$ ) as a HMR. We first detected HMRs in each individual separately, then intersected HMRs among individuals within each species to obtain the common HMRs with high confidence. We only kept the common HMRs with at least five CpG sites with each CpG site was covered by more than five reads for the downstream analyses [36].

### GWAS summary statistics

Details of the GWAS analysis in the cattle were described previously [63]. In brief, a linear mixed model, implemented in MMAP (<https://mmap.github.io/>), was employed to conduct single-marker GWAS analysis for 35 complex traits of 27,214 Holstein bulls using imputed sequence variants (~ 3 million SNPs). The model measured additive effects of genotypes while accounting for the population structure with a genomic relationship matrix. The phenotypes currently analyzed were de-regressed breeding values that have been accounted for all known systematic effects. We classified the cattle complex traits into three phenotypic categories, including 17 body type, 12 reproduction and 6 production traits. For human GWAS data, we obtained the summary statistics for 60 complex traits with an average sample size of 128,848 and an average SNP number of 5,905,874. We classified the 60 complex traits into five phenotype categories [64,65], including 14, 4, 10, 24 and 8 body development, reproduction, metabolism, brain-related and immune-related traits, respectively. Details of human GWAS studies are summarized in Supplementary Table S6.

### GWAS signal enrichment analysis based on detected epigenetic features

Since the complex phenotypes being studied are highly polygenic or even omnigenic [66], we employed the following sum-based marker-set test approach to examine the enrichment of GWAS signals in a given genomic features (e.g., a list of HMRs or genes). Previous studies demonstrated that this approach had higher power or at least equal to many commonly used marker-set test methods (e.g., count-based, score-based and covariance-based) in human [67], *Drosophila melanogaster* [68] and livestock [69-71], particularly in the highly polygenic traits.

$$T_{sum} = \sum_{i=1}^{m_f} t^2, \quad (1)$$

In which  $m_f$  is the number of markers within a genomic feature, and  $t^2$  is the square of  $t$  that was computed as the marker effect ( $b$ ) divided by the corresponding standard error. Here SNPs

within different elements (e.g., genes) of a genomic feature were often not in linkage disequilibrium (LD), as they may scattered distantly or even on different chromosomes. This method is similar to the popular linkage disequilibrium (LD) score regression [72], it analysed the genome-wide polygenic signals rather than a subset of SNPs that pass a certain significance threshold. It controlled LD patterns among SNPs and SNP-set sizes through applying the following cyclical permutation strategy, as described previously [67,68]. Briefly, we first ordered the test statistics (i.e.,  $t^2$ ) for all markers on the basis of their physical positions (i.e.,  $t_1^2, t_2^2, \dots, t_{m-1}^2, t_m^2$ ). We then randomly chose one test statistic (i.e.,  $t_k^2$ ) from this vector as the first, and shifted the remaining test statistics to new locations, while maintained their original orders (i.e.,  $t_k^2, t_{k+1}^2, \dots, t_m^2, t_1^2, \dots, t_{k-1}^2$ ). Thus, we uncoupled associations of SNPs with a genomic feature while retaining the correlation patterns among test statistics of SNPs. We computed a new summary statistic for the genomic feature being analysed on the basis of its original chromosome position. We repeated the permutation procedure 10,000 times for each genomic feature, and obtained an empirical  $P$ -value using one-tailed tests of the proportion of random summary statistics greater than that observed. We applied two types of marker-set test approaches in the current study based on different null hypotheses: 1) genome-wide marker-set test aimed to compare the makers within a genomic feature to random markers that were drawn from the whole genome (i.e., genic and intergenic regions), 2) genic marker-set test aimed to compare markers within a gene-set to random markers that were drawn from exclusively genic regions. The current marker-set test method together with multiple quantitative genomic tools were implemented in the QGG package (<http://psoerensen.github.io/qgg/>).

### Features of genome evolution

We computed the SNP density of C/A, C/G, and C/T over all the CpG sits to represent CpG mutation rates in the recent populations [21,49]. We examined recombination maps of human [50] and

cattle [51], where we used the recombination rate of the SNP that was within or in closest proximity to a promoter to represent the corresponding gene. We computed the relative recombination rates as the average recombination rate of a target gene-set divided by the overall average recombination rate of the genome.

### Gene-set functional annotation and motif enrichment analysis

We conducted the functional enrichment analysis for gene lists using R package clusterProfiler [73], where a hypergeometric test, based on the current GO database, was employed. We applied HOMER (<http://homer.ucsd.edu/homer/motif/>) to conduct the motif enrichment analysis for promoters of interest considering all the promoters in the genome as background. We adjusted *P*-values for multiple testing using the FDR method [74].

### Data availability

All the cattle sperm methylomes have been submitted to NCBI under GEO accession ID GSE119263 and GSE106538. The two human sperm methylome and one mouse sperm methylome data were obtained from Gene Expression Omnibus under accession ID GSE49624 and ([http://www.nodai-genome.org/mouse\\_en.html](http://www.nodai-genome.org/mouse_en.html)), respectively. All genomic annotation files, homologous gene-pairs, VCF files, and *dn/ds* ratios for human, cattle and mouse were downloaded from Ensembl database (<https://uswest.ensembl.org/index.html>). The GO annotation database can be publicly accessed (<https://bioconductor.org/packages/release/data/annotation/html/org.Bt.eg.db.html>).

### Acknowledgments

We thank Reuben Anderson, Alexandre Dimtchev, Zhiguang Li, Hongzheng Dai, Suzanne N. Martos, Yang Gao, Teng Li, and Zhibin Wang for technical assistances. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

### Competing interests

The authors declare that they have no competing interests

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work was supported in part by the Agriculture and Food Research Initiative (AFRI) grant numbers [2013-67015-20951] and [2016-67015-24886] from the USDA National Institute of Food and Agriculture (NIFA) Animal Genome and Reproduction Programs and the grant number [US-4997-17] from the United States - Israel Binational Agricultural Research and Development (BARD) Fund.

### ORCID

Derek M. Bickhart  <http://orcid.org/0000-0003-2223-9285>  
 Steven G. Schroeder  <http://orcid.org/0000-0001-9103-5150>  
 Benjamin D. Rosen  <http://orcid.org/0000-0001-9395-8346>  
 John B. Cole  <http://orcid.org/0000-0003-1242-4401>  
 Curtis P. Van Tassell  <http://orcid.org/0000-0002-8416-2087>  
 George E. Liu  <http://orcid.org/0000-0003-0192-6705>

### References

- [1] Barlow DP. Methylation and imprinting: from host defense to gene regulation? *Science*. 1993;260:309–311.
- [2] Bestor TH. Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes. *Am J Hum Genet*. 1998;62:1269–1273.
- [3] Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*. 2004;431:96.
- [4] Sharp AJ, Stathaki E, Migliavacca E, et al. DNA methylation profiles of human active and inactive X chromosomes. *Genome Res*. 2011;21:1592–1600.
- [5] Jenkins TG, Carrell DT. The sperm epigenome and potential implications for the developing embryo. *Reproduction*. 2012;143:727–734.
- [6] Jenkins TG, Aston KI, Pflueger C, et al. Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. *PLoS Genet*. 2014;10:e1004458.
- [7] Aston KI, Uren PJ, Jenkins TG, et al. Aberrant sperm DNA methylation predicts male fertility status and embryo quality. *Fertil Steril*. 2015;104(1388–97):e1–5.
- [8] Atsem S, Reichenbach J, Potabattula R, et al. Paternal age effects on sperm FOXP1 and KCNA7 methylation and transmission into the next generation. *Hum Mol Genet*. 2016;25:4996–5005.

- [9] Jenkins TG, Aston KI, Meyer TD, et al. Decreased fecundity and sperm DNA methylation patterns. *Fertil Steril*. 2016;105(51–7):e3.
- [10] Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet*. 2018;19:129.
- [11] Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. *Science*. 2001;293:1068–1070.
- [12] Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*. 2014;157:95–109.
- [13] Wang L, Zhang J, Duan J, et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell*. 2014;157:979–991.
- [14] Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016;17:61.
- [15] Van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016;7:11115.
- [16] Li Y, Zheng H, Wang Q, et al. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol*. 2018;19:18.
- [17] Molaro A, Hodges E, Fang F, et al. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell*. 2011;146:1029–1041.
- [18] Qu J, Hodges E, Molaro A, et al. Evolutionary expansion of DNA hypomethylation in the mammalian germline genome. *Genome Res*. 2018;28:145–158.
- [19] Gu X, Su Z. Tissue-driven hypothesis of genomic evolution and sequence-expression correlations. *Proc Nat Acad Sci*. 2007;104:2779–2784.
- [20] Hernando-Herraez I, Garcia-Perez R, Sharp AJ, et al. DNA methylation: insights into human evolution. *PLoS Genet*. 2015;11:e1005661.
- [21] Xia J, Han L, Zhao Z. Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. *BMC Genomics*. 2012;13:S7.
- [22] Weigmann K. Lifestyle in the sperm: there is growing evidence that epigenetic marks can be inherited. But what is the nature of the information they store and over how many generations do they prevail? *EMBO Rep*. 2014;15:1233–1237.
- [23] Yi SV. Insights into epigenome evolution from animal and plant methylomes. *Genome Biol Evol*. 2017;9:3189–3201.
- [24] Elsik CG, Tellam RL, Worley KC. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*. 2009;324:522–528.
- [25] Hammoud SS, Low DH, Yi C, et al. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell*. 2014;15:239–253.
- [26] Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics*. 2009;1:239–259.
- [27] Wagner JR, Busche S, Ge B, et al. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol*. 2014;15:R37.
- [28] Levine M, Tjian R. Transcription regulation and animal diversity. *Nature*. 2003;424:147.
- [29] Schlosberg CE, VanderKraats ND, Edwards JR. Modeling complex patterns of differential DNA methylation that associate with gene expression changes. *Nucleic Acids Res*. 2017;45:5100–5111.
- [30] Band MR, Larson JH, Rebeiz M, et al. An ordered comparative map of the cattle and human genomes. *Genome Res*. 2000;10:1359–1368.
- [31] Kobayashi H, Sakurai T, Imai M, et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet*. 2012;8:e1002440.
- [32] Hammoud SS, Nix DA, Zhang H, et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature*. 2009;460:473.
- [33] Xue Z, Huang K, Cai C, et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*. 2013;500:593.
- [34] Graf A, Krebs S, Zakhartchenko V, et al. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc Nat Acad Sci*. 2014;111:4139–4144.
- [35] Samans B, Yang Y, Krebs S, et al. Uniformity of nucleosome preservation pattern in Mammalian sperm and its connection to repetitive DNA elements. *Dev Cell*. 2014;30:23–35.
- [36] Zhou Y, Connor EE, Bickhart DM, et al. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. *GigaScience*. 2018;7.
- [37] Saitou M, Kurimoto K. Paternal nucleosomes: are they retained in developmental promoters or gene deserts? *Dev Cell*. 2014;30:6–8.
- [38] Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science*. 2015;347:1260419.
- [39] Hockett M, Almeida R, Rohrbach N, et al. Effects of induced clinical mastitis during preovulation on endocrine and follicular function. *J Dairy Sci*. 2005;88:2422–2431.
- [40] Hertl J, Gröhn Y, Leach J, et al. Effects of clinical mastitis caused by gram-positive and gram-negative bacteria and other organisms on the probability of conception in New York State Holstein dairy cows. *J Dairy Sci*. 2010;93:1551–1560.
- [41] Reese J, Das SK, Paria BC, et al. Global gene expression analysis to identify molecular markers of uterine receptivity and embryo implantation. *J Biol Chem*. 2001;276:44137–44145.
- [42] Kristiansen TZ, Nielsen MM, Blagoev B, et al. Mouse embryonic fibroblasts derived from Odin deficient mice display a hyperproliferative phenotype. *DNA Res*. 2004;11:285–292.

- [43] Gorenshiteyn D, Zaslavsky E, Fribourg M, et al. Interactive big data resource to elucidate human immune pathways and diseases. *Immunity*. 2015;43:605–614.
- [44] Lai CS, Fisher SE, Hurst JA, et al. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 2001;413:519.
- [45] Ostrowska M, Górka BŻ, Słoniewski K, et al. Expression of PC, PCK1, PCK2, LDHB, FBP1 and G6PC genes in the liver of cows in the transition from pregnancy to lactation. *Animal Sci Pap Rep*. 2013;31:281–290.
- [46] Ehrlich M, Norris KF, Wang RY, et al. DNA cytosine methylation and heat-induced deamination. *Biosci Rep*. 1986;6:387–393.
- [47] Zhao Z, Jiang C. Methylation-dependent transition rates are dependent on local sequence lengths and genomic regions. *Mol Biol Evol*. 2006;24:23–25.
- [48] Jiang C, Han L, Su B, et al. Features and trend of loss of promoter-associated CpG islands in the human and mouse genomes. *Mol Biol Evol*. 2007;24:1991–2000.
- [49] Stamatoyannopoulos JA, Adzhubei I, Thurman RE, et al. Human mutation rate associated with DNA replication timing. *Nat Genet*. 2009;41:393.
- [50] Kong A, Thorleifsson G, Gudbjartsson DF, et al. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*. 2010;467:1099.
- [51] Ma L, O'Connell JR, VanRaden PM, et al. Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet*. 2015;11:e1005387.
- [52] Jiang L, Zhang J, Wang -J-J, et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*. 2013;153:773–784.
- [53] Milekic M, Xin Y, O'donnell A, et al. Age-related sperm DNA methylation changes are transmitted to offspring and associated with abnormal behavior and dysregulated gene expression. *Mol Psychiatry*. 2015;20:995–1001.
- [54] Donkin I, Barrès R. Sperm epigenetics and influence of environmental factors. *Mol Metab*. 2018;14:1–11.
- [55] Cohen NM, Kenigsberg E, Tanay A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell*. 2011;145:773–786.
- [56] Liu Y, Sarkar A, Kheradpour P, et al. Evidence of reduced recombination rate in human regulatory domains. *Genome Biol*. 2017;18:193.
- [57] Hurst LD, Smith NG. Do essential genes evolve slowly? *Curr Biol*. 1999;9:747–750.
- [58] Jordan IK, Rogozin IB, Wolf YI, et al. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res*. 2002;12:962–968.
- [59] Behrman EL, Howick VM, Kapun M, et al. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc R Soc B*. 2018;285:20172599.
- [60] Sun D, Huh I, Zinzow-Kramer WM, et al. Rapid regulatory evolution of a nonrecombining autosome linked to divergent behavioral phenotypes. *Proc Nat Acad Sci*. 2018;115:2794–2799.
- [61] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
- [62] Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–1572.
- [63] Jiang J, Cole JM, Da Y, et al. Fast Bayesian fine-mapping of 35 production, reproduction and body conformation traits with imputed sequences of 27K Holstein bulls. *bioRxiv*. 2018;428227.
- [64] Lu Q, Powles RL, Abdallah S, et al. Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet*. 2017;13:e1006933.
- [65] Hao X, Zeng P, Zhang S, et al. Identifying and exploiting trait-relevant tissues with multiple functional annotations in genome-wide association studies. *PLoS Genet*. 2018;14:e1007186.
- [66] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177–1186.
- [67] Rohde PD, Demontis D, Cuyabano BCD, et al. Covariance Association Test (CVAT) identifies genetic markers associated with schizophrenia in functionally associated biological processes. *Genetics*. 2016;203:1901–1913.
- [68] Sørensen IF, Edwards SM, Rohde PD, et al. Multiple trait covariance association test identifies gene ontology categories associated with chill coma recovery time in *Drosophila melanogaster*. *Sci Rep*. 2017;7.
- [69] Sarup P, Jensen J, Ostersen T, et al. Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. *BMC Genet*. 2016;17:11.
- [70] Fang L, Sahana G, Ma P, et al. Exploring the genetic architecture and improving genomic prediction accuracy for mastitis and milk production traits in dairy cattle by mapping variants to hepatic transcriptomic regions responsive to intra-mammary infection. *Genet Sel Evol*. 2017;49:44.
- [71] Fang L, Sørensen P, Sahana G, et al. MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle. *Sci Rep*. 2018;8:9345.
- [72] Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228.
- [73] Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–287.
- [74] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B (Methodological)*. 1995;57:289–300.