

RESEARCH ARTICLE

Open Access



# Analyses of inter-individual variations of sperm DNA methylation and their potential implications in cattle

Shuli Liu<sup>1,2†</sup>, Lingzhao Fang<sup>2,3,4†</sup>, Yang Zhou<sup>5</sup>, Daniel J.A. Santos<sup>3</sup>, Ruidong Xiang<sup>6,7</sup>, Hans D. Daetwyler<sup>7,8</sup>, Amanda J. Chamberlain<sup>7</sup>, John B. Cole<sup>2</sup>, Cong-jun Li<sup>2</sup>, Ying Yu<sup>1</sup>, Li Ma<sup>3</sup>, Shengli Zhang<sup>1\*</sup> and George E. Liu<sup>2\*</sup> 

## Abstract

**Background:** DNA methylation has been shown to be involved in many biological processes, including X chromosome inactivation in females, paternal genomic imprinting, and others.

**Results:** Based on the correlation patterns of methylation levels of neighboring CpG sites among 28 sperm whole genome bisulfite sequencing (WGBS) data (486 × coverage), we obtained 31,272 methylation haplotype blocks (MHBs). Among them, we defined conserved methylated regions (CMRs), variably methylated regions (VMRs) and highly variably methylated regions (HVMRs) among individuals, and showed that HVMRs might play roles in transcriptional regulation and function in complex traits variation and adaptive evolution by integrating evidence from traditional and molecular quantitative trait loci (QTL), and selection signatures. Using a weighted correlation network analysis (WGCNA), we also detected a co-regulated module of HVMRs that was significantly associated with reproduction traits, and enriched for glycosyltransferase genes, which play critical roles in spermatogenesis and fertilization. Additionally, we identified 46 VMRs significantly associated with reproduction traits, nine of which were regulated by cis-SNPs, implying the possible intrinsic relationships among genomic variations, DNA methylation, and phenotypes. These significant VMRs were co-localized ( $\pm 10$  kb) with genes related to sperm motility and reproduction, including *ZFP36L1*, *CRISP2* and *HGF*. We provided further evidence that rs109326022 within a predominant QTL on BTA18 might influence the reproduction traits through regulating the methylation level of nearby genes *JOSD2* and *ASPDH* in sperm.

**Conclusion:** In summary, our results demonstrated associations of sperm DNA methylation with reproduction traits, highlighting the potential of epigenomic information in genomic improvement programs for cattle.

**Keywords:** Sperm DNA methylation, Methylation haplotype blocks, Variably methylated regions, Reproduction traits, Cattle

## Background

Emerging evidence shows that the sperm DNA methylation contributes to not only male fertility but also to early embryo development [1–4]. DNA methylation in sperm has been shown to be involved in many biological processes, including X chromosome inactivation in females,

paternal genomic imprinting, silencing of transposable elements and DNA compaction [5–8]. Some alterations of sperm DNA methylation may persist into the early embryo and thus influence the transcriptome and epigenome in somatic tissues [9, 10], leading to variation in phenotypes of offspring [11]. In dairy cattle breeding, we use the breeding value (e.g., predicted transmitting ability, PTA) of a sire to measure his contribution to complex traits of offspring, after correcting for all known systematic effects. An elite bull often has thousands of daughters due to artificial insemination, yielding high reliable phenotypes (i.e., estimated breeding values). This offers a valuable source

\* Correspondence: [zhangslcau@cau.edu.cn](mailto:zhangslcau@cau.edu.cn); [george.liu@usda.gov](mailto:george.liu@usda.gov)

<sup>†</sup>Shuli Liu and Lingzhao Fang contributed equally to this work.

<sup>1</sup>College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

<sup>2</sup>USDA-ARS, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705, USA

Full list of author information is available at the end of the article



for understanding the relationships between sperm DNA methylation and complex traits in mammals, particularly in males.

Variation of DNA methylation among individuals has been speculated to affect susceptibility to complex diseases and resistance to drug treatment in human [12–14]. The epigenetic polymorphism, termed “variably methylated regions” or VMRs, were found to be enriched in various functional genomic features, like enhancers, CpG shores, 3'UTR etc., indicating their potential roles in transcriptional regulation [15, 16]. In addition, inter-individual methylation variations have been demonstrated to be associated with tissue-specific function and environmental adaptation [15]. For instance, VMRs within co-methylated networks in fibroblasts were enriched for four clusters of HOX genes. Furthermore, both genetic factors and environmental exposures like diet, stress, toxic exposure and exercise contribute to epigenetic variation [17–19]. Analysis of VMRs in human neonatal blood samples further indicated that VMRs were best explained mainly by either environmental factors and genotype interaction (GxE) or their additive effects (G + E) [20]. Additionally, SNPs involved in the significant GxE models were highly enriched with signals of genome wide association studies (GWAS) for complex diseases [20]. Studies have also revealed that, by targeting VMRs, the statistical power can be improved in epigenetic signature detection using epigenetic association studies (EWAS) [21]. While VMRs were studied in human and model organisms, to our knowledge, no studies have been published to investigate inter-individual variation of DNA methylation in cattle, particularly in sperm, which is of importance in dairy cattle due to the wide use of artificial insemination technology.

There are different methods to detect VMRs. In previous human studies, VMRs were determined by either merging the adjacent highly variable CpG sites within predefined windows [15, 22] or using the highly variable restriction enzyme (*Msp* I) fragments from reduced representation bisulfite sequencing. However, methylation levels of each CpG site within VMRs may be variable, raising a question of which CpG site to choose. For example, the most variable CpG site (tagCpG) within each VMR was often selected to represent it [15, 22]. This strategy may miss the information provided by other CpG sites. Some researchers directly utilized the single CpG site [21], which may be influenced by the technical noise and sensitivity in measuring single CpG methylation [23]. Because adjacent CpG sites tend to show the coordinated methylation statuses due to the progressivity of the DNA methylation and/or demethylation enzymes (DNMT1, DNMT3A/B and TET proteins), these co-

methylated CpG sites may form methylation haplotypes, called methylation haplotype blocks (MHBs) [23]. A previous study showed that MHBs tend to be enriched in VMRs and functional regions [23]. Here, we hypothesized that the utilization of MHBs could improve the definition of VMRs by concentrating on biologically relevant regions, and enhance statistical power by reducing the multiple testing burden compared to the single-CpG method.

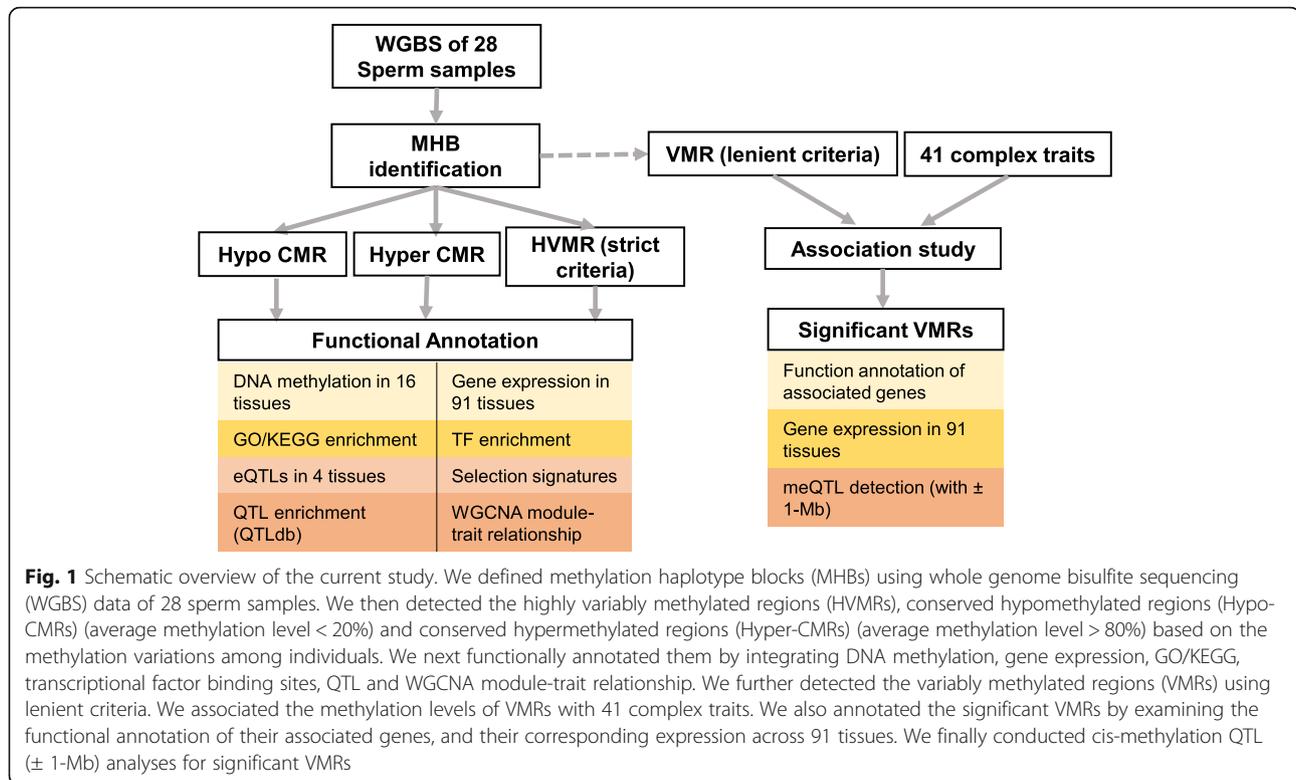
In this study, we functionally annotated the VMRs using a range of other information, including gene expression, predicted transcription factor binding sites, traditional quantitative trait loci (QTL), gene expression QTL and selection signatures (Fig. 1). We further associated VMRs with 41 complex traits and further explore the cis-SNPs that may regulate the detected significant epigenetic signatures within  $\pm 1$ -Mb. To our knowledge, this is the first attempt to reveal the potential impact of epigenomics on phenotypes on genome wide level.

## Results

### Identification and characterization of methylation haplotype blocks

Our correlation analysis of 28 WGBS data within various genomic elements, revealed that global methylation was highly conserved among individuals, i.e. Pearson correlations ranged from 0.914 to 0.995 (Additional file 1: Figure S1a). The top conserved genomic elements included 5' UTRs, CG islands and promoters, while introns, 3'UTRs and exons were relatively dynamic among individuals.

We calculated a pairwise “linkage disequilibrium” of CpG methylation (LD,  $r^2$ ) as previously reported [23], and partitioned the mappable genome (coverage  $\geq 10$ ; minimal size: 80 bp) into blocks using LD ( $r^2$ ) cutoff of 0.5 with at least 3 CpG sites within a block. We identified 31, 272 MHBs (Additional file 2: Table S1) with an average size of 52 bp (Fig. 2a), and an average of 12 CpG sites per 100 bp (Fig. 2b). Pearson correlation analysis showed that methylation levels within MHBs were less correlated among individuals (0.52 ~ 0.86) compared to those of various genomic elements (Additional file 1: Figure S1b), which was concordant with the previous observation that MHBs were variable among individuals and highly enriched in VMR [15]. The MHBs also overlapped with multiple known genomic elements (Fig. 2c). Among all the MHBs, 64.6, 35.4% were located in intergenic regions or transcribed regions, respectively. The MHBs were highly enriched in CpG islands, 5'UTRs, exons and promoters (1000 times of permutation test using RegioneR [24],  $P < 0.001$ ), indicating that they may play important roles in transcriptional regulation (Fig. 2d). Based on the 15 chromatin states in bovine rumen cells predicted using histone modifications (H3K4me3, H3K4me1, H3K27ac, H3K27me3) and other epigenome information (ATAC-



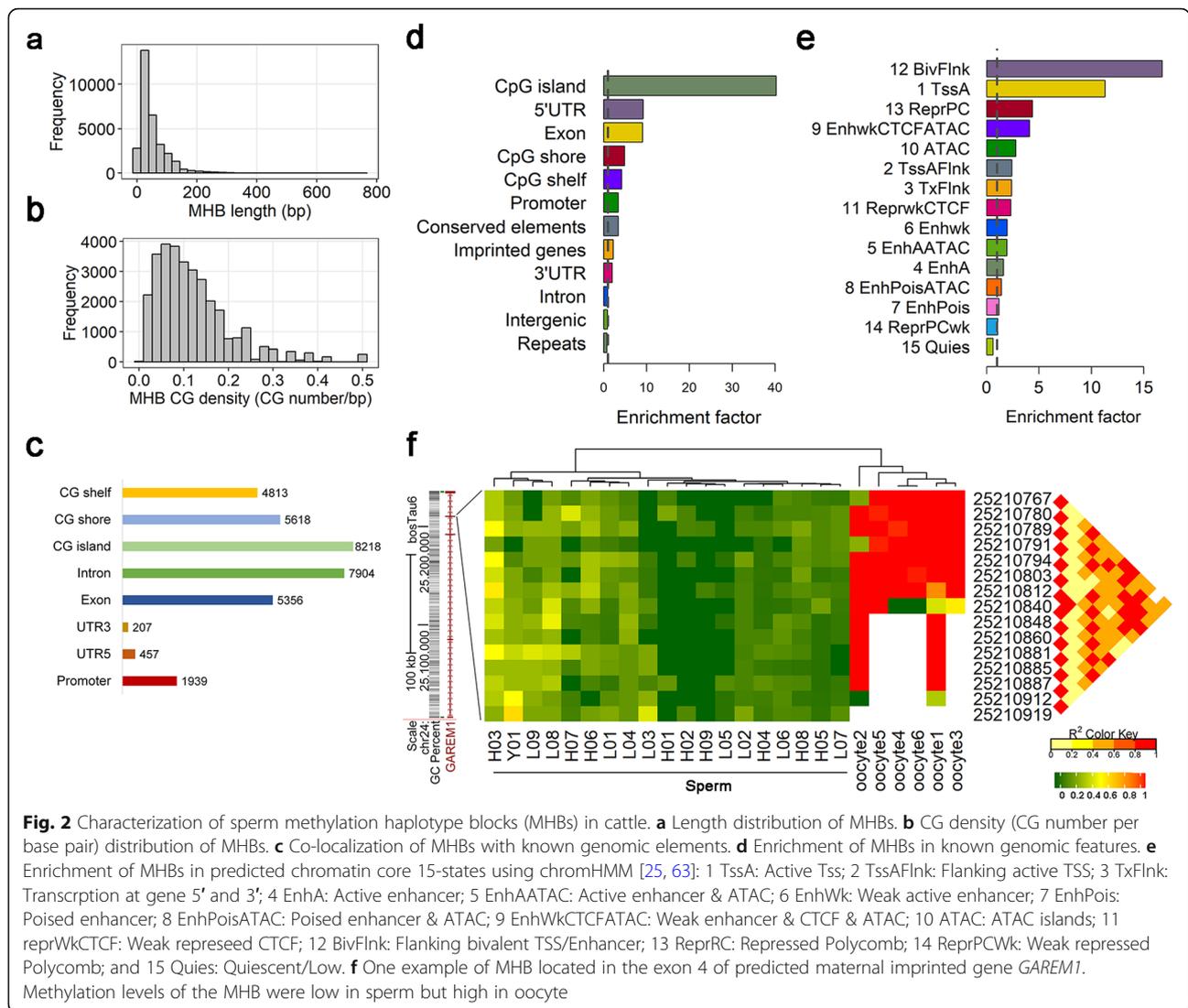
seq and CTCF binding sites) [25], we observed MHBs had a significant enrichment for flanking bivalent TSS/enhancer (enrichment factor: 17) and active TSS i.e. promoters (enrichment factor: 11) (Fig. 2e). We also observed an enrichment of MHBs in imprinted genes (enrichment factor: 2.22) like *HOXA* genes, *IGF2*, and *IGF2R*. One example is the methylation block containing 15 CpG sites in exon 4 of a predicted maternal imprinted gene *GAREMI* (Fig. 2f). We observed distinct methylation patterns of this region between sperm and oocytes [26], where sperm were consistently lowly methylated while oocytes were highly methylated. Collectively, blocks with coordinately methylated 5mC were likely to capture the epigenetic signatures associated with transcriptional regulation.

To explore the conservation of MHBs among species, we compared the detected MHBs between cattle and human by converting the MHB coordinates to the human hg19 genome using the liftOver tool in the UCSC browser. It is noted that human MHBs were predicted using human somatic tissues instead of sperm, which were published before [23]. Out of all the bovine MHBs, 51.8% were successfully converted with the minimal match of 0.8. A total of 1952 bovine MHBs were overlapped with those from human [23], and associated with 1701 human genes, while the cattle-specific MHBs were associated with 5832 genes. Interestingly, genes overlapping with common MHBs between human and cattle (HCMHBs) showed an enrichment of

GO terms related to early embryonic development, while those associated with cattle-specific MHBs (CMHBs) were mainly enriched in the development of nervous system (FDR < 0.05, Additional file 1: Figure S2a). We also found HCMHBs overlapped with more genes (Additional file 1: Figure S2b) and were more conserved among mammals (indicated as the PhastCon scores, <https://genome.ucsc.edu/goldenPath/help/phastCons.html>) than CMHBs (Additional file 1: Figure S2c). Further study is required to understand the biological mechanisms underlying the evolution of MHBs among species.

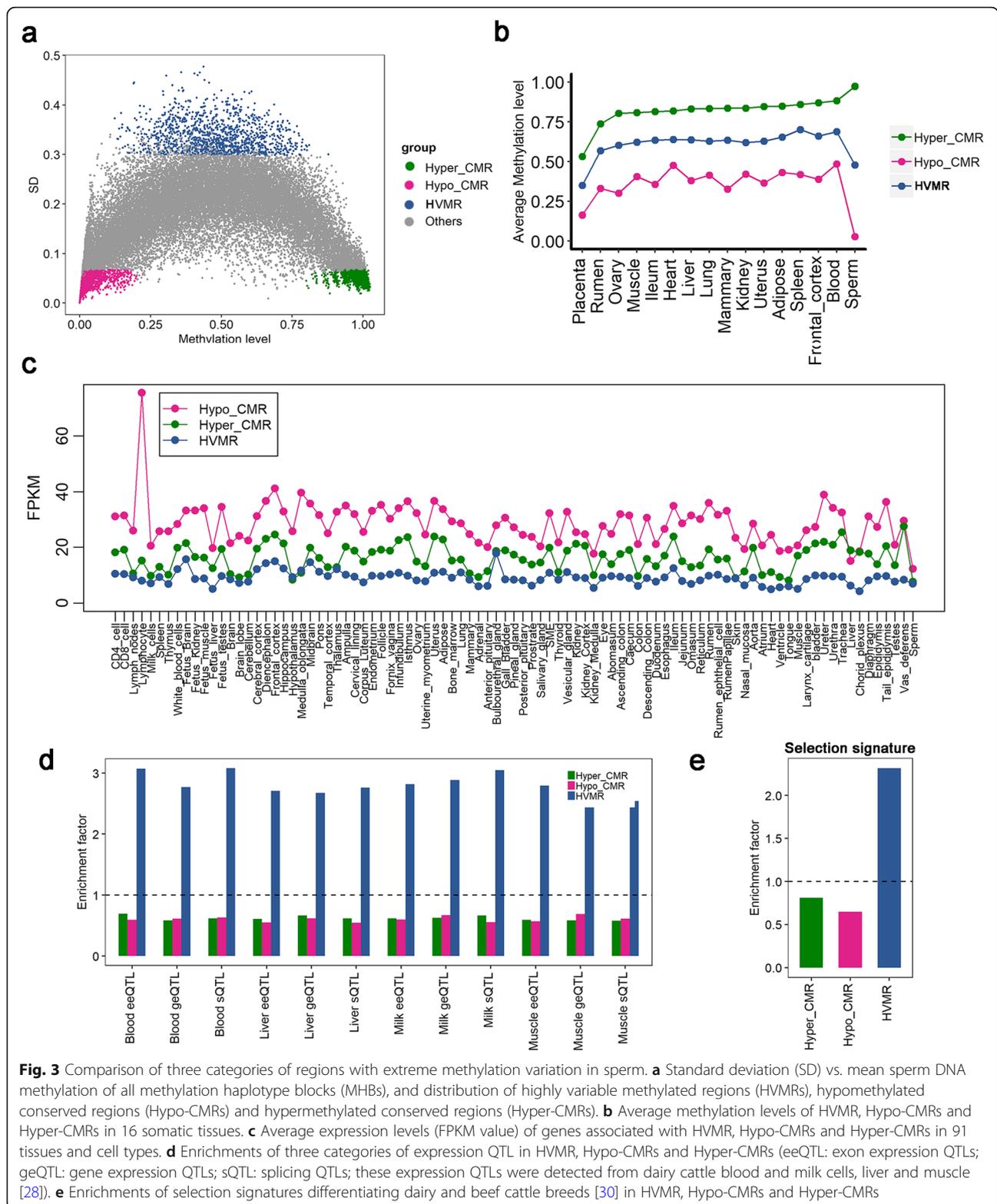
#### Inter-individual variation and conservation in sperm DNA methylome

MHBs were previously shown to have an extremely high enrichment in VMRs [23]. To characterize the methylation dynamics among individual bulls, we defined three categories of regions from MHBs in terms of their methylation variations (See Methods), 1) highly variable methylation regions (HVMRs,  $n = 1681$ ) with extremely high methylation variations, 2) conserved hypomethylated regions (hypo-CMRs, Average methylation level  $\leq 0.2$ ,  $n = 3371$ ), 3) conserved hypermethylated regions (hyper-CMRs, Average methylation level  $\geq 0.8$ ,  $n = 1594$ ) (Additional file 2: Table S2). As expected, moderately methylated MHBs were more likely to be variable among individuals than highly or lowly methylated MHBs (Fig. 3a). We further observed that methylation



differences between HVMRs and CMRs persist into multiple somatic tissues, such as rumen, ovary and placenta (Fig. 3b). We collected bovine transcriptomes of 91 tissues/cells from published data of previous studies. Similarly, expression differences between HVMR- and CMR-associated genes (genes overlapped with MHB regions in promoters and gene bodies) were also consistent in 91 bovine tissues/cells. Genes associated with hypo-CMRs generally had the highest expression, followed by genes associated with hyper-CMRs and genes associated with HVMRs (Fig. 3c). Functional annotation further revealed that genes associated with hypo-CMRs were engaged in basic cell-function, including transcription, DNA binding and nuclear chromatin (Additional file 1: Figure S3a), suggesting that these genes were mainly house-keeping genes. For example, Hypo-CMRs were enriched for developmental motifs, like HOXD13, and motifs involved in cell proliferation and differentiation, like MYB, KHDRBS2 and SRSF10. Hyper-CMRs were enriched

in motifs associated with hemopoietic development and alternative splicing, like MZF1 and CELF2. In contrast, we speculated that HVMRs may be more likely to harbor tissue-specific expressed genes. Furthermore, our transcription factor binding motif analysis validated that the HVMRs were enriched for motifs with divergent functions, such as ZNF711 for cognitive disability, PBX3 for leukemia and PKNOX1 for adult spermatogenesis [27] (Additional file 1: Figure S3b). Additionally, human orthologous genes in HVMRs were also enriched in tissue-specific expressed genes (enrichment factor: 1.84;  $P = 5.52 \times 10^{-6}$ ; Fisher's exact test), while human orthologous genes in hypo-CMRs and hyper-CMRs were enriched in house-keeping functions (enrichment factor: 1.49 and 1.64;  $P = 7.01 \times 10^{-8}$  and  $6.86 \times 10^{-7}$ ) (Additional file 1: Figure S3c). We further overlapped HVMRs, hypo- and hyper-CMRs with multiple types of expression QTLs, including splicing QTLs (sQTLs), gene expression QTLs (geQTLs) and exon expression QTLs



**Fig. 3** Comparison of three categories of regions with extreme methylation variation in sperm. **a** Standard deviation (SD) vs. mean sperm DNA methylation of all methylation haplotype blocks (MHBs), and distribution of highly variable methylated regions (HVMRs), hypomethylated conserved regions (Hypo-CMRs) and hypermethylated conserved regions (Hyper-CMRs). **b** Average methylation levels of HVMR, Hypo-CMRs and Hyper-CMRs in 16 somatic tissues. **c** Average expression levels (FPKM value) of genes associated with HVMR, Hypo-CMRs and Hyper-CMRs in 91 tissues and cell types. **d** Enrichments of three categories of expression QTL in HVMR, Hypo-CMRs and Hyper-CMRs (eeQTL: exon expression QTLs; geQTL: gene expression QTLs; sQTL: splicing QTLs; these expression QTLs were detected from dairy cattle blood and milk cells, liver and muscle [28]). **e** Enrichments of selection signatures differentiating dairy and beef cattle breeds [30] in HVMR, Hypo-CMRs and Hyper-CMRs

(eeQTLs), from previous analyses [28, 29]. All three types of QTLs were highly enriched in HVMRs ( $P < 2.2 \times 10^{-16}$ , Fisher's exact test), but depleted in hypo-CMRs and hyper-

CMRs ( $P < 2.2 \times 10^{-16}$ ) (Fig. 3d), which were consistent among all four tissues (muscle, liver, blood and milk cells). This result indicates that genomic regions associated with

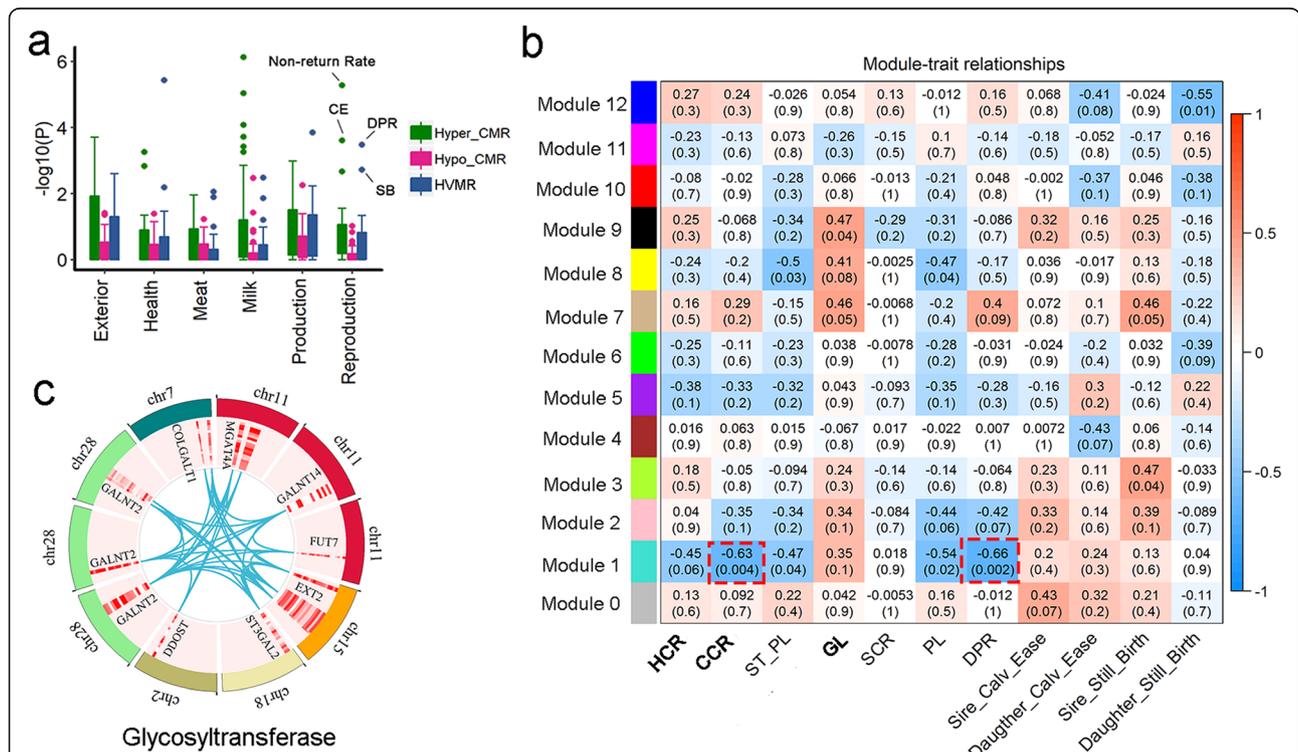
sperm HVMRs could play key roles in regulation of gene expression and splicing among individuals. We further found an enrichment of HVMRs ( $P < 2.2 \times 10^{-16}$ ) in selection signatures with higher frequency in dairy rather than in beef breeds [30], suggesting that HVMRs might play roles in positive selection and adaptive evolution (Fig. 3e). Collectively, our data revealed distinct DNA methylation variation patterns in sperm might influence the transcriptional regulation and evolution.

**HVMRs were associated with QTLs of reproduction traits**

To explore the relationship between HVMRs in sperm and complex traits, we first examined the QTL regions of six categories of traits ( $n = 232$ ) (Exterior, health, Meat, milk, production and reproduction traits) from the Cattle QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). We observed that both hyper-CMRs and HVMRs had higher enrichments for QTL signals of complex traits than hypo-CMRs. Of note, HVMRs tend to be specifically and significantly enriched for QTL signals of reproduction traits, with the top associated traits were daughter pregnancy rate

(DPR; FDR = 0.03) and stillbirth (SB; FDR = 0.1). While hyper-CMRs were highly associated with a range of complex traits such as milk production traits, non-return rate and calving ease (CE) (Fig. 4a).

To further investigate the relationship between HVMRs and reproduction traits, we grouped the co-methylated HVMRs into 12 distinct modules using WGCNA analysis [31], and associated each module with 11 reproduction traits in our 19 bulls of similar age (1~2-year-old), after correcting for the genetic relatedness (Fig. 4b). We observed Module 1 was significantly ( $P < 0.05$ ) negatively correlated with several reproduction traits, including heifer conception rate (HCR), cow conception rate (CCR), standard length of productive life (ST\_PL), length of productive life (PL), and DPR, but positively correlated with gestation length (GL) and sire still birth. This was consistent with their negative genetic correlations as previously reported [4]. Interestingly, the 15 co-methylated MHBs in Module 1 were enriched for glycosyltransferase genes (FDR = 0.0046) (Fig. 4c). Although the exact mechanisms remain elusive, published results reported that glycosyltransferases were the main enzymes in glycosylation and responsible for the synthesis of glycans which play pivotal roles in spermatogenesis [32, 33].



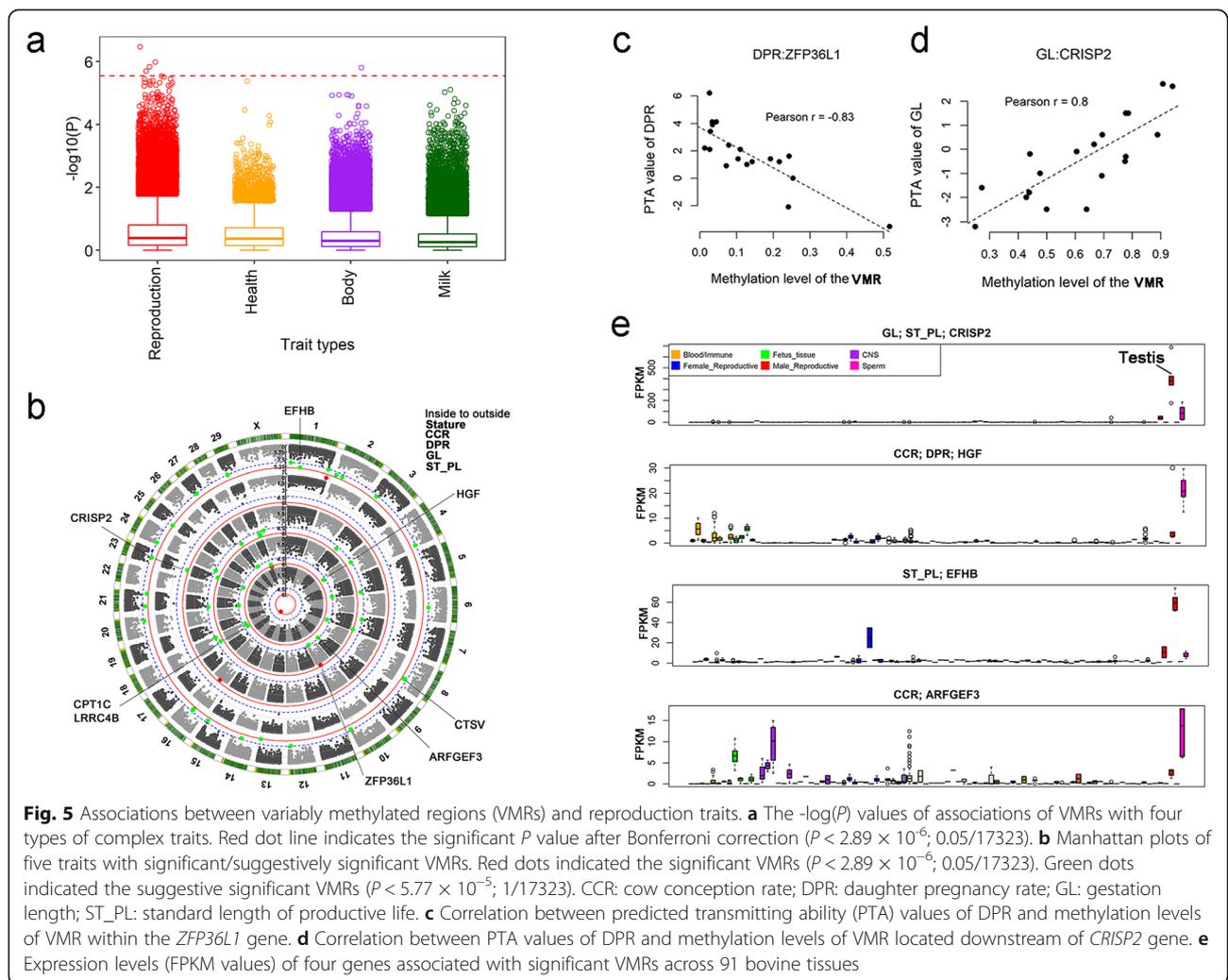
**Fig. 4** Relationship between methylation variations in sperm and complex traits. **a** Enrichments of six QTL categories (including 232 traits) from Cattle QTL database in HVMRs, Hypo-CMRs and Hyper-CMRs. CE: calving ease; DPR: daughter pregnancy rate; SB: still birth. **b** Module-trait relationships using a weighted correlation network analysis (WGCNA) (Only reproduction traits were tested). Elements in red dash box were two traits most significantly associated with module 1. HCR: heifer conception rate; CCR: cow conception rate; ST\_PL: standard length of productive life; GL: gestation length; SCR: sire conception rate; PL: length of productive life; DPR: daughter pregnancy rate. **c** Co-methylated regions with 15 MHBs in module 1 enriched for glycosyltransferase genes. M1, M2 and M3 represents three MHBs located with the GALNT2 gene

Previous studies have shown that glycosyltransferases are crucial for spermatozoa maturation in epididymis and sperm survival in the female reproductive tract [34, 35]. Sperm glycans continue to be modified by glycosyltransferases and carry out functions in female reproductive tracts like acrosome reaction, protection from innate and adaptive female immunity and passage through the cervical mucus [34]. Additionally, glycosyltransferase activities are potentially involved in modification of the glycan on the zona pellucida that boost the its ability to bind spermatozoa [36].

**Association analyses between VMRs and reproduction traits**

To further explore the association of sperm methylation with complex traits, we tested the association of VMRs with 41 complex traits individually using 19 samples from bulls with similar ages (1 to 2-year-old). Unlike HVMRs, VMRs were defined using a less strict criteria (the *p* value of chi-square test less than  $1 \times 10^{-4}$ ) to avoid missing important signals (See

Methods). This resulted in 17,323 VMRs for subsequent analysis, accounting for 55.4% of all identified MHBs. Methylation levels of VMRs were transferred from  $\beta$  values to M values for association tests [37]. We found that VMRs were more likely to be associated with reproduction traits than other traits (Fig. 5a). After Bonferroni correction, we detected 5 significant ( $P < 2.89 \times 10^{-6}$ , 0.05/17323) VMRs and 41 suggestively significant VMRs for five traits ( $P < 5.77 \times 10^{-5}$ , 1/17323), of which four were reproduction traits, including DPR, CCR, GL, and ST\_PL, and the remaining one was related to stature. (Fig. 5b, Additional file 2: Table S3). The results were consistent with the above mentioned WGCNA results. About half ( $n = 25$ ) of these detected VMRs were within 10 kb of annotated genes, and some of them had known functions in male or female fertility. The most significant VMR associated with DPR was within exon 2 of the *ZFP36L1* gene. We observed a highly negative correlation (Pearson  $r = -0.83$ ) between methylation levels of this VMR

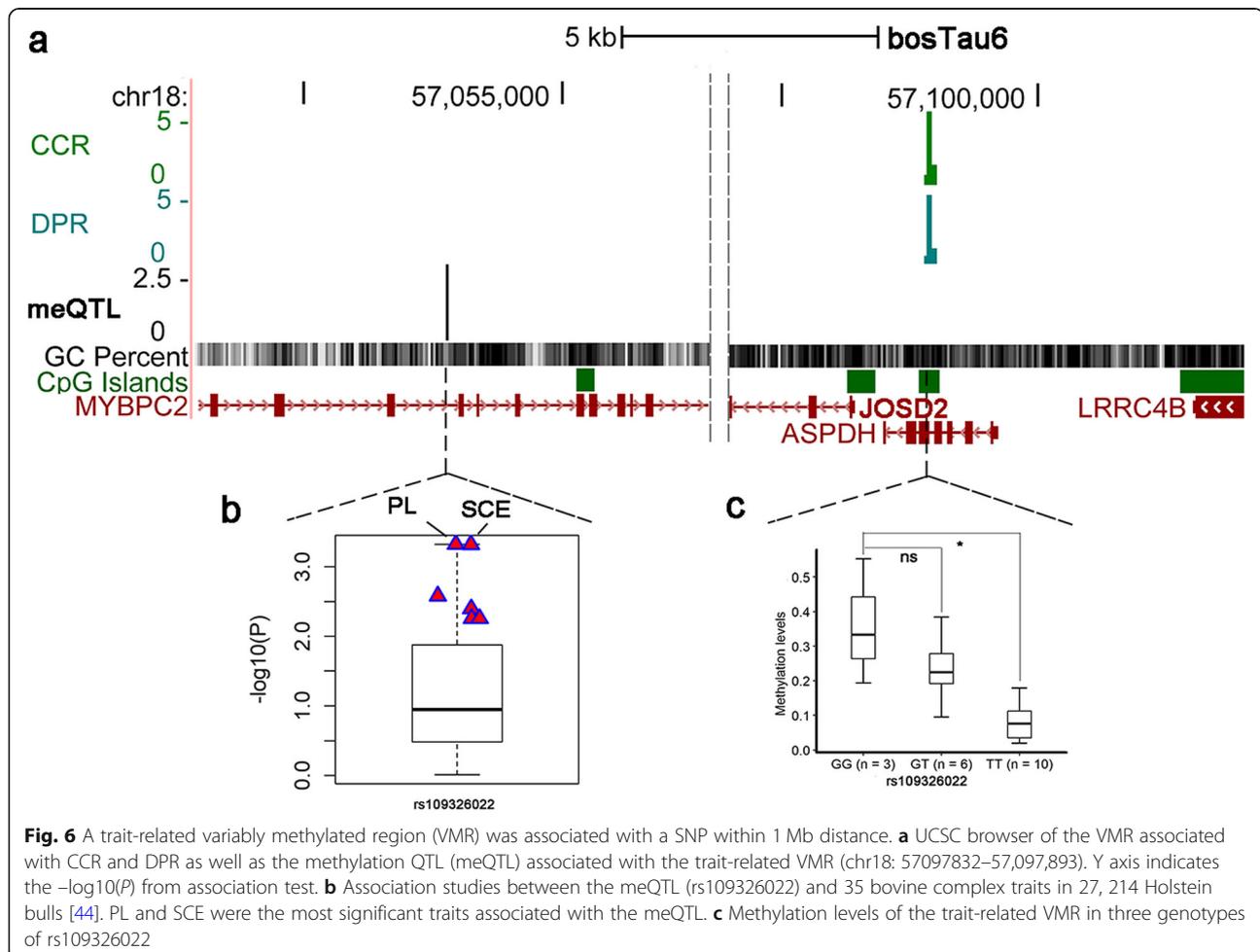


and DPR (Fig. 5c). The *ZNF36L1* gene is crucial for female fertility, and the disruption of this gene will result in embryo lethality [38]. Another example was the suggestively significant VMR associated with GL, which is located ~5.9 kb downstream of the *CRISP2* gene. We observed a positive correlation (Pearson  $r = 0.8$ ) between methylation levels of this VMR and GL (Fig. 5d). Although the mechanisms are not currently understood, existing literature demonstrated that *CRISP2* is a testicular sperm protein involved in spermatogenesis and it participates in acrosome reaction and gamete fusion [39–41]. We examined the transcriptome of 91 bovine tissues, and observed four genes associated with significant VMRs (*CRISP2*, *HGF*, *EFHB* and *ARFGF3*) showing high expression patterns almost exclusively in sperm and testis (Fig. 5e).

#### Examples of the trait-associated VMRs which were influenced by genetic variations

To dissect the impact of genetic variations on trait-associated VMRs, we obtained 80 K SNPs genotypes for the 19 individuals of similar age being studied. Using an R

package MatrixEQTL [42], we tested SNPs within 1 Mb (59 SNPs on average tested for each VMR) surrounding the detected significant or suggestively significant VMRs. Out of the 46 VMRs, nine were associated with at least one cis-SNP (FDR < 0.2). By overlapping these significant methylation QTLs (meQTLs) with the Cattle QTL database, we observed that meQTLs of three trait-associated VMRs were located within QTLs of reproduction traits (conception rate, still birth, calving ease, DPR, PL etc.). An interesting example is a trait-associated VMR (chr18: 57097832–57,097,893) located in the exon 5 of *ASPDH* and upstream (1.5 Kb) of *JOSD2* (Fig. 6a). rs109326022 is the most significant SNP associated with the VMR methylation level. Among the three genotypes, individuals with GG have the highest DNA methylation level but lowest PTA values for DPR and CCR, and those with TT have the lowest DNA methylation level but highest PTA values (Fig. 6c). This cis-meQTL was located within QTLs of PL, calving ease, still birth and calf size from Cattle QTL database. It was also detected as an eeQTL for *JOSD2* in blood, liver and milk cells [30]. On the other hand, a previous DNA methylation study identified *ASPDH* as a gene



enriched in low-fertility sires [43]. To further validate the SNP effects on a range of complex traits, we examined the associations of rs109326022 with 35 complex traits in 27, 214 Holstein bulls [44]. This cis-meQTL was most significantly associated with PL and SCE among all tested traits (Fig. 6b), which might indicate the co-regulatory, intrinsic relationships among the cis-meQTL, sperm DNA methylation, and fertility traits. However, this cis-meQTL did not reach genome-wide significance for PL and SCE in the single-marker GWAS due to the very strict threshold and its small effect.

## Discussion

To our knowledge, this is the first report to characterize the inter-individual variation of sperm DNA methylation and to explore their associations with complex traits in cattle. We demonstrated that HVMRs were distinct from CMRs in terms of methylation levels, expression patterns and their association with complex traits. HVMRs may associate with tissue-specific gene expression and play roles in transcriptional regulation. Our results showed that the sperm epigenetic variations were associated with reproduction traits in cattle.

In our association studies, about half of the significant/suggestively significant VMRs were within 10 kb of annotated genes (Fig. 5b). Some of these genes were functionally important in reproduction. Besides *ZFP36L1* and *CRISP2* mentioned before, we detected the *HGF* gene, whose VMR was significantly associated with DPR. The VMR was located about 4400 bp downstream of *HGF*. Within 91 collected tissues, we found that *HGF* was specifically highly expressed in sperm. A previous study has shown that *HGF* is expressed by Sertoli cells and active during all phases of prenatal and postnatal testis development [45]. *HGF* and its only acceptor c-Met are involved in testis and ovary differentiation. *HGF* also has a role in mediating the spermatogenesis and sperm quality in different aspects [46, 47]. The level of *HGF* was reported to be changed in a comparison between healthy and infertility individuals [48]. In addition, we found one of our significant VMRs (chr18: 57097832–57,097,893) was also located within a differentially methylated region (DMR) (chr18: 57097376–57,098,221) in previous analysis by comparing bulls with high and low male fertility [43]. This significant VMR was identified to associate with CCR and DPR simultaneously. We also identified another nearby VMR (chr18: 56560453–56,560,476) associated with DPR in BTA18 (Additional file 1: Figure S4). These VMRs were located within a very prominent QTL for various fertility traits on chromosome 18 (BTA18, roughly located between 50 Mb and 60 Mb) [49]. This was also supported by our previous study, which reported sperm DNA methylation alterations in this region were associated GL, sire

conception rate (SCR), body depth (BDE), and CCR [4]. Therefore, the influence of the prominent QTL on fertility traits could be a combination of both genetic and epigenetic mechanisms.

Our results indicated the sperm DNA methylation variation associates with reproduction traits in bulls, which were estimated based on the reproduction performance of their daughters, such as DPR, CCR, GL and PL. There are two potential mechanisms could lead to these associations. One is the transgenerational inheritance of sperm epigenome. While DNA methylation erasure occurred after fertilization, induced transgenerational epimutations appear to be protected from it. For example, researchers have shown that male mice fed with unbalanced diets could lead to the metabolic disease in the offspring, coupled with sperm epigenetics alteration [10, 50, 51]. Sperm DNA methylation and sperm RNA could both persist into embryo and even adult tissues, impacting the phenotypes of offspring. Another potential mechanism is that the genetic factors, such as the genetic variants like SNPs associated with DNA methylation (meQTLs), are transmitted to offspring via sperm DNA. We have shown that the genetic effects on DNA methylation could not be dismissed, as about 20% of trait-associated VMRs were influenced by genetic variations. A monozygous and heterozygous twin study reported that contribution of additive genetic factors on methylation variable sites were 23.0% on average [52]. Another study demonstrated that 44% of the methylation variation was best explained by genetic factors [20]. Additionally, meQTLs were previously suggested to be conserved among tissue types and developmental stages [53, 54]. Therefore, although our study performed an initial exploration between sperm DNA methylation and sire fertility traits, it will be interesting for future work to study the mechanisms how sire sperm epigenetic signatures may impact reproduction traits in female progenies. For example, comparing epigenetic information between sire sperm and embryo could provide insights on retained DNA methylation marks.

In our study, we identified VMRs based on MHBs, which were robust and sensitive [23]. Focusing on MHBs helped us to narrow the scope and improve the statistical power for subsequent methylation association analyses. Especially, we found that MHBs were enriched in functionally important elements, like CpG islands, gene promoters, and imprinted genes, which were in concordance with previous characterization of MHBs in human and mice [23, 55]. We revealed extreme enrichment of MHBs in flanking bivalent TSS/enhancer (BivFlnk) and active TSS (TssA) using annotated chromatin states in rumen epithelial cells. BivFlnk colocalizes both active

(H3K4me3) and repressive (H3K27me3) histone modifications, and associates with genes of developmental importance [56]. TssA are active promoters and are associated with developmentally important genes [56]. Functional regions like active promoters and bivalent TSS were stable among cell types or tissues [57, 58]. Because of their conservation among tissues, it is possible to roughly annotate the MHBs in sperm using chromatin states in rumen epithelial cells.

Reproduction traits are complex and influenced by both genetic and epigenetic factors. Distinct from human and mouse, because of artificial insemination, we can measure the paternal contribution of sires to their offspring highly reliably. One advantage of our study was the high reliability of phenotype. The mean reliability for PTA estimation of GL, DPR, CCR, ST\_PL and Stature was higher than 88%. Also we used sperm as our target to analyze the correlation between DNA methylation and phenotypes. Sperm is the only vector to deliver paternal genetic and epigenetic information to offspring, contributing to the variation of phenotypes in offspring (like female reproduction traits and milk traits). In our study, we also selected individuals with lowest relatedness and controlled their ages to around 1~2-year-old to limit confounding factors.

One limitation of our study is the small sample size ( $n = 19$ ) to explore the preliminary associations between VMRs and complex traits. Even though WGBS costs less than before and has become more prevalent, it is still unaffordable to have a large number of samples tested. Therefore, in the near future it is urgent to design either PCR- or array-based high-throughput DNA methylation assays, for example, a low-density bovine methylation array covering important functional regions similar to the Infinium human methylation arrays. This would allow for the confirmation of our findings in a larger sample size.

## Conclusions

The significant enrichment of QTLs, eQTLs and selection signatures in HVMRs indicated the potential roles of methylation variation in sperm on transcriptional regulation, as well as complex trait and adaptive evolution. Our preliminary co-methylation analysis and methylation association study also suggested the potential impacts of sperm methylation variation on reproduction traits, despite their elusive natures. Additionally, we found that the effect of a prominent QTL region on BTA18 on female fertility traits could be related to both genetic and epigenetic mechanisms. In summary, our study of sperm DNA methylation variation provides novel insights into the biological basis underlying complex traits in cattle,

and supplies valuable epigenetic hypotheses for future explorations.

## Methods

### Data processing and methylation haplotype blocks detection

A total of 28 sperm samples were used, and their age and coverage are listed in Additional file 2: Table S4. Twenty-six sperm samples were collected from 23 (1 to 5-year-old) fertile, health and representative U.S. Holstein AI bulls, and the other two sperm samples were collected from two 7-year-old Chinese Holstein AI bulls. All of semen samples passed QC tests (including microscopic examination of sperm count, motility, and abnormality, other laboratory tests) to qualify for commercial distributions. Semen samples were collected from bulls using a standardized procedure with artificial vaginas. Genomic DNA was isolated according to the QIAamp DNA Mini-Kit protocol (QIAGEN, Valencia, CA, USA). DNA quality was assessed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and spectrophotometer (NanoDrop Technologies, Rockland, DE) for DNA concentration, degradation, purify (OD260/280) and potential contamination. The qualified genomic DNA from sperm samples were used to construct libraries as described in Zhou et al. [57]. Programs FastQC v 0.11.2 and Trim Galore v 0.4.0 were used to obtain the read quality and filter the sequences, respectively. Clean reads were subsequently mapped to the reference genome (UMD3.1) using bowtie2 under the Bismark software (0.14.5) with default parameters. The methylcytosine information was extracted using bismark\_methylation\_extractor after deduplicating the duplicated reads.

We identified the MHBs using MONOD2 as described [23]. Briefly, we split the bovine UMD3.1 genome into non-overlapping “sequenceable and mappable” segments (mean size: 2.9 Mb and total size: 2.63Gb) using the combined WGBS data from 28 sperm samples. The mapped reads were converted into methylation haplotypes within each mappable segment. We then calculated MHBs based on the correlation patterns of methylation levels of neighboring CpG sites. Candidate MHBs were defined as the genome regions in which the  $r^2$  value of two adjacent CpG sites was no less than 0.5. MHBs with more than 2 CpG sites were kept for subsequent analysis.

### Enrichment analysis of methylation haplotype blocks for functional genomic regions

Genomic elements such as exons, introns, 5'UTRs, 3'UTRs and gene bodies were downloaded from Ensembl. Repeats and CpG islands were collected using the UCSC Table browser. Promoters were defined as the regions from upstream - 2 kb of TSS to TSS. Fifteen chromatin states were

estimated using 4 histone marks (H3K4me3, H3K4me1, H3K27ac, H3K27me3), ATAC-seq and CTCF-seq from rumen epithelial primary cells in our previous study [25]. Fifteen chromatin states are active TSS, flanking active TSS, strong transcription, weak transcription, enhancers, bivalent enhancer etc., as shown in Roadmap project ([https://egg2.wustl.edu/roadmap/web\\_portal/chr\\_state\\_learning.html#core\\_15state](https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state)). Enrichment analysis was performed using R package regioneR [24] (Permutation test: 1000) and the mappable segments were used as the background. Expression QTL data including sQTL, geQTL and eeQTL in blood and milk cells, liver and muscle were retrieved from the previous study [28]. The summary data of selection signatures differentiating dairy and beef cattle were obtained from [30] using the 1000 Bull Genomes Project data (Run6) [59]. All the genome coordinates were based on UMD3.1/btau6.

#### Identification of highly variable methylated regions and conserved methylated regions

Average methylation level of MHBs were calculated using a weighted methylation level method as described [60]. Methylation levels of regions which didn't meet  $5 \times$  coverage for CpG sites were assigned to "NA". We filtered out the MHBs where more than 13 individuals had methylation levels of "NA". After filtering, 29,542 MHBs were kept for subsequent analysis. We then identified the HVMRs by overlapping the results of standard deviation (SD)-based method and chi-square test method. For the SD-based method, we firstly calculated the median SD for the MHBs. We then compared the SD of the methylation levels of each MHB to the median SD using the chi-square test for variance. We used a significant threshold of 0.01. To control for the family-wise error rate, MHBs with a  $P \leq 3.39 \times 10^{-7}$  (Bonferroni corrected) and SD larger than median level were identified as candidate HVMRs. For the chi-square test method, we treated each CpG result (methylated or unmethylated) in an MHB as an individual observation as previously described [16]. Candidate HVMRs were identified using the threshold of  $P \leq 3.39 \times 10^{-7}$ . Regions identified by both above methods were termed HVMRs. Hypo-CMRs were regions identified using SD method, of which  $P \leq 3.39 \times 10^{-7}$ , SD was smaller than median level and average methylation levels among individuals  $\leq 0.2$ . Hyper-CMRs were regions identified using SD method, of which  $P \leq 3.39 \times 10^{-7}$ , SD was smaller than median level and average methylation levels among individuals  $\geq 0.8$ .

#### Methylation association studies

We collected the individuals ( $n = 19$ ) with similar ages (1~2-year-old) to process the methylation association studies (Additional file 2: Table S4). We

found the methylation levels of MHBs tend to be dynamic among individuals. To increase the power of the association studies, we conducted two steps of quality control. We first filtered out the MHBs ( $n = 9,331$ ) relatively conserved among individuals ( $P > 3.39 \times 10^{-9}$ , Bonferroni corrected) using the chi-square test method. Then we excluded the MHBs ( $n = 2,888$ ) of which methylation levels in 19 individuals had NA values. After filtering, we called these inter-individual variable MHBs as VMRs. Finally, we obtained 17,323 VMRs for association studies. Since average methylation levels can be skewed, to avoid undue influence from outliers, we transferred them to the M-value using the  $\beta$ -value to M-value method [37]. We also collected the phenotypes of 41 bovine complex traits (PTA values of daughter's traits and EBV values of sire conception rate) for the individuals. The statistics summary for these traits are shown in Additional file 2: Table S5. Detailed trait description and trait measurements can be found at <https://www.uscdecb.com/>. To adjust the relationship among the individuals, we conducted the principle component analysis (PCA) using genotype data of 80 K SNP array. We included the first two components (PC1 and PC2) into our association analyses. We used linear regression models:  $y = PC1 + PC2 + Meth$ ; where  $y$  is the PTA values/estimated breeding value of 41 traits, Meth is the M-value of VMRs. After Bonferroni correction, we reported significant VMRs ( $P < 2.89 \times 10^{-6}$ ; 0.05/17323) and suggestive significant VMRs ( $P < 5.77 \times 10^{-5}$ ; 1/17323). All analyses were performed using R (3.5.3).

#### Gene expression across 91 bovine tissues

We obtained 723 transcriptomes of 91 bovine tissues, of which 567 were collected from NCBI SRA/GEO databases and 156 samples were generated locally. Accessions for all datasets were SRP042639, GSE41637, SRP102212, SRP122763, SRP067373, SRP111067, GSE108840, GSE74076, ERP109534, GSE63509, SRP136662, GSE131849, GSE128075 and GSE129416.

#### Identification of cis-meQTL

We tested the SNPs within 1 Mb distance from VMRs in association with methylation. A total of 79,294 probes from 19 individuals were processed. After filtering out the probes with only two genotypes and one of them contained less than 3 individuals, we kept 68,921 probes for subsequent analysis. Methylation levels of VMRs were transferred from  $\beta$  values to M values. We performed association analyses between cis-SNPs and M values of VMRs using MatrixEQTL package [42]. SNPs with  $FDR < 0.2$  were significant cis-meQTL for target VMRs.

### Other downstream bioinformatics analysis

We conducted gene functional annotation enrichments using online software, DAVID v6.8 [61]. We used HOMER [62] for motif discovery and prediction of TF binding sites considering the MHB as background. We conducted QTL enrichment analyses, with a hypergeometric test, for hypo- and hyper-CMRs and HVMRs by using cattle QTLdb (Release 37, Dec. 23, 2018). We arbitrarily considered genes overlapping or closest to the lead SNP in each QTL as the candidate genes for a trait. We only chose complex traits with more than five candidate genes to perform the enrichment analysis. *P*-value were adjusted using the FDR method.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-019-6228-6>.

**Additional file 1: Figure S1.** Correlation among individuals. (a) Pearson correlation among individuals of known genome features. (b) Heatmap of Pearson correlation in methylation haplotype blocks (MHBs) among individuals. **Figure S2.** Comparison of MHBs detected in cattle and human. (a) Functional enrichment of genes associated with cattle specific MHB (CMHB) and human & cattle shared MHB (HCMHB). BP: Biological Process; CC: Cellular Component. (b) Gene density (gene number per kb) in CMHBs and HCMHBs. (c) Distribution of Phastcon scores of CMHBs and HCMHBs. **Figure S3.** Characterization of highly variable methylated regions (HVMRs), hypomethylated conserved regions (Hypo CMRs) and hypermethylated conserved regions (Hyper CMRs). (a) Functional enrichment of genes associated with Hypo CMRs. KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: Biological Process; CC: Cellular Component. (b) Motif enrichments of HVMRs, Hypo CMRs and Hyper CMRs. (c) Enrichment of human orthologous genes associated with HVMRs, Hypo CMRs and Hyper CMRs in tissue-specific genes. **Figure S4.** A trait-related VMR was associated with an SNP within 1 Mb distance. (a) The UCSC browser of the epigenetic markers associated with CCR and DPR as well as the methylation QTLs (meQTLs) associated with the trait-related VMR (chr18: 56560453–56,560,476). (b) Association studies between the meQTL (rs41893756) and 35 bovine complex traits in 27, 214 Holstein bulls. PL and SCE were the most significant traits associated with the meQTL. (c) Methylation levels of the trait-related VMR in two genotypes of rs41893756.

**Additional file 2: Table S1.** Identified methylation haplotype blocks (MHBs) in sperm DNA methylation. **Table S2.** Identified highly variable methylated regions (HVMRs), hypomethylated conserved regions (Hypo CMRs) and hypermethylated conserved regions (Hyper CMRs). **Table S3.** Significant VMRs associated with bovine complex traits and the most significant meQTL within 1 Mb of these VMRs. Some meQTL were within the reproduction-related QTLs (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). **Table S4.** Sperm sample description. **Table S5.** Statistics summary of predicted transmitting ability (PTA) values and the accuracy for Bovine complex traits in association studies with VMRs.

### Abbreviations

BDE: Body depth; BTA: *bos taurus* autosome; CCR: Cow conception rate; CE: Calving ease; CMHB: Cattle-specific MHB; CMR: Conserved methylated region; DPR: Daughter pregnancy rate; eeQTL: Exon expression QTL; FDR: False discovery rate; geQTL: Gene expression QTL; GO: Gene ontology; GWAS: Genome wide association studies; HCMHB: Common MHB between human and cattle; HCR: Heifer conception rate; HVMR: Highly variably methylated region; hyper-CMR: Conserved hypermethylated region; hypo-CMR: Conserved hypomethylated region; LD: Linkage disequilibrium; meQTL: Methylation QTL; MHB: Methylation haplotype block; PCA: Principle component analysis; PL: Productive life; PTA: Predicted transmitting ability; EWAS: Epigenetic association studies; QTL: Quantitative trait loci; SB: Stillbirth;

SCR: Sire conception rate; sQTL: Splicing QTL; ST\_PL: Standard length of productive life; TFBS: Transcription factor binding sites; UTR: Untranslated region; VMR: Variably methylated region; WGBS: Whole genome bisulfite sequencing; WGCNA: Weighted correlation network analysis

### Acknowledgements

We thank Reuben Anderson, Alexandre Dimtchev for technical assistance. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. 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.

### Authors' contributions

SL, LF, SZ and GEL conceived and designed the experiments. RX, HDD, AJC, JBC and CJL collected samples and/or generated data. SL, LF, YZ, DS, YY and LM performed computational and statistical analyses. SL, LF, LM, SZ and GEL wrote the paper. All authors read and approved the final manuscript.

### Funding

This work was supported in part by AFRI grant numbers 2013–67015-20951, 2016–67015-24886, and 2019–67015-29321 from the USDA National Institute of Food and Agriculture (NIFA) Animal Genome and Reproduction Programs and BARD grant number US-4997-17 from the US-Israel Binational Agricultural Research and Development (BARD) Fund. The funders played no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

### Availability of data and materials

All high-throughput sequencing data were deposited in NCBI GEO database under accession numbers GSE119263, GSE106538, and GSE131851.

### Ethics approval and consent to participate

Semen samples from U.S. Holstein bulls were collected with approval of the US Department of Agriculture Agricultural Research Service Institutional Animal Care and Use Committee under Protocol 16–016. Semen samples from China Holstein bulls were collected with approval of the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University under Protocol DK996, which included written informed consent to use the animals from the owners of the animals.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests except that George Liu is a member of the editorial board (Associate Editor) of this journal.

### Author details

<sup>1</sup>College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. <sup>2</sup>USDA-ARS, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705, USA. <sup>3</sup>Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA. <sup>4</sup>Medical Research Council Human Genetics Unit at the Medical Research Council Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. <sup>5</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Education Ministry of China, Huazhong Agricultural University, Wuhan 430070, Hubei, China. <sup>6</sup>Faculty of Veterinary & Agricultural Science, The University of Melbourne, Parkville, Victoria 3052, Australia. <sup>7</sup>Agriculture Victoria, AgriBio, Centre for AgriBiosciences, Bundoora, Victoria 3083, Australia. <sup>8</sup>School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia.

Received: 19 June 2019 Accepted: 28 October 2019

Published online: 21 November 2019

## References

- Aston KI, Uren PJ, Jenkins TG, Horsager A, Cairns BR, Smith AD, Carrell DT. Aberrant sperm DNA methylation predicts male fertility status and embryo quality. *Fertil Steril*. 2015;104(6):1388–1397.e1381–1385.
- Wyck S, Herrera C, Requena CE, Bittner L, Hajkova P, Bollwein H, Santoro R. Oxidative stress in sperm affects the epigenetic reprogramming in early embryonic development. *Epigenetics Chromatin*. 2018;11(1):60.
- Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev*. 2014;28(8):812–28.
- Fang L, Jiang J, Li B, Zhou Y, Freebern E, Vanraden PM, Cole JB, Liu GE, Ma L. Genetic and epigenetic architecture of paternal origin contribute to gestation length in cattle. *Commun Biol*. 2019;2:100.
- Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BA, Smith AD, Dean MD, Good JM. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics*. 2016;203(4):1841–57.
- Stewart KR, Veselovska L, Kelsey G. Establishment and functions of DNA methylation in the germline. *Epigenomics*. 2016;8(10):1399–413.
- Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, Bourc'his D. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev*. 2015;29(12):1256–70.
- Miller D, Brinkworth M, Iles D. Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. *Reproduction*. 2010;139(2):287–301.
- Ben Maamar M, Nilsson E, Sadler-Riggelman I, Beck D, McCarrey JR, Skinner MK. Developmental origins of transgenerational sperm DNA methylation epimutations following ancestral DDT exposure. *Dev Biol*. 2019;445(2):280–93.
- Wei Y, Yang CR, Wei YP, Zhao ZA, Hou Y, Schatten H, Sun QY. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proc Natl Acad Sci U S A*. 2014;111(5):1873–8.
- Milekic MH, Xin Y, O'Donnell A, Kumar KK, Bradley-Moore M, Malaspina D, Moore H, Brunner D, Ge Y, Edwards J, 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(8):995–1001.
- Chu AY, Tin A, Schlosser P, Ko YA, Qiu C, Yoehanes R, Grams ME, Liang L, Gluck CA, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun*. 2017;8(1):1286.
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12(8):529–41.
- Binney E, Smith GD, Grealley JM. Epigenome-wide association studies and the interpretation of disease -omics. *PLoS Genet*. 2016;12(6):e1006105.
- Garg P, Joshi RS, Watson C, Sharp AJ. A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. *PLoS Genet*. 2018;14(10):e1007707.
- Chatterjee A, Stockwell PA, Rodger EJ, Duncan EJ, Parry MF, Weeks RJ, Morison IM. Genome-wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic variation. *Sci Rep*. 2015;5:17328.
- Horsburgh S, Robson-Ansley P, Adams R, Smith C. Exercise and inflammation-related epigenetic modifications: focus on DNA methylation. *Exerc Immunol Rev*. 2015;21:26–41.
- Ronn T, Volkov P, Davegardh C, Dayeh T, Hall E, Olsson AH, Nilsson E, Tornberg A, Dekker Nitert M, Eriksson KF, et al. A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet*. 2013;9(6):e1003572.
- El-Dahr SS. DNA methylation links intrauterine stress with abnormal nephrogenesis. *Nat Rev Nephrol*. 2019;15(4):196–7.
- Czamura D, Eraslan G, Page CM, Lahti J, Lahti-Pulkkinen M, Hamalainen E, Kajantie E, Laivuori H, Villa PM, Reynolds RM, et al. Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. *Nat Commun*. 2019;10(1):2548.
- Hachiya T, Furukawa R, Shiwa Y, Ohmomo H, Ono K, Katsuoka F, Nagasaki M, Yasuda J, Fuse N, Kinoshita K, et al. Genome-wide identification of inter-individually variable DNA methylation sites improves the efficacy of epigenetic association studies. *NPJ Genomic Med*. 2017;2:11.
- Davis CN, Slutske WS, Martin NG, Agrawal A, Lynskey MT. Genetic and environmental influences on gambling disorder liability: a replication and combined analysis of two twin studies. *Psychol Med*. 2019;49(10):1705–12.
- Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet*. 2017;49(4):635–42.
- Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the associational analysis of genomic regions based on permutation tests. *Bioinformatics*. 2016;32(2):289–91.
- Fang L, Liu S, Liu M, Kang X, Lin S, Li B, Connor EE, Baldwin RL, Tenesa A, Ma L, et al. Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. *BMC Biol*. 2019;17(1):68.
- Jiang Z, Lin J, Dong H, Zheng X, Marjani SL, Duan J, Ouyang Z, Chen J, Tian XC. DNA methylomes of bovine gametes and in vivo produced preimplantation embryos. *Biol Reprod*. 2018;99(5):949–59.
- Kawai Y, Oda A, Kanai Y, Goitsuka R. Germ cell-intrinsic requirement for the homeodomain transcription factor Pknox1/Prep1 in adult spermatogenesis. *PLoS One*. 2018;13(1):e0190702.
- Xiang R, Hayes BJ, Vander Jagt CJ, MacLeod IM, Khansefid M, Bowman PJ, Yuan Z, Prowse-Wilkins CP, Reich CM, Mason BA, et al. Genome variants associated with RNA splicing variations in bovine are extensively shared between tissues. *BMC Genomics*. 2018;19(1):521.
- van den Berg I, Hayes BJ, Chamberlain AJ, Goddard ME. Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. *BMC Genomics*. 2019;20(1):291.
- Xiang R, Berg Ivd, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, Bolormaa S, Liu Z, Rochford SJ, Reich CM, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A*. 2019;116(39):19398–408.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
- Stanley P, Batista F, Huang HH. Roles for Glycans in mammalian development and spermatogenesis. *FASEB J*. 2013;27(1\_supplement):211.2.
- Huang HH, Stanley P. A testis-specific regulator of complex and hybrid N-glycan synthesis. *J Cell Biol*. 2010;190(5):893–910.
- Teclé E, Gagneux P. Sugar-coated sperm: unraveling the functions of the mammalian sperm glycocalyx. *Mol Reprod Dev*. 2015;82(9):635–50.
- Takasaki N, Tachibana K, Ogasawara S, Matsuzaki H, Hagiuda J, Ishikawa H, Mochida K, Inoue K, Ogonuki N, Ogura A, et al. A heterozygous mutation of GALNTL5 affects male infertility with impairment of sperm motility. *Proc Natl Acad Sci U S A*. 2014;111(3):1120–5.
- Tulsiani DR, Chayko CA, Orgebin-Crist MC, Araki Y. Temporal surge of glycosyltransferase activities in the genital tract of the hamster during the estrous cycle. *Biol Reprod*. 1996;54(5):1032–7.
- Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
- Stumpo DJ, Byrd NA, Phillips RS, Ghosh S, Maronpot RR, Castranio T, Meyers EN, Mishina Y, Blackshear PJ. Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36L1, a gene encoding a CCCH tandem zinc finger protein of the Tristetraprolin family. *Mol Cell Biol*. 2004;24(14):6445–55.
- Jamsai D, Rijal S, Bianco DM, O'Connor AE, Merriner DJ, Smith SJ, Gibbs GM, O'Bryan MK. A novel protein, sperm head and tail associated protein (SHTAP), interacts with cysteine-rich secretory protein 2 (CRISP2) during spermatogenesis in the mouse. *Biol Cell*. 2009;102(2):93–106.
- Brukman NG, Miyata H, Torres P, Lombardo D, Caramelo JJ, Ikawa M, Da Ros VG, Cuasnicu PS. Fertilization defects in sperm from cysteine-rich secretory protein 2 (Crisp2) knockout mice: implications for fertility disorders. *Mol Hum Reprod*. 2016;22(4):240–51.
- Lim S, Kierzek M, O'Connor AE, Brenker C, Merriner DJ, Okuda H, Volpert M, Gaikwad A, Bianco D, Potter D, et al. CRISP2 is a regulator of multiple aspects of sperm function and male fertility. *Endocrinology*. 2019;160(4):915–24.
- Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28(10):1353–8.
- Kropp J, Carrillo JA, Namous H, Daniels A, Salih SM, Song J, Khatib H. Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. *BMC Genomics*. 2017;18(1):280.

44. Jiang J, Cole JB, Freebern E, Da Y, VanRaden PM, Ma L. Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun Biol.* 2019;2:212.
45. Ricci G, Catizone A. Pleiotropic activities of HGF/c-met system in testicular physiology: paracrine and endocrine implications. *Front Endocrinol.* 2014;5:38.
46. Catizone A, Ricci G, Galdieri M. Functional role of hepatocyte growth factor receptor during sperm maturation. *J Androl.* 2002;23(6):911–8.
47. Wiltshire EJ, Flaherty SP, Couper RT. Hepatocyte growth factor in human semen and its association with semen parameters. *Hum Reprod.* 2000;15(7):1525–8.
48. Depuydt CE, De Potter CR, Zalata A, Baekelandt E, Bosmans E, Comhaire FH. Levels of hepatocyte growth factor/scatter factor (HGF/SF) in seminal plasma of patients with andrological diseases. *J Androl.* 1998;19(2):175–82.
49. Muller MP, Rothhammer S, Seichter D, Russ I, Hinrichs D, Tetens J, Thaller G, Medugorac I. Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *J Dairy Sci.* 2017;100(3):1987–2006.
50. Mejos KK, Kim HW, Lim EM, Chang N. Effects of parental folate deficiency on the folate content, global DNA methylation, and expressions of FRalpha, IGF-2 and IGF-1R in the postnatal rat liver. *Nutr Res Pract.* 2013;7(4):281–6.
51. Schuster A, Skinner MK, Yan W. Ancestral vinclozolin exposure alters the epigenetic transgenerational inheritance of sperm small noncoding RNAs. *Environ Epigenet.* 2016;2(1):dvw001.
52. Hannon E, Knox O, Sugden K, Burrage J, Wong CCY, Belsky DW, Corcoran DL, Arseneault L, Moffitt TE, Caspi A, et al. Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLoS Genet.* 2018;14(8):e1007544.
53. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics.* 2014;15:145.
54. Lin D, Chen J, Perrone-Bizzozero N, Bustillo JR, Du Y, Calhoun VD, Liu J. Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia. *Genome Med.* 2018;10(1):13.
55. Ma Q, Xu Z, Lu H, Xu Z, Zhou Y, Yuan B, Ci W. Distal regulatory elements identified by methylation and hydroxymethylation haplotype blocks from mouse brain. *Epigenetics Chromatin.* 2018;11(1):75.
56. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell.* 2006;125(2):315–26.
57. Zhou Y, Connor EE, Bickhart DM, Li C, Baldwin RL, Schroeder SG, Rosen BD, Yang L, Van Tassell CP, Liu GE. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. *Gigascience.* 2018;7(5). <https://doi.org/10.1093/gigascience/giy039>.
58. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.
59. Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, Liao X, Djari A, Rodriguez SC, Grohs C. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet.* 2014;46(8):858.
60. Schultz MD, Schmitz RJ, Ecker JR. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* 2012;28(12):583–5.
61. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
62. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576–89.
63. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012;9(3):215–6.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

