

ARTICLE

<https://doi.org/10.1038/s42003-019-0341-6>

OPEN

# Genetic and epigenetic architecture of paternal origin contribute to gestation length in cattle

Lingzhao Fang<sup>1,2</sup>, Jicai Jiang<sup>1</sup>, Bingjie Li<sup>2</sup>, Yang Zhou<sup>3</sup>, Ellen Freebern<sup>1</sup>, Paul M. Vanraden<sup>2</sup>, John B. Cole<sup>2</sup>, George E. Liu<sup>2</sup> & Li Ma<sup>1</sup>

The length of gestation can affect offspring health and performance. Both maternal and fetal effects contribute to gestation length; however, paternal contributions to gestation length remain elusive. Using genome-wide association study (GWAS) in 27,214 Holstein bulls with millions of gestation records, here we identify nine paternal genomic loci associated with cattle gestation length. We demonstrate that these GWAS signals are enriched in pathways relevant to embryonic development, and in differentially methylated regions between sperm samples with long and short gestation length. We reveal that gestation length shares genetic and epigenetic architecture in sperm with calving ability, body depth, and conception rate. While several candidate genes are detected in our fine-mapping analysis, we provide evidence indicating *ZNF613* as a promising candidate for cattle gestation length. Collectively, our findings support that the paternal genome and epigenome can impact gestation length potentially through regulation of the embryonic development.

<sup>1</sup>Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA. <sup>2</sup>Animal Genomics and Improvement Laboratory, BARC, Agricultural Research Service, USDA, Beltsville, MD 20705, USA. <sup>3</sup>Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Education Ministry of China, Huazhong Agricultural University, 430070 Wuhan, Hubei, China. These authors contributed equally: Lingzhao Fang, Jicai Jiang. Correspondence and requests for materials should be addressed to G.E.L. (email: [George.Liu@ars.usda.gov](mailto:George.Liu@ars.usda.gov)) or to L.M. (email: [lima@umd.edu](mailto:lima@umd.edu))

**G**estation length measures the fetal development period from conception to subsequent parturition in mammals, which is crucial for mammalian development. The events occurred in gestation period can have important consequences for the health, productivity, and fertility of the offspring<sup>1</sup>. The abnormality of gestation length can lead to either preterm or post-term birth, resulting in acute and long-term adverse health outcomes in humans<sup>2–4</sup>. In cattle, the length of gestation highly correlates with health, production, and reproduction performances<sup>5</sup>. For instance, prolonged gestation length has been reported to be associated with increased fetal weight, reduced pregnancy rate, and more difficult calving in dairy cattle<sup>5,6</sup>. Furthermore, gestation is a unique immunological state (i.e., the immune clock of pregnancy) that may help predict preterm birth (i.e., shortened gestation length)<sup>7</sup> that can influence the risk of developing immune-related diseases<sup>8</sup>. Moreover, gestation length as a trait also has direct applications to the dairy industry, because more precise expected dates can be used to assist herd management practices on health and nutritional aspects<sup>5</sup>.

Gestation length is a complex phenotype affected by many genetic and environmental factors, including the progesterone rise, prenatal growth, maternal age, and maternal and fetal immune systems<sup>9–11</sup>. It has been reported that gestation length is a highly heritable trait with a heritability of 30–50% in humans and cattle<sup>10,12,13</sup>. By measuring properties of pregnancy, gestation length also has a complex genetic architecture with direct contributions from maternal and fetal genomes, and likely an indirect, paternal influence through regulation of the fetal development. Many studies in humans have explored the maternal genetic factors that were associated with gestation length<sup>4,14,15</sup>, but few of them have investigated the indirect impacts of paternal genetics on gestation length, possibly due to the limited availability of data.

The U.S. dairy industry has a long history of collecting phenotypic records of dairy cattle. Based on millions of mating and dairy production records, the U.S. dairy cattle database has archived a large amount of reliable phenotypes on gestation length for thousands of service bulls. This data resource provides valuable information to study the paternal impacts on gestation length in mammals. A previous study reported that the heritability of gestation length for service sires (the paternal contribution) was about 47 and 33% when mated to heifers and cows, respectively<sup>10</sup>. On the other hand, the epigenetic information in sperm has been shown to influence the embryonic development through regulating gene expression in embryos<sup>16–18</sup>, warranting interest to explore whether the epigenetic alterations in sperm can also contribute to gestation length.

In this study, we seek to investigate the paternal genetic contribution to gestation length in cattle through associating ~3 million imputed sequence variants with gestation length in a large sample of 27,214 Holstein bulls. In addition, we characterized genome-wide DNA methylation alterations in sperm that were associated with gestation length and three genetically correlated traits of economic importance, sire calving ease (SCE), body depth (BDE), and cow conception rate (CCR), by comparing sperm methylomes of 18 representative animals with extreme phenotypes<sup>19</sup>. Moreover, we integrated the genetic variants of gestation length with DNA methylation alterations in sperm. We further validated our findings by examining publicly available transcriptome data across 87 adult and embryonic tissues<sup>20</sup>. Finally, we provided genetic, epigenetic, and selection evidence implicating a candidate gene *ZNF613* for gestation length. Collectively, our results illustrated the importance of paternal genome and epigenome to gestation length through regulation of embryonic development, and provided insights into the genetic and biological mechanisms underpinning gestation length. We

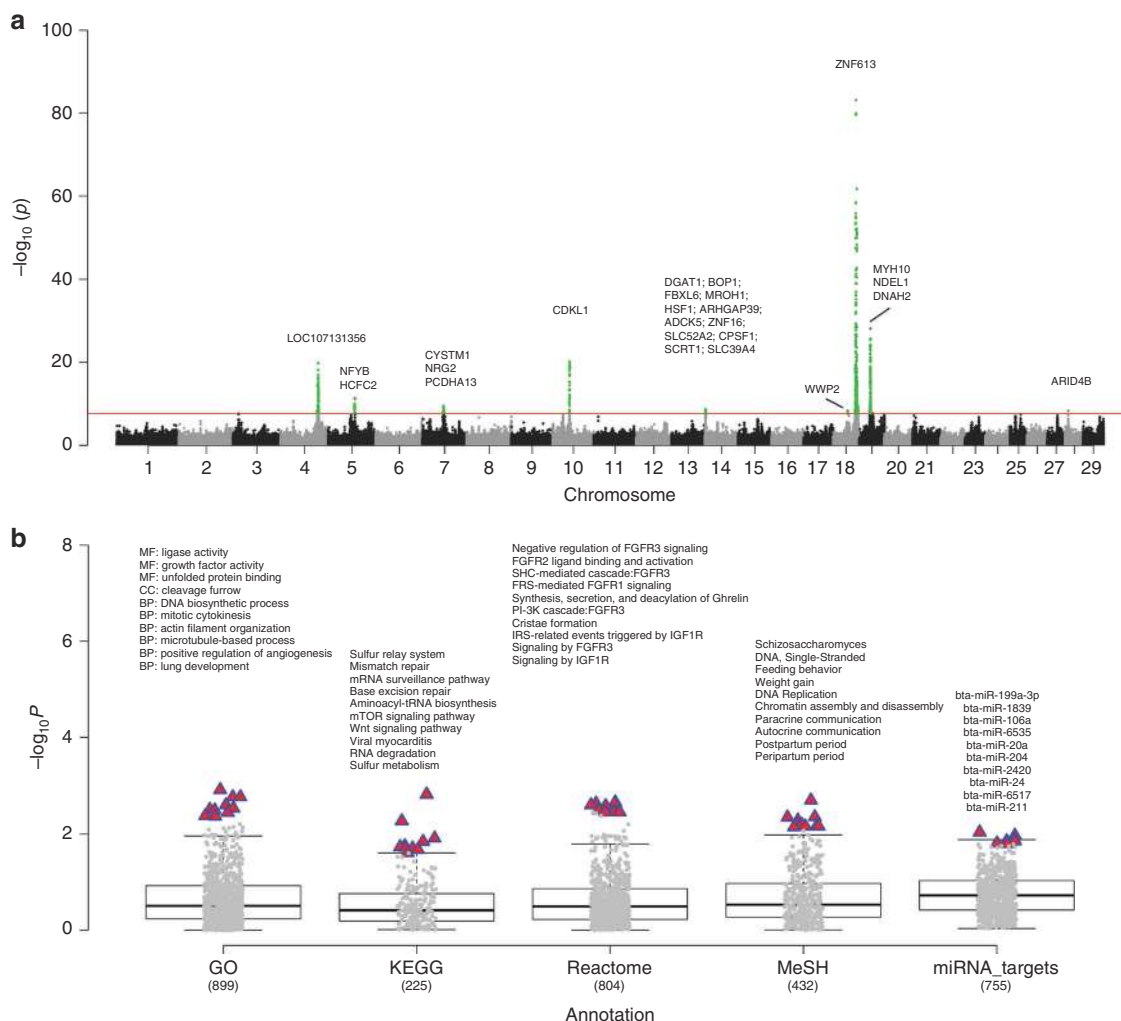
believe that our findings in cattle can provide valuable knowledge for other mammals, including human and rodents.

## Results

### The complex genetic architecture underlying gestation length.

A total of 27,214 Holstein bulls with ~3 million imputed sequence variants and highly reliable phenotypes were included in the current analyses. In total, our single-marker GWAS revealed nine quantitative trait loci (QTL) that were located in the *Bos taurus* chromosome (BTA) 4, 5, 7, 10, 14, 18, 19, and 28, respectively (Fig. 1a). BTA 18 had two QTLs with one of them being the most significant, consistent with a previous study that used a smaller sample size of 4743 bulls<sup>21</sup>. Our following fine-mapping analyses on these nine QTL regions determined 25 candidate genes (posterior probability of causality > 0.05) for gestation length, including multiple genes participating in the embryonic development (e.g., *HSF1*<sup>22</sup>, *MYH10*<sup>23,24</sup>, *NDEL1*<sup>25</sup>, and *NRG2*<sup>26</sup>), immune responses (e.g., *HCFC2*, and *CYSTM1*), DNA processing (e.g., *WWP2*, *CDKL1*, *ZNF613*, and *CPSF1*), and cell differentiation (e.g., *ZNF16* and *ARID4B*) (Fig. 1a; Supplementary Data 1). By examining the Bovine Gene Atlas data that measured the transcriptome across 87 tissues in cattle (<http://www.innatedb.com/>)<sup>20</sup>, we found that 5 out of 25 fine-mapped genes exhibited the highest expression level in placenta, including *SLC39A4*, *WWP2*, *DNAH2*, *ZNF613*, and *MYH10*. We further found five fine-mapped genes having the highest expression in the immune and growth-related glands (e.g., thymus, anterior pituitary and thyroid), including *NFBY*, *NRG2*, *FBXL6*, *ZNF613*, and *ARID4B*, among which *NFBY* and *ZNF613* were also highly expressed in the embryonic tissues (i.e., fetal tongue surface) (Supplementary Data 1). Through examining the human GWAS Catalog (<https://www.ebi.ac.uk/gwas/home>), we found that the human orthologues of three fine-mapped genes, *DNAH2*, *CYSTM1*, and *WWP2*, have been reported to be significantly associated with vascular endothelial growth factor levels, intelligence (neurogenesis and myelination), and menarche in human, respectively, suggesting their important roles in development and fertility (Supplementary Data 1). Together, all these results suggest that our fine-mapped genes likely affect gestation in a tissue-specific manner and potentially through the regulation of the fetal genome and development.

Based on the omnigenetic model of complex phenotypes<sup>27</sup>, we conducted GWAS signal enrichment analyses to determine the core molecular interaction networks that are engaged in regulating gestation length. We employed five commonly used gene annotation sources, including Gene ontology (GO), Kyoto Encyclopedia of genes and genomes (KEGG) pathway, Reactome metabolism pathway, Medical subject headings (MeSH), and miRNA-target networks (miRBase). As shown in Fig. 1b (See Supplementary Data 2 for details), various biological processes and pathways may affect gestation length, including multiple embryonic developmental processes (e.g., positive regulation of angiogenesis, Wnt signaling pathway, mTOR signaling pathway, and mRNA surveillance pathway), growth factors signaling pathways (e.g., fibroblast growth factor receptors (FGFRs), ghrelin synthesis and insulin-like growth factor 1 receptor (IGF1R) regulation), and DNA processing pathways (e.g., DNA biosynthetic process, DNA replication and mismatch repair). These results were consistent with previous findings that DNA damage was repaired very effectively during pregnancy in humans<sup>28</sup>. Analyses on the basis of MeSH revealed that gestation length was strongly associated with feeding behaviors, postpartum and peripartum periods, and autocrine and paracrine communications, in agreement with a previous study that reported the important roles of autocrine and paracrine signaling

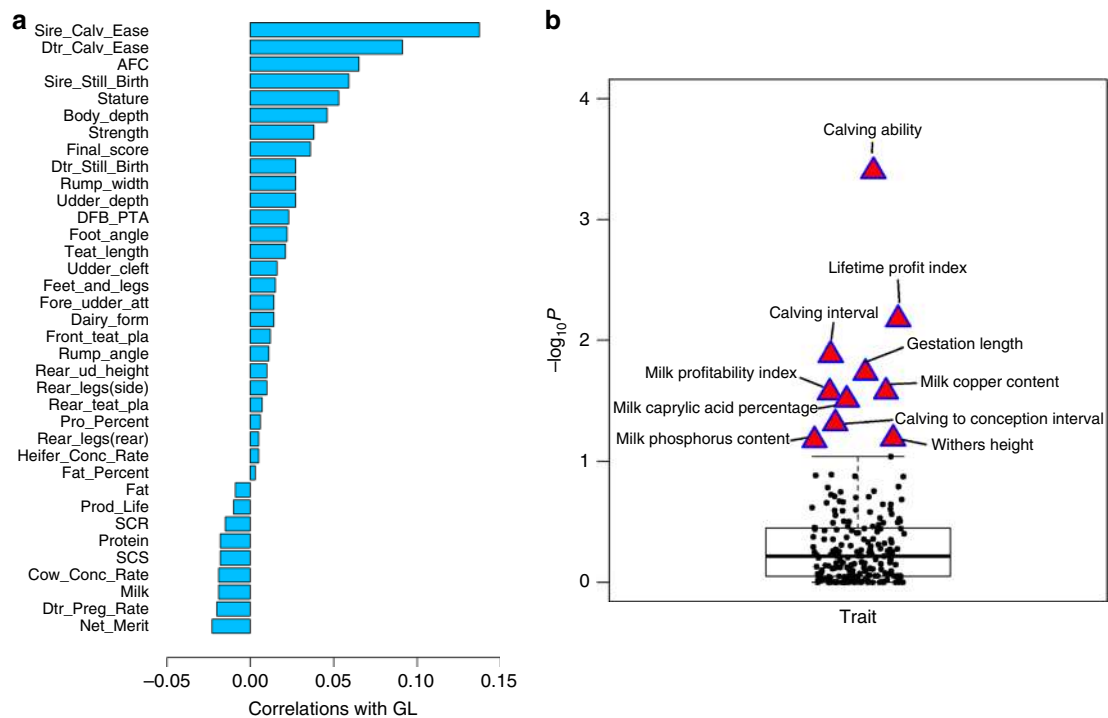


**Fig. 1** Single-marker GWAS, gene-level fine-mapping, and GWAS signal enrichment of gestation length. **a** Manhattan plot of all the imputed sequence variants being tested. Red line denotes the genome-wide significance of  $P$  equals to  $1.91e-08$ . All the candidate genes are determined with the posterior probability of causality  $>0.05$  based on gene-level fine mapping (BMAP). **b** GWAS signal enrichment on the basis of five gene annotation sources, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, Reactome metabolism pathway, Medical Subject Headings (MeSH), and miRNA-target networks. The values at x axis are the number of genomic features (i.e., gene lists) being tested in the corresponding annotation sources. The red triangle denotes the top 10 items with the highest enrichments (i.e.,  $-\log_{10}P$ ) in each annotation source. The names of these items are shown in the figure

in the embryonic skeletal development through regulating the IGF1R signaling pathways<sup>29</sup>. Our miRNA-target network analyses showed that 84 out of the 755 tested miRNAs were significantly ( $P < 0.05$ ) involved in gestation length (Supplementary Data 2), and their targets were enriched ( $FDR < 0.05$ ) in the regulation of action cytoskeleton, miRNA surveillance pathways, multiple growth-related hormone metabolism (e.g., adrenergic signaling and parathyroid hormone synthesis, secretion and action), and immune responses (e.g., leukocyte transendothelial migration and bacterial invasion of epithelial cells) (Supplementary Fig. 1). Of note were the enrichments of all the miRNA-targets, which were significantly ( $P < 0.01$ ;  $t$ -test) higher than all the other four annotation databases (Fig. 1b; Supplementary Data 2). Previous studies showed that maternal plasma miRNAs can be used as a biomarker during pregnancy to predict preterm birth and pregnancy loss in human<sup>30,31</sup>, and the altered expression of circulating miRNAs has also been proposed to be associated with pregnancy in cattle<sup>32</sup>. We further validated that 8 and 4 out of the 84 significant miRNAs were differentially

expressed in milk<sup>33</sup> and plasma<sup>34</sup>, respectively, during early pregnancy in cattle, including bta-miR-20a, bta-miR-106b, bta-miR-100, bta-miR-143, bta-miR-99b, bta-miR-125b, bta-miR-125a, bta-miR-93, bta-miR-99a-5p, bta-miR-99b, bta-miR-125b, and bta-miR-29a. Our findings were consistent with previous findings that miRNAs played an important role in pregnancy, especially for fetal growth and regulation of the immune system<sup>35,36</sup>.

Combining gestation length with other economically important dairy traits, we showed that gestation length was significantly positively correlated with calving ability (i.e., service sire and dam effects on still birth and calving ease, which were combined to measure calving ability) and body conformation traits (e.g., stature and body depth), whereas it was significantly, yet negatively correlated with conception/pregnancy rates and milk production performance (Fig. 2a). Through conducting GWAS signal enrichment analysis using QTLs of 208 complex traits in the Cattle QTLdb<sup>37</sup> (<https://www.animalgenome.org/cgi-bin/QTLdb/index>), we confirmed that gestation length was not only



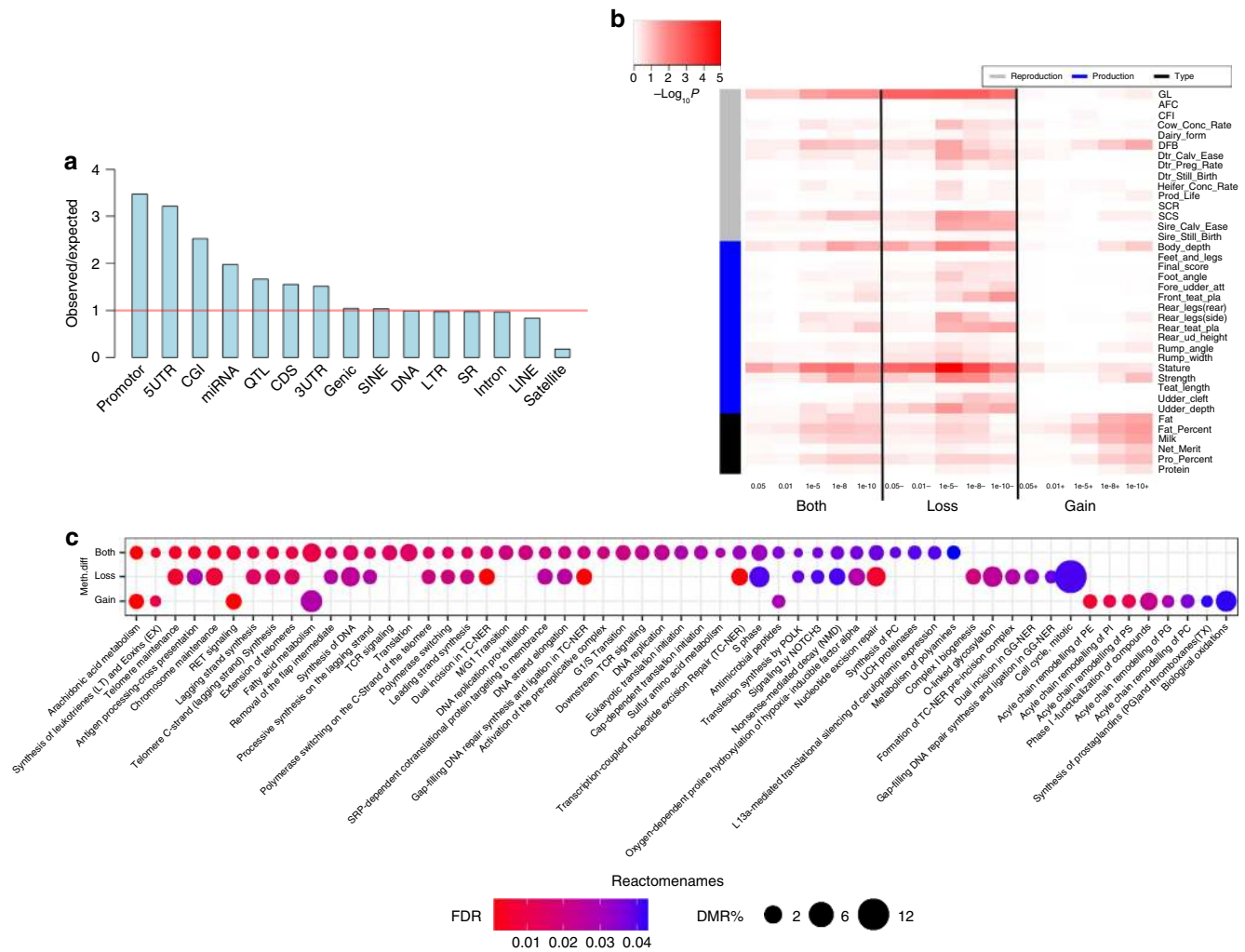
**Fig. 2** Genetic relationships of gestation length and other dairy traits of economic importance. **a** Genetic correlations between gestation length and 35 dairy traits in the U.S. Holstein population. The genetic correlations are approximately computed using the effects of all tested variants in GWAS. **b** GWAS signal enrichment of gestation length based on the Cattle QTLdb (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). Each dot denotes a list of genes that are associated with a complex trait in Cattle QTLdb. The top 10 traits with the highest enrichments (i.e.,  $-\log_{10}P$ ) are shown in the figure

highly associated with calving ability and withers height, but also highly associated with milk caprylic acid percentage, as well as milk phosphorus and copper content (Fig. 2b; Supplementary Data 2). This was consistent with other studies that indicated phosphorus and copper deficiency can affect prenatal development and pregnancy outcomes in human<sup>38,39</sup>. These results suggest that gestation length may share the underlying genetic basis with many dairy traits of economic importance, reflecting the biological and genetic complexity as well as economic implications of gestation length in cattle.

### Sperm methylation changes associated with gestation length.

By comparing sperm methylomes between bulls with high and low gestation length, we aimed to determine the differentially methylated regions (DMR) in sperm (adjusted  $P$  value ( $q$ ) < 0.01 and the absolute difference in methylation >5%), which may contribute to gestation length through regulating the fetal development. In total, we detected 66,318 out of 593,035 tested regions as significant DMRs, and a QQ plot for the DMR analysis was shown in Supplementary Fig. 2. We found that many of these gestation length associated DMRs were clustered in BTA 13, 18, and 25 as compared to other chromosomes (Supplementary Fig. 3). We further observed that gestation length associated DMRs intersected many genomic elements, but were more likely to be enriched in promoters, CpG islands, miRNAs, and QTLs of gestation length (i.e.,  $\pm 1$  Mb around the top associated SNPs) (Fig. 3a). Since there was an inflation of the test statistics in the QQ plot of the DMR analysis, our analysis was not focused on the significant DMRs. Instead, we defined multiple sets of DMRs with different  $q$  value cutoffs, and then investigated the enrichment of GWAS signals in each DMR set using GWAS signal enrichment analysis (see Methods). Our results revealed that GWAS signals were significantly

( $P < 0.05$ ) enriched in gestation length associated DMRs. This enrichment was also significant for other traits, including days to first breeding after calving (DFB), somatic cell score (SCS, which is highly related to mastitis<sup>40</sup>), multiple body type traits (e.g., strength and stature), and milk production traits (e.g., milk and fat yields). Of note, we found that DMRs that lost methylation in animals with higher gestation length had stronger enrichment (i.e., smaller  $P$ -value) than those that gained methylation across gestation length and multiple reproduction and body type traits (Fig. 3b). However, milk production traits exhibited an opposite trend, where DMRs that gained methylation in animals with higher gestation length had more significant enrichment than those that lost methylation (Fig. 3b). Our DMR-set enrichment using Reactome pathways further validated that DMRs that lost methylation in animals with higher gestation length were significantly enriched in pathways related to telomere and chromosome maintenance, translation, and DNA damage repairs, suggesting their potential role in pregnancy and embryonic development<sup>41,42</sup> (Fig. 3c; Supplementary Data 3). DMRs that gained methylation were significantly engaged in lipid metabolism (e.g., arachidonic acid metabolism and fatty acid metabolism), indicating their important roles in the regulation of milk production (Fig. 3c; Supplementary Data 3). Furthermore, we found that 17 out of the 25 fine-mapped genes were overlapped (gene body and promoter) with DMRs (Supplementary Data 4), and eight of them were reported to be transcriptionally active during early embryonic development (from four cells to blastocyst) before implantation<sup>43</sup>, including *ARID4B*, *HCFC2*, *NFYB*, *CPSF1*, *WWP2*, *ZNF613*, *NDEL1*, and *PCDHA13* (Supplementary Data 4). All of these results provide evidence that sperm methylation alterations influence gestation length through regulating fetal development, as well as suggest that epigenetic



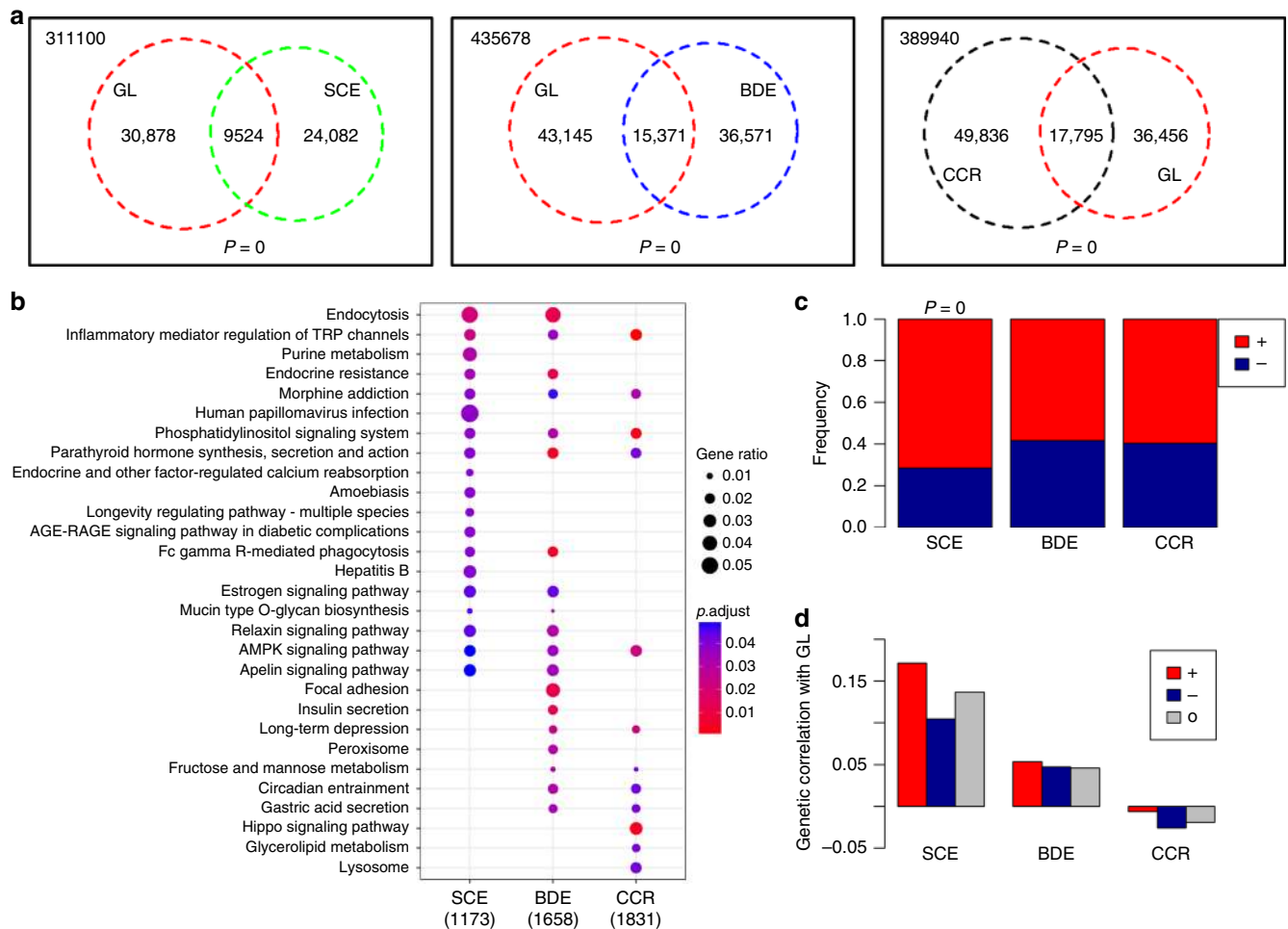
**Fig. 3** Sperm DNA methylation alterations associated with gestation length. **a** Enrichment of differentially methylated regions (DMRs) across multiple genomic elements. The enrichment of DMRs on a genomic element is computed as the ratio between the observed density of DMRs in this particular genomic element and the expected density of DMRs in the entire genome. **b** GWAS signal enrichment of gestation length based on DMRs that were determined by comparing animals with high gestation length to those with low gestation length. DMRs are first defined based on five different  $q$  values cutoffs (i.e., 0.05, 0.01, 1e-5, 1e-8, and 1e-10) and the absolute value of methylation difference >5% (Both). DMRs were then divided into two subsets according to the sign of methylation difference, i.e., <-5% (Loss) and > 5% (Gain). **c** DMR-set enrichment analysis based on the Reactome pathway database

alterations in germline cells induced by environmental perturbations may impact complex phenotypes through a transgenerational inheritance model<sup>44</sup>, potentially contributing to the genetic architecture underlying complex traits and diseases.

Since sire calving ease (SCE) and body depth (BDE) and cow conception rate (CCR) were genetically correlated with gestation length (Fig. 2a), we hypothesized that they also share the epigenetic architecture in sperm with gestation length. To test this, we further detected DMRs associated with SCE, BDE, and CCR based on the comparisons of high-SCE vs. low-SCE bulls, high BDE vs. low BDE, and high CCR vs. low CCR, respectively. We found that gestation length associated DMRs significantly overlapped with DMRs that were associated with SCE, BDE and CCR, respectively (Fig. 4a). There were 9524, 15,371, and 17,795 DMRs shared in gestation length & SCE, gestation length & BDE, and gestation length & CCR, respectively, which were used as three interesting DMR groups for further comparisons (Fig. 4a).

Based on DMRs in these three groups, we defined three groups of genes that overlapped (gene body and promoter) with them. We found that genes in the three groups were commonly and significantly ( $FDR < 0.05$ ) engaged in parathyroid hormone metabolism, inflammatory mediator regulation of TRP channels, phosphatidylinositol signaling system, and AMPK signaling pathway (Fig. 4b). Notably, for the shared DMRs in gestation length & SCE group, genes were selectively and significantly engaged in the longevity regulating pathway, endocrine-regulated calcium reabsorption and amoebiasis (Fig. 4b), implying their important roles in pregnancy and calving<sup>45-48</sup>. For the shared DMRs in the gestation length & BDE group, genes were significantly engaged in the focal adhesion, insulin secretion and peroxisome (Fig. 4b), suggesting their roles in the regulation of embryo growth<sup>49,50</sup>. For the shared DMRs in the gestation length & CCR group, genes were selectively and significantly engaged in the hippo signaling pathway, glycerolipid metabolism and lysosome (Fig. 4b), indicating their potential roles in the



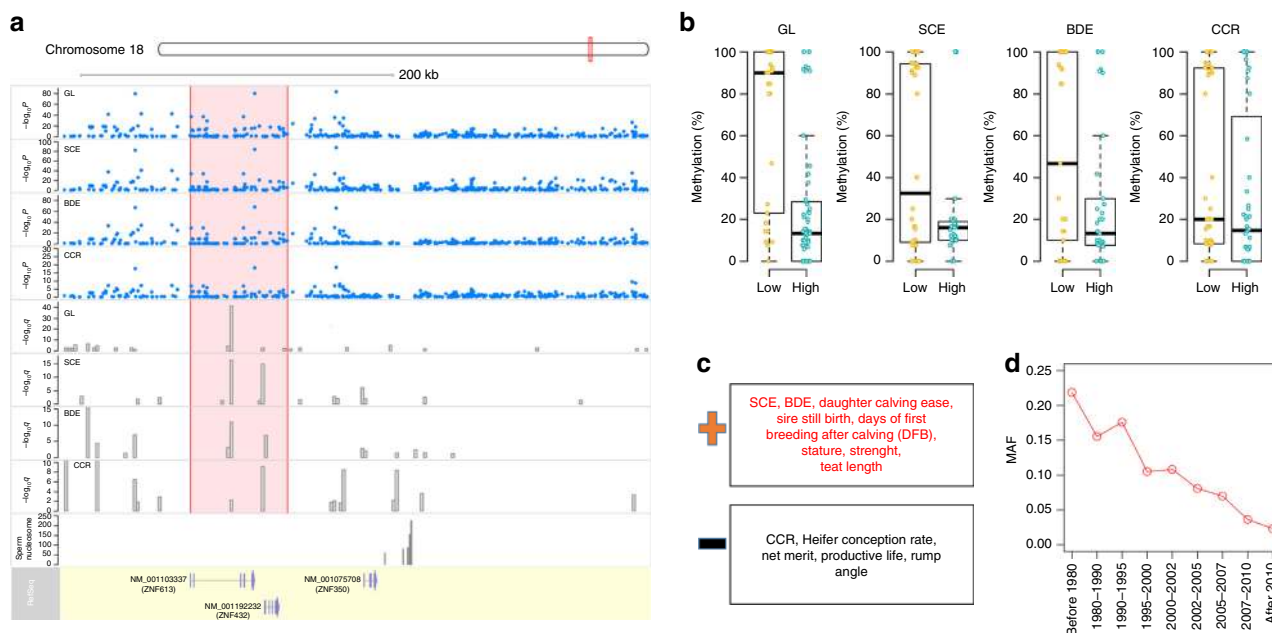


**Fig. 4** Relationships of sperm methylation alterations associated with gestation length, sire calving ease (SCE), body depth (BDE), cow conception rate (CCR). **a** Intersections between differentially methylated regions (DMRs) of gestation length and DMRs of SCE, BDE and CCR, respectively.  $P$ -value is calculated using a Fisher-exact test. **b** KEGG pathway functional enrichment of genes that are overlapped with shared DMRs in gestation length & SCE, gestation length & BDE, and gestation length & CCR, respectively. **c** Frequency of shared DMRs with the same (+) or opposite (-) change directions in gestation length & SCE, gestation length & BDE, and gestation length & CCR, respectively. **d** Comparisons of genetic correlations of gestation length & SCE, gestation length & BDE, and gestation length & CCR within their shared DMRs of same (+) or opposite (-) change directions, and over the entire genome (O), respectively

regulation of conception and fertility<sup>51</sup>. In Fig. 4c, for the shared DMRs in the gestation length & SCE group, 71% of the DMRs had the same direction of change for gestation length and SCE (Fig. 4c). In contrast, lower proportions of DMRs (58 and 60%) had the same change direction for gestation length & SCE and gestation length & CCR groups (Fig. 4c). This finding was in line with the genetic evidence that gestation length was more genetically correlated with SCE than with BDE and CCR. Furthermore, we observed that the genetic correlation between SCE and gestation length was higher in their shared DMRs of the same change direction, compared to those of opposite directions or the entire genome background (Fig. 4d). These results further supported that the epigenetic alterations in sperm associated with the genetic architecture underlying gestation length and other dairy traits.

**Multiple evidence implicating *ZNF613* for gestation length.** The most significant QTL of gestation length was on BTA18, which was also significantly associated with SCE, BDE, and CCR

(Supplementary Fig. 4). This was consistent with previous studies that reported this QTL to be associated with many fertility and body conformation traits in cattle, including gestation length<sup>21</sup>, stillbirth<sup>52</sup>, calving ability<sup>52–58</sup>, longevity<sup>59</sup>, calf birth weight<sup>19,60</sup>, young stock survival<sup>61</sup>, conformation<sup>54–56,62</sup>, and udder types<sup>55</sup>. However, the high extent of linkage disequilibrium and lack of functional annotation in this QTL region hampered the efforts to determine the causal gene and variants. Here, the fine-mapping analysis in our current and previous studies<sup>63</sup> identified *ZNF613* as the candidate gene for all of the four traits, including gestation length, SCE, BDE and CCR, which was consistent with previous studies in the Nordic dairy cattle population that proposed *ZNF613* to be associated with calving difficulty and longevity<sup>59,61</sup>. Of note, *ZNF613* had a common DMR on its second intron across all these four traits, and this common DMR was also the top one for gestation length in the corresponding genomic region (Fig. 5a). We further observed that animals with higher gestation length, SCE, BDE, and CCR had lower methylation levels in this particular DMR (Fig. 5b), which implied that the loss of methylation in the second intron of *ZNF613* could be associated with a



**Fig. 5** Genetic, epigenetic, and selection evidence implicating the *ZNF613* gene on gestation length. **a** GWAS signals (point-plot) and differentially methylated regions (DMR, bar plot) around *ZNF613* for gestation length, sire calving ease (SCE), body depth (BDE) and cow conception rate (CCR). Sperm-retained nucleosome is also shown below DMRs. **b** Methylation levels of the shared DMR (the second intron of *ZNF613*) in the two compared groups across these four traits. **c** Comparison of signs of variant effects (i.e.,  $b$ , based on signal-marker GWAS) within *ZNF613* between gestation length and 13 dairy traits, for which *ZNF613* is also a candidate gene based on our previous fine mapping analyses. + denotes the same direction, and - denotes the opposite direction. **d** The changes of minor allele frequency (MAF) of the lead variant (chr18:58141989;  $P = 7.97 \times 10^{-84}$ ) within *ZNF613* over the years from 1952 to 2012

prolonged gestation, a more difficult calving, a bigger body size, but a higher conception rate. By examining the Bovine Gene Atlas data, we found that *ZNF613* was expressed in 85 out of the 87 tissues, but was highly expressed in the thyroid, placenta above cotyledon, fetal tongue surface, anterior pituitary and hippocampus, suggesting its potential role in the regulation of embryonic development. We further validated that *ZNF613* was highly expressed in the embryonic brain and ovary, intercaruncular tissue (e.g., placenta), thyroid and testes among 174 tissues in sheep<sup>64</sup> (Supplementary Data 5). Its orthologous gene in mouse, *zfp157*, was highly expressed in the early conceptus and embryo ectoderm, and was associated with phenotypes related to endocrine/exocrine glands, integument and reproductive system (<http://www.informatics.jax.org/marker/MGI:1919404>). In addition, *ZNF613* was reported to be transcriptionally active at the blastocyst stage before implantation in cattle<sup>43</sup>. We further confirmed this by observing that nucleosomes were retained around *ZNF613* in the cattle mature sperm<sup>65,66</sup> (Fig. 5b), as sperm-retained nucleosomes package genes for the embryonic development<sup>67</sup>. Thus, it is tempting to propose that *ZNF613* functions at the very early embryonic developmental stage after fertilization, and that its epigenetic marks in sperm play important roles in the regulation of gestation length by influencing the fetal development, as well as that *ZNF613* is related to many fertility and conformation traits potentially due to its effect on gestation length.

Our previous fine-mapping study on 35 complex traits with the same dataset demonstrated that *ZNF613* was also the fine-mapped gene for eight dairy traits (average posterior probability of causality = 0.418), including days of first breeding after calving (DFB), heifer conception rate, net merit, productive life, rump angle, sire still birth, stature, strength and teat length<sup>63</sup>. By

comparing the sign of marker effects in *ZNF613* across all these associated traits, we found that gestation length was under the same selection direction with SCE, daughter calving ease, sire still birth, DFB, stature, strength and teat length, while gestation length was under the opposite selection direction with CCR, heifer conception rate, net merit and productive life (Fig. 5c). This finding was consistent with the epigenetic evidence that animals with higher methylation levels on the second intron of *ZNF613* had a prolonged gestation, a more difficult calving and a bigger body size; however these animals had a higher conception rate. We found that the minor allele frequency (MAF) of the top associated SNP (chr18:58141989;  $P = 7.97 \times 10^{-84}$ ) in *ZNF613* was 0.07 in the current U.S. Holstein population. However, the MAF of this SNP decreased dramatically over the years from 1952 to 2012 (Fig. 5d) due to the strong selection against it, suggesting that current and future selection may quickly remove this undesired variant out of the cattle population. These results demonstrated an example where selection shaped the genetic architecture of complex traits.

## Discussion

The data generated from the dairy industry is a valuable resource that provides a unique opportunity to investigate the paternal contribution to pregnancy related traits such as gestation length. In this study, we combined both genomics and sperm epigenomics data to dissect the paternal effects on gestation. Our GWAS study identified nine QTLs of gestation length passing the genome-wide significance level. Using a Bayesian fine-mapping analysis, we identified 25 candidate genes of gestation length. By investigating multiple sources of functional genomics data, we showed strong evidence supporting *ZNF613* as the candidate gene

for the most significant QTL on BTA 18. Despite the evidence we showed in this study, direct functional studies are still needed to validate the candidate genes reported here in the future.

In conclusion, our study is the first to explore the paternal genetic and epigenetic contributions to gestation length in a large cattle population. We demonstrate that paternal effects on gestation length may occur through regulating the embryo development, and the underlying epigenetic architecture of gestation length in sperm correlates with its genetic architecture. Gestation length shares both genetic and epigenetic architecture with other dairy traits of economic importance, such as SCE, BDE and CCR. In addition, we provide genetic, epigenetic, and selection evidence implicating *ZNF613* as the candidate gene for the major QTL on BTA18, indicating that both epigenetic alterations and selection pressures could contribute to the genetic architecture underlying complex traits. Our study also demonstrates the usefulness of integrating multiple layers of biological information to understand the phenotypic variation of complex traits.

## Methods

**Single-marker GWAS.** The phenotype and genotype data have been described in previous studies<sup>63,68</sup>. The phenotypes were de-regressed breeding values (predicted transmitting abilities or PTA) with high reliability for 27,214 Holstein bulls, which have been adjusted for known systematic effects including herd, year, season, and parity<sup>10</sup>. The high-density SNP genotypes ( $n = 312$  K) of all the individuals were imputed to sequence variants ( $n = 3,148,506$ ) with an imputation accuracy of 96.7%<sup>68</sup> using reference data from Run 5 of the 1000 Bull Genomes Project<sup>69</sup>. The imputation was conducted with the FindHap software (<https://aipl.arsusda.gov/software/findhap>). A total of 2,619,418 imputed variants with minor allele frequency (MAF) > 0.01 and Hardy–Weinberg Equilibrium (HWE) test ( $P > 1e-06$ ) were kept for further analyses.

Details of the single-marker GWAS analyses were described in<sup>63</sup>. Briefly, a linear mixed model, implemented in MMAP (<https://mmap.github.io/>), was employed to test for association of the imputed sequence variants:

$$y = \mu + Xb + g + e$$

where  $y$  is the de-regressed PTA,  $\mu$  is the overall mean,  $X$  is the genotype of a candidate marker (coded as 0, 1, or 2),  $b$  is the marker effect,  $g \sim N(0, \sigma_g^2 G)$  is the polygenic effect accounting for familial relationship and population structure, and  $e \sim N(0, \sigma_e^2 R)$  is the residual.  $G$  is the genomic relationship matrix<sup>70</sup>, which was built using HD markers with MAF > 0.01.  $R$  is a diagonal matrix with  $R_{ii} = 1/r_i^2 - 1$ , where  $r_i^2$  is the reliability of phenotype for the  $i$ th individual. A tested variant with  $P < 1.91e-08$  (Bonferroni correction) was considered significant at the genome-wide level.

Besides gestation length, we have also analyzed 35 other dairy traits in a previous GWAS<sup>63</sup>. These 35 traits were clustered into three groups based on their genetic correlations, including 17 body type, 12 reproduction, and 6 production traits. The pair-wise genetic correlations between gestation length and other 35 traits were approximately computed using the Pearson's correlation of effects ( $b$ ) of all tested variants<sup>71</sup>.

**Gene-level fine-mapping analysis.** The genomic regions used for fine-mapping analyses were defined by extending the QTL regions (i.e., the minimal regions covering all significant SNPs) 1 Mb upstream and downstream. We conducted fine-mapping analysis for each of these candidate regions by employing our Bayesian fine-mapping approach (BFMAP) (<http://terpconnect.umd.edu/~jiang18/bfmap/>)<sup>63</sup>, which follows a similar framework as Huang et al.<sup>72</sup>. BFMAP attempted to determine independent association signals in a genomic region by assessing a posterior probability of causality (PPC) for each variant within this particular region. The PPC of a gene was calculated as the sum of PPCs of all variants that were located within 2 Kb upstream and downstream of the corresponding gene. BFMAP has been shown to have at least equal power as the commonly used fine-mapping approaches<sup>63</sup>, such as PAINTOR<sup>73</sup> and CAVIARBF<sup>74</sup>.

**Public gene annotation sources and genomic features.** We used org.Bt.eg.db v. 3.6.0, reactome.db v. 1.64.0, packages, MeSH v. 1.10.0, which have been implemented in Bioconductor v. 3.7. (<https://bioconductor.org/packages/release/data/annotation/html/org.Bt.eg.db.html>) to link genes to GO terms, KEGG pathways, Reactome pathways, and MeSH terms. A total of 755 miRNA-target networks were built as previously described<sup>75</sup>. Briefly, 755 cattle miRNAs were obtained from miRbase (<http://www.mirbase.org/>). The miRmap software<sup>76</sup> was used to predict the targets of each miRNA, and only the top 25% of predicted targets were considered. For a given trait in the Cattle QTLdb (release 35, April 29 2018, <https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>), we considered genes that were located inside or in the closest proximity to the QTL regions as associated genes with this particular trait. We excluded traditional QTL mapping results in the QTLdb due to their large QTL regions. In the end, we kept the GO terms ( $n = 889$ ), KEGG pathways ( $n = 225$ ), Reactome pathways ( $n = 804$ ), miRNA-target networks ( $n = 755$ ), and trait-associated gene networks ( $n = 208$ ) that comprised of at least 10 genes as genomic features for further GWAS signal enrichment analyses. For instance, a GO term with  $\geq 10$  genes can be considered as a genomic feature.

**GWAS signal enrichment analysis.** Because the complex phenotypes being analyzed are highly polygenic or even omnigenic<sup>27,77</sup>, we employed the following sum-based marker-set test approach (<http://psoerensen.github.io/qgg/>) to determine whether the GWAS signals were enriched in a predefined genomic feature (i.e., a gene list defined using the above annotation sources, including GO terms, KEGG, and Reactome pathways, and miRNA-targets, trait-associated gene networks, and differentially methylation regions). This enrichment analysis is based on all variants in a GWAS study rather than the top variants passing genome-wide significance level. Previous studies have demonstrated that this approach has at least equal power to many commonly used GWAS signal enrichment methods in human<sup>78</sup>, *Drosophila melanogaster*<sup>79</sup> and livestock species<sup>80–82</sup>, particularly in highly polygenic phenotypes. The sum-based statistics can be expressed as

$$T_{sum} = \sum_{i=1}^{m_j} b_i^2$$

where  $m_j$  is the number of genomic variants in a pre-defined genomic feature, and  $b$  is the variant effect. Here, SNPs located in different genes within a genomic feature (e.g., a biological pathway) were often not in linkage disequilibrium (LD). This approach is similar to the popular LD score regression in human studies<sup>83</sup>, and it controls LD patterns among variants and variant-set sizes through employing the following genotype cyclical permutation strategy<sup>78,79</sup>. In brief, we first ordered the test statistics (i.e.,  $b^2$ ) for all variants according to their chromosome positions (i.e.,  $b_1^2, b_2^2, \dots, b_{m-1}^2, b_m^2$ ). We then randomly chose one test statistic (i.e.,  $b_k^2$ ) from this vector as the first, and shifted the remaining test statistics to new locations, while retained their original orders (i.e.,  $b_k^2, b_{k+1}^2, \dots, b_{m-1}^2, b_m^2, b_1^2, \dots, b_{k-1}^2$ ) to maintain the correlation patterns among variants. We calculated a new summary statistic for the tested genomic feature using its original genomic positions. We repeated this permutation procedure 10,000 times for each genomic feature being tested, and obtained an empirical  $P$ -value by using a one-tailed test of the proportion of random summary statistics greater than that observed.

**Sperm WGBS data and bioinformatics analyses.** No animal experiments were conducted in this study, and ethics committee approval was thus not required. All sperm methylation data were generated in previous studies, and references were provided where animal data were used.

The sperm whole-genome bisulfite sequencing (WGBS) data were generated in previous studies<sup>66</sup>. All the semen samples used in this study were collected from bulls by an artificial insemination company using a standardized procedure with artificial vaginas. Briefly, one semen straw (0.5 ml) often contains 10–40 million sperm, and is transported and stored in a liquid nitrogen tank. After thawing a semen straw, PBS buffer was used to wash away the extender for three times by mild centrifugations. Visual examination of washed sperm samples was conducted under a microscope, and over 90% of sperm cells were found morphologically normal. At first, eight semen straws were sampled from eight representative Holstein bulls, among which four animals have high PTA of gestation length and the other four with low PTA of gestation length (Supplementary Table 1). For other traits such as SCE, BDE, and CCR, we also collected six semen samples with two groups (i.e., high PTA vs. low PTA) for each trait, and each group then had three biological replicates. In total, all of these sperm samples were from 18 fertile, age-matched and representative animals, because several animals had extreme values for multiple traits. The reliabilities of PTA for all of the 18 animals were greater than 0.97, 0.93, 0.87 and 0.83 for gestation length, SCE, BDE and CCR, respectively. Genomic DNA was isolated using QIAamp DNA Mini Kit protocol (QIAGEN, Valencia, CA, USA). The quality of isolated DNA was evaluated using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The sequence libraries were constructed using all of the qualified genomic DNA<sup>66</sup>, and were then sequenced using HiSeq X 10 (Illumina, San Diego, CA, USA) with a 150 bp paired-end technology.

FastQC v0.11.2 (<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/)) were employed to check and clean the raw data, respectively<sup>66</sup>. Generally, adapters and reads with low quality ( $Q < 20$ ) or shorter than 20 bp were removed. All the cleaned data were mapped to the cattle reference genome (UMD 3.1) using bowtie2<sup>84</sup> with an average mapping rate of 69.23% (ranging from 53.10 to 78.90%). The total number of mapped reads per sample ranged from 134,984,436 to 227,534,395 with an average of 171,982,028. Bismark software<sup>85</sup> was applied to extract methylcytosine information. Only the loci that were covered by



at least 10 clean reads were kept for further analyses. More details have been described previously<sup>66</sup>.

Because methylation alterations that are associated with complex traits often exhibit spatial correlation patterns<sup>86</sup>, DMR (differentially methylated region) instead of DMC (differentially methylated cytosine) were determined by using methylKit<sup>87</sup>. The entire genome was first tiled into windows of 2000 bp in length and 2000 bp step-size, and then methylation levels on those tiles were summarized. A logistic regression model implemented in the *calculateDiffMeth* function was employed to detect DMR: information from each sample is specified (the number of methylated Cs and number of unmethylated Cs at a given region), and a logistic regression test is applied to compare fraction of methylated Cs across the two groups under comparison. *P*-values were calculated through comparing the model fitness of alternative models to the null model, and were corrected to *q*-values for multiple testing using the SLIM method<sup>88</sup>. Since methylation alterations associated with complex phenotypes are generally very weak<sup>86,89</sup>, the absolute value of difference in methylation >5% and different *q* cutoffs (i.e., 0.05, 0.01, 1e-5, 1e-8, and 1e-10) were used to define DMR for the downstream analyses.

**DMR-set enrichment analysis.** The following count-based approach was used to test whether a Reactome pathway was enriched for DMR,

$$T_{count} = \sum_{i=1}^{m_f} I(q_i < q_0), \quad (2)$$

where  $m_f$  is the total number of 2000 bp-tiles being tested that were overlapped (at least 1 bp) with genes in a pathway,  $q_i$  is the *q* value for the *i*th tested tile,  $q_0$  is an arbitrarily chosen threshold, and *I* is an indicator function that takes value 1 when  $q_i < q_0$ , and value zero otherwise. Here  $q_0 = 0.01$  was used tentatively, and gene regions were extended 10 Kb upstream and downstream to cover potential regulatory regions. Under the null hypothesis (i.e., DMR are distributed in the genome randomly),  $T_{count}$  was assumed to follow a hypergeometric distribution:  $T_{count} \sim \text{Hyper}(m, m_g, m_f)$  where *m* is the total number of tiles being tested in the entire genome,  $m_g$  is the total number of DMR detected in the entire genome, and  $m_f$  is the number of tiles being tested in a pathway. The null hypothesis (i.e., no enrichment) will be rejected when the adjusted *P*-value is less than 0.05. Here, a tile was considered belonging to a pathway if the tile intersected any gene ( $\pm 10$  Kb up- and down-stream) in the particular pathway.

Functional enrichment analyses for gene lists in DMRs were conducted using the R package clusterProfiler<sup>90</sup>, where a hypergeometric test was employed using the current KEGG database. *P*-values were adjusted for multiple testing using the FDR method<sup>91</sup>, and FDR < 0.05 was considered as significant.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All the 18 cattle sperm methylomes have been submitted to NCBI with accession number GSE119263. All genomic annotation files of cattle (UMD 3.1.1) are available for download from Ensembl database (<https://uswest.ensembl.org/index.html>). The authors confirm that the original genotype data are owned by the Council on Dairy Cattle Breeding (CDCB). A request to CDCB is necessary for getting data on research, which may be sent to: João Dürr, CDCB Chief Executive Officer ([joao.durr@cdcbs.us](mailto:joao.durr@cdcbs.us)). All other data have been shown in the manuscript and supplementary data.

Received: 29 August 2018 Accepted: 6 February 2019

Published online: 14 March 2019

## References

1. Arnott, G. et al. Board invited review: the importance of the gestation period for welfare of calves: maternal stressors and difficult births. *J. Anim. Sci.* **90**, 5021–5034 (2012).
2. Lemons, J. A. et al. Very low birth weight outcomes of the National Institute of Child health and human development neonatal research network, January 1995 through December 1996. *Pediatrics* **107**, e1–e1 (2001).
3. Yoshida, S. et al. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. *J. Glob. Health* **6**, 010508 (2016).
4. Huusko, J. M. et al. Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLoS Genet.* **14**, e1007394 (2018).
5. Vieira-Neto, A., Galvão, K., Thatcher, W. & Santos, J. Association among gestation length and health, production, and reproduction in Holstein cows and implications for their offspring. *J. Dairy Sci.* **100**, 3166–3181 (2017).
6. Nogalski, Z. & Piwczyński, D. Association of length of pregnancy with other reproductive traits in dairy cattle. *Asian-Australas. J. Anim. Sci.* **25**, 22 (2012).
7. Aghaepour, N. et al. An immune clock of human pregnancy. *Sci. Immunol.* **2**, ean2946 (2017).
8. Goedicke-Fritz, S. et al. Preterm Birth affects the risk of developing immune-mediated diseases. *Front. Immunol.* **8**, 1266 (2017).
9. Jukic, A. M., Baird, D. D., Weinberg, C. R., McConaughey, D. R. & Wilcox, A. J. Length of human pregnancy and contributors to its natural variation. *Hum. Reprod.* **28**, 2848–2855 (2013).
10. Norman, H. et al. Genetic and environmental factors that affect gestation length in dairy cattle. *J. Dairy Sci.* **92**, 2259–2269 (2009).
11. Morel, M. D., Newcombe, J. & Holland, S. Factors affecting gestation length in the Thoroughbred mare. *Anim. Reprod. Sci.* **74**, 175–185 (2002).
12. Clausson, B., Lichtenstein, P. & Cnattingius, S. Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG* **107**, 375–381 (2000).
13. York, T. P. et al. Fetal and maternal genes' influence on gestational age in a quantitative genetic analysis of 244,000 Swedish births. *Am. J. Epidemiol.* **178**, 543–550 (2013).
14. Schierding, W. et al. GWAS on prolonged gestation (post-term birth): analysis of successive Finnish birth cohorts. *J. Med. Genet.* **55**, 55–63 (2017). [jmedgenet-2017-104880](https://doi.org/10.1093/jmedgenet/2017-104880).
15. Zhang, G. et al. Genetic associations with gestational duration and spontaneous preterm birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).
16. Carrell, D. T. & Hammoud, S. S. The human sperm epigenome and its potential role in embryonic development. *Mol. Hum. Reprod.* **16**, 37–47 (2009).
17. Jenkins, T. G. & Carrell, D. T. The sperm epigenome and potential implications for the developing embryo. *Reproduction* **143**, 727–734 (2012).
18. Teperek, M. et al. Sperm is epigenetically programmed to regulate gene transcription in embryos. *Genome Res.* **26**, 1034–1046 (2016).
19. Cole, J. et al. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* **92**, 2931–2946 (2009).
20. Harhay, G. P. et al. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. *Genome Biol.* **11**, R102 (2010).
21. Maltecca, C., Gray, K., Weigel, K., Cassady, J. & Ashwell, M. A genome-wide association study of direct gestation length in US Holstein and Italian Brown populations. *Anim. Genet.* **42**, 585–591 (2011).
22. Xiao, X. et al. HSF1 is required for extra-embryonic development, postnatal growth and protection during inflammatory responses in mice. *EMBO J.* **18**, 5943–5952 (1999).
23. Ma, X. & Adelstein, R. S. A point mutation in Myh10 causes major defects in heart development and body wall closure. *Circulation* **113**, 000455 (2014).
24. Ridge, L. A. et al. Non-muscle myosin IIB (Myh10) is required for epicardial function and coronary vessel formation during mammalian development. *PLoS Genet.* **13**, e1007068 (2017).
25. Sasaki, S. et al. Complete loss of Ndel1 results in neuronal migration defects and early embryonic lethality. *Mol. Cell. Biol.* **25**, 7812–7827 (2005).
26. Zhao, Y.-Y. et al. Neuregulin signaling in the heart: dynamic targeting of erbB4 to caveolar microdomains in cardiac myocytes. *Circ. Res.* **84**, 1380–1387 (1999).
27. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
28. Furness, D., Dekker, G. & Roberts, C. DNA damage and health in pregnancy. *J. Reprod. Immunol.* **89**, 153–162 (2011).
29. Wang, Y., Bikle, D. D. & Chang, W. Autocrine and paracrine actions of IGF-I signaling in skeletal development. *Bone Res.* **1**, 249 (2013).
30. Gray, C., McCowan, L. M., Patel, R., Taylor, R. S. & Vickers, M. H. Maternal plasma miRNAs as biomarkers during mid-pregnancy to predict later spontaneous preterm birth: a pilot study. *Sci. Rep.* **7**, 815 (2017).
31. Hosseini, M. K., Gunel, T., Gumusoglu, E., Benian, A. & Aydinli, K. MicroRNA expression profiling in placenta and maternal plasma in early pregnancy loss. *Mol. Med. Rep.* **17**, 4941–4952 (2018).
32. Ioannidis, J. & Donadeu, F. X. Changes in circulating microRNA levels can be identified as early as day 8 of pregnancy in cattle. *PLoS One* **12**, e0174892 (2017).
33. Schanzbach, C. I., Kirchner, B., Ulbrich, S. E. & Pfaffl, M. W. Can milk cell or skim milk miRNAs be used as biomarkers for early pregnancy detection in cattle? *PLoS One* **12**, e0172220 (2017).
34. Ioannidis, J. & Donadeu, F. X. Circulating miRNA signatures of early pregnancy in cattle. *BMC Genom.* **17**, 184 (2016).
35. Laresgoiti-Servitje, E. Pregnancy-related miRNAs participate in the regulation of the immune system during the gestational period. *J. Clin. Cell Immunol.* **6**, 2 (2015).
36. Cai, M., Kolluru, G. K. & Ahmed, A. Small molecule, big prospects: microRNA in pregnancy and its complications. *J. Pregnancy* **2017**, 6972732 (2017).

37. Hu, Z.-L. & Reecy, J. M. Animal QTLdb: beyond a repository. *Mamm. Genome* **18**, 1–4 (2007).
38. Reitz, R. E., Daane, T. A., Woods, J. R. & Weinstein, R. L. Calcium, magnesium, phosphorus, and parathyroid hormone interrelationships in pregnancy and newborn infants. *Obstet. Gynecol.* **50**, 701–705 (1977).
39. Keen, C. L. et al. Effect of copper deficiency on prenatal development and pregnancy outcome. *Am. J. Clin. Nutr.* **67**, 1003S–1011S (1998).
40. Heringstad, B., Gianola, D., Chang, Y., Ødegård, J. & Klemetsdal, G. Genetic associations between clinical mastitis and somatic cell score in early first-lactation cows. *J. Dairy Sci.* **89**, 2236–2244 (2006).
41. Hande, M. DNA repair factors and telomere-chromosome integrity in mammalian cells. *Cytogenet. Genome Res.* **104**, 116–122 (2004).
42. Ménéz, Y., Dale, B. & Cohen, M. DNA damage and repair in human oocytes and embryos: a review. *Zygote* **18**, 357–365 (2010).
43. Graf, A. et al. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc. Natl Acad. Sci. USA* **111**, 4139–4144 (2014).
44. Johannes, F. et al. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e1000530 (2009).
45. Armon, P. Amoebiasis in pregnancy and the puerperium. *BJOG* **85**, 264–269 (1978).
46. Pitkin, R. M. Endocrine regulation of calcium homeostasis during pregnancy. *Clin. Perinatol.* **10**, 575–592 (1983).
47. Tiezzi, F., Arceo, M. E., Cole, J. B. & Maltecca, C. Including gene networks to predict calving difficulty in Holstein, Brown Swiss and Jersey cattle. *BMC Genet.* **19**, 20 (2018).
48. de Maturana, E. L., Ugarte, E. & González-Recio, O. Impact of calving ease on functional longevity and herd amortization costs in Basque Holsteins using survival analysis. *J. Dairy Sci.* **90**, 4451–4457 (2007).
49. Ashworth, M., Leach, F. & Milner, R. Development of insulin secretion in the human fetus. *Arch. Dis. Child.* **48**, 151 (1973).
50. Shiokawa, S. et al. Functional role of focal adhesion kinase in the process of implantation. *Mol. Hum. Reprod.* **4**, 907–914 (1998).
51. Kawamura, K. et al. Hippo signaling disruption and Akt stimulation of ovarian follicles for infertility treatment. *Proc. Natl Acad. Sci. USA* **110**, 17474–17479 (2013).
52. Thomassen, J., Guldbrandtsen, B., Sørensen, P., Thomsen, B. & Lund, M. Quantitative trait loci affecting calving traits in Danish Holstein cattle. *J. Dairy Sci.* **91**, 2098–2105 (2008).
53. Müller, M.-P. et al. Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *J. Dairy Sci.* **100**, 1987–2006 (2017).
54. Mao, X. et al. Fine mapping of a calving QTL on Bos taurus autosome 18 in Holstein cattle. *J. Anim. Breed. Genet.* **133**, 207–218 (2016).
55. Brand, B. et al. Quantitative trait loci mapping of calving and conformation traits on Bos taurus autosome 18 in the German Holstein population. *J. Dairy Sci.* **93**, 1205–1215 (2010).
56. Pausch, H., et al. Genome-wide association study identifies two major loci affecting calving ease and growth related traits in cattle. *Genetics* **187**, 289–297 (2010).
57. Purfield, D., Bradley, D., Kearney, J. & Berry, D. Genome-wide association study for calving traits in Holstein–Friesian dairy cattle. *Animal* **8**, 224–235 (2014).
58. Purfield, D. C., Bradley, D. G., Evans, R. D., Kearney, F. J. & Berry, D. P. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. *Genet. Sel. Evol.* **47**, 47 (2015).
59. Zhang, Q., Guldbrandtsen, B., Thomassen, J. R., Lund, M. S. & Sahana, G. Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds. *J. Dairy Sci.* **99**, 7289–7298 (2016).
60. Cole, J., Waurich, B., Wensch-Dorendorf, M., Bickhart, D. & Swalve, H. A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. *J. Dairy Sci.* **97**, 3156–3172 (2014).
61. Wu, X., Guldbrandtsen, B., Nielsen, U. S., Lund, M. S. & Sahana, G. Association analysis for young stock survival index with imputed whole-genome sequence variants in Nordic Holstein cattle. *J. Dairy Sci.* **100**, 6356–6370 (2017).
62. Magee, D. A. et al. DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein–Friesian cattle. *BMC Genet.* **11**, 93 (2010).
63. Jiang J., Cole J. M., Da Y., VanRaden P. M., Ma L. Fast Bayesian fine-mapping of 35 production, reproduction and body conformation traits with imputed sequences of 27K Holstein bulls. *bioRxiv*, 428227 (2018).
64. Clark, E. L. et al. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLoS Genet.* **13**, e1006997 (2017).
65. Samans, B. et al. Uniformity of nucleosome preservation pattern in Mammalian sperm and its connection to repetitive DNA elements. *Dev. Cell.* **30**, 23–35 (2014).
66. Zhou, Y. et al. Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm. *Gigascience* **7**, giy039 (2018).
67. Hammoud, S. S. et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473 (2009).
68. VanRaden, P. M., Tooker, M. E., O’Connell, J. R., Cole, J. B. & Bickhart, D. M. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Sel. Evol.* **49**, 32 (2017).
69. Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858 (2014).
70. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
71. Zhu, X. et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Human. Genet.* **96**, 21–36 (2015).
72. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173 (2017).
73. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
74. Chen, W., McDonnell, S. K., Thibodeau, S. N., Tillmans, L. S. & Schaid, D. J. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. *Genetics* **116**, 188953 (2016).
75. Fang, L. 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.* **8**, 9345 (2018).
76. Vejnar, C. E. & Zdobnov, E. M. MiRmap: comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.* **40**, 11673–11683 (2012).
77. Kemper, K. E. & Goddard, M. E. Understanding and predicting complex traits: knowledge from cattle. *Hum. Mol. Genet.* **21**, R45–R51 (2012).
78. Rohde, P. D., Demontis, D., Cuyabano, B. C. D., Børglum, A. D. & Sørensen, P. Covariance association test (CVAT) identifies genetic markers associated with schizophrenia in functionally associated biological processes. *Genetics* **203**, 1901–1913 (2016).
79. Sørensen, I. F., Edwards, S. M., Rohde, P. D. & Sørensen, P. Multiple trait covariance association test identifies gene ontology categories associated with chill coma recovery time in *Drosophila melanogaster*. *Sci. Rep.* **7**, 2413 (2017).
80. Sarup, P., Jensen, J., Ostensen, T., Henryon, M. & Sørensen, P. Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. *BMC Genet.* **17**, 11 (2016).
81. Fang, L. 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.* **49**, 44 (2017).
82. Fang, L. et al. Integrating sequence-based GWAS and RNA-Seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. *Sci. Rep.* **7**, 45560 (2017).
83. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228 (2015).
84. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
85. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
86. Guo, S. et al. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635 (2017).
87. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
88. Wang, H.-Q., Tuominen, L. K. & Tsai, C.-J. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* **27**, 225–231 (2010).
89. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. *Nat. Rev. Genet.* **19**, 129 (2018).
90. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
91. Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**:289–300 (1995).

## Acknowledgements

This work was supported in part by AFRI grants number 2013-67015-20951 and number 2016-67015-24886 from the USDA National Institute of Food and Agriculture (NIFA) and BARD grant number US-4997-17 from the US-Israel Binational Agricultural

Research and Development (BARD) Fund. We thank the 1000 Bull Genomes Project for providing genome references for sequence imputation.

### Author contributions

L.F., J.J., G.E.L., and L.M. conceived and designed the experiments. P.M.V. and J.B.C. provided genotype and phenotype data, Y.Z. and G.E.L. collected samples and/or generated sperm methylation data. L.F., J.J., and E.F. performed computational and statistical analyses. L.F., J.J., B.L., and L.M. wrote the paper. All authors read and approved the final manuscript.

### Additional information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42003-019-0341-6>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019