

# SCIENTIFIC REPORTS



OPEN

## Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility

Xiuge Wang<sup>1</sup>, Chunhong Yang<sup>1</sup>, Fang Guo<sup>1</sup>, Yaran Zhang<sup>1</sup>, Zhihua Ju<sup>1</sup>, Qiang Jiang<sup>1</sup>, Xueming Zhao<sup>2</sup>, Yong Liu<sup>1</sup>, Han Zhao<sup>3</sup>, Jinpeng Wang<sup>1</sup>, Yan Sun<sup>1</sup>, Changfa Wang<sup>1</sup>, Huabin Zhu<sup>2</sup> & Jinming Huang<sup>1,3</sup>







Sperm motility is the main index used to assess the quality of bull semen. It may also be used to evaluate the fertility potential of bulls. Protein-coding mRNA and long noncoding RNA (lncRNA) participate in the regulation of spermatogenesis. Here, we employed strand-specific RNA sequencing to profile the semen transcriptome (mRNA and lncRNA) of six paired full-sibling Holstein bulls with divergent sperm motility and to determine the functions of mRNA and lncRNA in sperm motility. Among 20,875 protein-encoding genes detected in semen, 19 were differentially expressed between the high sperm motility group (H: H1, H2, and H3) and low sperm motility group (L: L1, L2, and L3). Of the 11,561 lncRNAs identified in sperm, 2,517 were differentially expressed between the H and L groups. We found that TCONS\_00041733 lncRNA targets the node gene *EFNA1* (*ephrin A1*), involved in male reproductive physiology. Our study provides a global mRNA and lncRNA transcriptome of bull semen, as well as novel insights into the regulation of neighboring protein coding by lncRNAs and the influence of mRNAs on sperm motility.

Bull fertility is the most important factor that influences the genetic quality of the dairy herd and plays a key role in the genetic improvement of the modern cattle population. Nevertheless, over the past several decades, the dairy cow industry has neglected bull fertility in favor of milk production as its main breeding objective<sup>1</sup>. Semen quality is usually assessed on the basis of several characteristics, such as ejaculate volume, density, motility, and deformity rate. Among these characteristics, sperm motility is the major index applied to determine whether fresh semen can be used to produce frozen-thawed semen. Inadequate sperm motility is one of the critical reasons for male subfertility or infertility<sup>2</sup>. Failure in semen production due to poor sperm motility can result in significant economic loss. Usually, this is a practical problem that there is a time interval (10–12 months) from the genomic evaluation of bull to semen quality can be evaluated<sup>3</sup>. In particular, full-sibling Holstein bulls may have different sperm motilities despite their similar genetic backgrounds and estimated genomic breeding values. Estimated heritability for sperm motility of dairy bull is 0.43<sup>4</sup>. However, the molecular mechanisms underlying the difference in the sperm motilities of different bull individuals remain poorly understood.

In several genome-wide association studies, candidate genes and single nucleotide polymorphisms underlying sperm motility have been identified in bulls<sup>5–9</sup>. Moreover, high-throughput technologies, including transcriptomics and proteomics, have revealed a lot of differences between the spermatozoa of high and low fertility bulls<sup>10–12</sup>. However, there is currently little insights on which regulatory mechanisms affect the difference of bull sperm motility. Therefore, the possible reason underlying the variation in sperm motility among bulls must be investigated to improve bull fertility.

A large number of unique coding and noncoding transcripts are specifically retained in mature sperm cells and have potential regulatory roles in fertilization, early embryo development, and spermatogenic processes<sup>13,14</sup>.

<sup>1</sup>Dairy Cattle Research Center, Shandong Academy of Agricultural Sciences, Jinan, 250131, P.R. China. <sup>2</sup>Embryo Biotechnology and Reproduction Laboratory, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100193, P.R. China. <sup>3</sup>College of Life Sciences, Shandong Normal University, Jinan, 250014, P.R. China. Xiuge Wang and Chunhong Yang contributed equally. Correspondence and requests for materials should be addressed to J.H. (email: [huangjinm@sina.com](mailto:huangjinm@sina.com))

Full-sibling bulls 1	Full-sibling bulls 2	Full-sibling bulls 3	Group
 H1 (gEBV: 154)	 H2 (gEBV: 138)	 H3 (gEBV: 133)	H (H1, H2, H3) (68.67% ± 1.15%)
 L1 (gEBV: 153)	 L2 (gEBV: 139)	 L3 (gEBV: 134)	

**Figure 1.** Three groups of paired full-sibling Holstein bulls. gEBV: genomic estimated breeding value. H: high sperm motility group, L: low sperm motility group. Data are presented as mean ± SE.

Notably, nonprotein-coding transcripts account for a greater percentage of the transcriptomic profile than protein-coding genes. In addition to various small noncoding RNAs (e.g. miRNAs, siRNA, and piRNAs), the transcriptome can generate a large class of long noncoding RNAs (lncRNAs) that exceed 200 nucleotides in length. lncRNAs were originally considered as “junk” genomic sequences but have recently been acknowledged to regulate gene expression. However, a growing body of evidence demonstrates that up to 90% of the noncoding transcripts in the human genome have important and diverse biological roles<sup>15</sup>. For example, lncRNAs regulate different genes at the transcriptional, posttranscriptional, and epigenetic levels by interacting with DNA, RNA, and proteins<sup>16–18</sup>. The forms of lncRNAs may vary in accordance with their genomic locations and include overlapping, intergenic, intronic, sense, and antisense transcripts<sup>19</sup>.

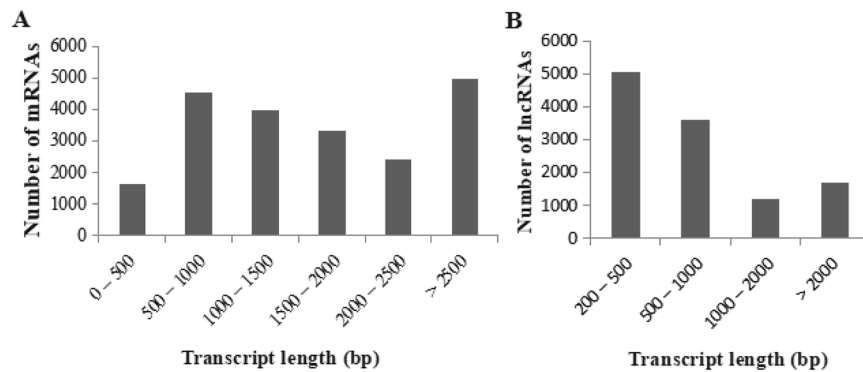
Recent RNA sequencing (RNA-Seq) studies have shown that several lncRNAs may regulate testis development and spermatogenesis through currently unknown mechanisms. For example, Mrhl lncRNA affects spermatogenesis<sup>20</sup>, and HongrES2 regulates sperm maturation<sup>21</sup>. Tsx is critical for the progression of spermatocyte meiosis and is specifically expressed in pachytene spermatocytes<sup>22</sup>. Dmrt1 is a testis-specific functional lncRNA that participates in the transition of germ cells from mitosis to meiosis<sup>23</sup>. However, the functions and regulatory mechanisms of lncRNAs remain unknown<sup>16,24</sup>. Moreover, studies on functional lncRNAs that affect bull sperm motility have not been reported. We attempted to address this research gap in our present study. We designed this study to obtain preliminary knowledge on the molecular mechanism that controls sperm motility. We performed the high-throughput sequencing of the bull semen transcriptome to identify the functional transcripts that are involved in bull sperm motility. Our data also provide new insights into the functional effects of lncRNAs on improving sperm motility and have implications for the development of the lncRNA-based molecular breeding of dairy cows.

## Results

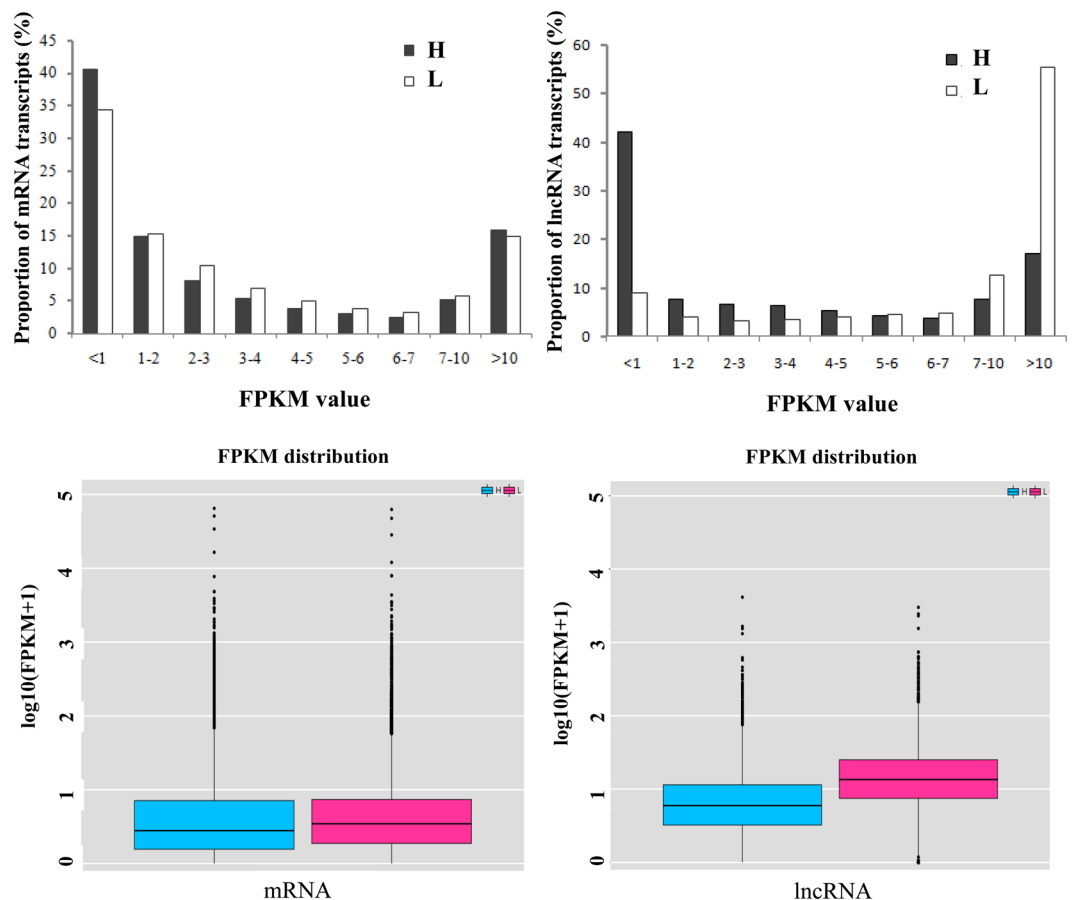
**Transcriptome sequencing of mRNAs and lncRNAs.** Three paired full-sibling Holstein bulls were selected from a bull station (Fig. 1). To construct the mRNA and lncRNA transcriptome profile of bull semen, we constructed the RNA-Seq sequencing libraries of six Holstein semen samples collected from the H and L groups. Through Illumina paired-end sequencing after strict filtering and removing low-quality raw reads, we acquired approximately 536 million clean reads that yielded 80.5 Gb high quality data. All Q20 (%) values of the read sequences in six samples exceeded 98% (Supplementary Tables S1–S2).

**Expressed patterns of mRNAs and lncRNAs.** We detected 2,0875 protein-coding genes when the remaining tRNAs, miscRNAs, and other smaller RNAs were filtered out. We used CPC and CNCI and strict filtering criteria to predict the coding potential of assembled transcripts. Finally, we identified 11,561 candidate novel lncRNAs that we further categorized into four categories (Supplementary Fig. S1). Statistical analysis showed that the transcripts of the candidate lncRNAs were shorter than those of mRNAs. The length range of lncRNAs was 200–1000 bp, which was narrower than the length range of mRNAs (Fig. 2). Approximately 36.2% of the protein-coding transcripts contained more than 10 exons, whereas no more than seven exons were detected in lncRNAs.

We estimated mRNA and lncRNA expression levels through the FPKM method by using the Cuffdiff program in Cufflinks package. We estimated the expression levels of lncRNA transcripts through FPKM. Approximately 80% of lncRNA transcripts had FPKM values of less than 2, and only approximately 10% had FPKM values of more than 10 (Fig. 3).



**Figure 2.** Genomic characteristics of mRNAs and lncRNAs. (A) Length distribution of coding mRNAs. (B) Length distribution of lncRNAs.



**Figure 3.** Expression levels of mRNAs and lncRNAs in the high sperm motility group (H) and low sperm motility group (L). Expression levels of mRNAs and lncRNAs in the H and L groups were compared through the FPKM method. Boxplots of FPKM distribution were calculated using all the mRNAs and lncRNAs expression data respectively. FPKM: fragments per kilobase of exon per million mapped fragments.

**Differentially expressed mRNAs and lncRNAs.** Interestingly, the expression levels of lncRNA transcripts were considerably higher than those of mRNAs (Fig. 3). Among 20,875 coding genes, 19 were differentially expressed between the H and L groups, including 9 up-regulated genes and 10 down-regulated genes ( $FDR \leq 0.05$ ,  $|\log_2(\text{fold change})| \geq 1$ ) (Table 1; Supplementary Fig. S2; Supplementary Table S3). Among 11,561 lncRNAs, 2,517 were differentially expressed between the H and L groups. These lncRNAs included 57 up-regulated genes and 2,460 down-regulated genes ( $FDR \leq 0.05$  and  $|\log_2(\text{fold change})| \geq 1$ ) (Supplementary Fig. S2; Supplementary Table S4). The top 20 significantly differentially expressed lncRNAs were shown in Table 2. To validate the results from the RNA-Seq data, we verified the expression levels of ten differentially expressed genes that were randomly selected from sequencing data, including five mRNAs and five lncRNAs, through qRT-PCR analysis. Our

gene	Ensembl Gene ID	locus	H(FPKM)	L(FPKM)	log <sub>2</sub> (fold change)	FDR value
NKX1-2	ENSBTAG00000020918	chr26:44416759-44420132	0.00	5.46	−inf	0.01
AQP2	ENSBTAG00000008374	chr5:30094769-30103142	46.69	5.40	3.11	0.02
SUBH2BV	ENSBTAG00000012775	chr20:8803855-8804333	369.16	104.00	1.83	0.02
SHD	ENSBTAG00000039696	chr7:21002692-21010652	87.36	261.12	−1.58	0.02
EFNA1	ENSBTAG00000020244	chr3:15521493-15528212	107.24	17.86	2.59	0.02
SESN3	ENSBTAG00000004034	chr15:15502202-15530863	28.12	9.61	1.55	0.02
RPS27	ENSBTAG00000013866	chr3:16505731-16507247	1323.87	512.66	1.37	0.02
U1	ENSBTAG00000048270	chr18:14877691-14877845	329.93	0.00	inf	0.03
ENSBTAG00000038351	ENSBTAG00000038351	chr5:441592-442266	0.00	158.13	−inf	0.03
ENSBTAG00000046783	ENSBTAG00000046783	chr2:94202054-94202354	21.49	0.00	inf	0.03
PCDH8	ENSBTAG00000018179	chr12:10938750-10943209	2.09	11.90	−2.51	0.04
RPL30	ENSBTAG00000040051	chr14:68467562-68470952	474.20	226.19	1.07	0.04
RBMX	ENSBTAG00000047652	chr15:46287139-46288701	4.61	18.43	−2.00	0.04
PRDM13	ENSBTAG00000006095	chr9:50873668-50881408	1.84	11.47	−2.64	0.04
FZD1	ENSBTAG00000002107	chr4:8490675-8494829	1.41	6.80	−2.27	0.05
GTSF1L	ENSBTAG00000011233	chr13:72984199-72985097	318.49	142.18	1.16	0.05
MLPH	ENSBTAG00000000634	chr3:117584072-117637812	1.62	7.29	−2.17	0.05
SLAIN2	ENSBTAG00000021963	chr6:68695297-68741224	16.18	40.23	−1.31	0.05
RBM46	ENSBTAG00000001696	chr17:2523017-2567620	4.14	16.58	−2.00	0.05

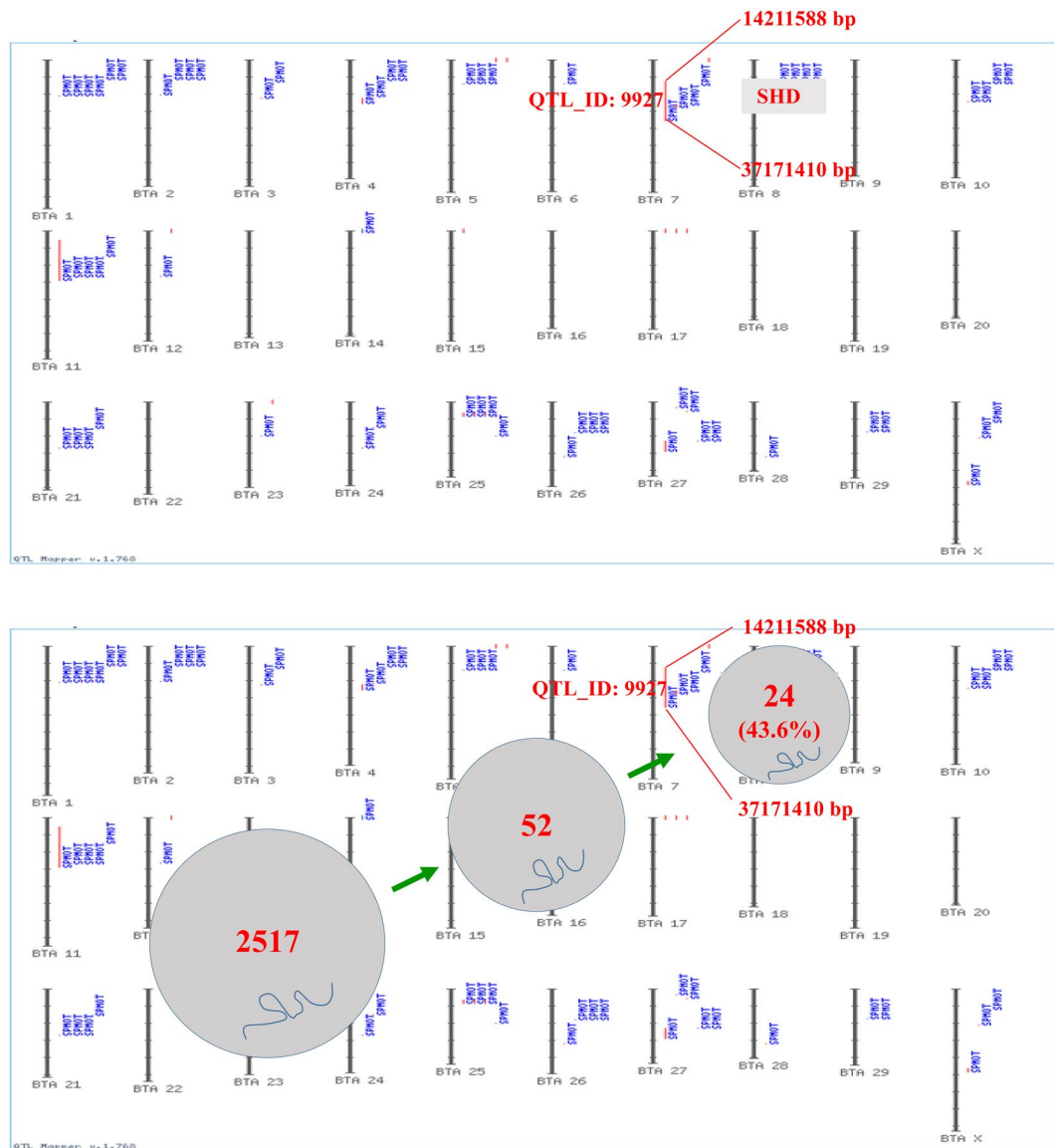
**Table 1.** The 19 significantly differentially expressed genes. Note: inf means positive infinity; −inf means negative infinity.

lncRNAs	locus	H(FPKM)	L(FPKM)	log <sub>2</sub> (fold change)	FDR value
TCONS_00000538	chr1:75937027-75937410	0.00	80.73	−inf	0.01
TCONS_00003123	chr10:55232037-55232318	0.00	153.91	−inf	0.01
TCONS_00005494	chr11:62030429-62030804	0.00	45.96	−inf	0.01
TCONS_00007217	chr11:97395420-97396011	0.00	7.90	−inf	0.01
TCONS_00008747	chr12:87789106-87792905	20.13	0.00	inf	0.01
TCONS_00009226	chr13:46843141-46843647	0.00	21.71	−inf	0.01
TCONS_00011237	chr14:6662538-6663034	0.00	18.56	−inf	0.01
TCONS_00012048	chr14:6722290-6722813	0.00	20.09	−inf	0.01
TCONS_00012181	chr14:23115327-23115826	0.00	13.76	−inf	0.01
TCONS_00012775	chr15:19466827-19467304	0.00	19.09	−inf	0.01
TCONS_00013706	chr15:85036155-85036516	0.00	14.22	−inf	0.01
TCONS_00014744	chr16:2115585-2116115	15.22	0.00	inf	0.01
TCONS_00014816	chr16:7166876-7167359	0.00	16.80	−inf	0.01
TCONS_00015123	chr16:43687434-43687736	0.00	32.20	−inf	0.01
TCONS_00019053	chr18:41851514-41851878	0.00	35.39	−inf	0.01
TCONS_00020130	chr18:16232385-16233355	10.72	0.00	inf	0.01
TCONS_00021031	chr18:64429371-64429857	24.37	0.00	inf	0.01
TCONS_00021657	chr19:34327860-34328423	0.00	16.26	−inf	0.01
TCONS_00024090	chr2:18693367-18693630	0.00	31.45	−inf	0.01
TCONS_00025506	chr2:24451811-24452412	0.00	15.76	−inf	0.01

**Table 2.** The top 20 significantly differentially expressed lncRNAs. Note: inf means positive infinity; −inf means negative infinity.

analysis indicated that the gene expression patterns validated through qRT-PCR are consistent with the results of RNA-Seq high-throughput sequencing (Supplementary Fig. S3; Supplementary Tables S5–S6).

**Chromosomal mapping of mRNAs and lncRNAs.** First, we analyzed the chromosomal distribution of 19 differentially expressed coding genes and 2,517 differentially expressed lncRNAs. Statistical results showed that differentially expressed coding genes were mostly located on chromosomes 3, 5 and 15 and differentially expressed lncRNAs were mostly located on chromosomes 1, 2 and 6 (Supplementary Fig. S4). Then, we mapped differentially expressed mRNAs and lncRNAs to cattle quantitative trait loci (QTLs) (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). We found that differentially expressed coding gene SHD could be mapped onto



**Figure 4.** Mapping of differentially expressed mRNAs and lncRNAs on QTLs related to sperm motility.

a QTL (QTL\_ID: 9927) related to sperm motility on chromosomes 7. Moreover, 52 out of 2,517 lncRNAs were located on cattle QTLs related to sperm motility, among which 46.2% (24/52) lncRNAs were mapped to the QTL (QTL\_ID: 9927) on chromosome 7 (Fig. 4, Supplementary Table S7).

## Discussion

With the development of accurate and cost-effective high-throughput genomic analysis technology, many critical genes responsible for species characteristics and regulatory pathways that underlie biological processes have been unraveled<sup>25</sup>. Herein, we constructed the expression profiles of mRNAs and lncRNAs in the H and L groups and comprehensively analyzed and identified the candidate key genes and functional pathways that are related to sperm motility. The basic profiles of the genes revealed notable differences between the H and L groups.

As a result, we detected 19 differentially expressed mRNAs and 2,517 differentially expressed lncRNAs between the H and L groups. We also identified 5 differentially expressed genes including RBMX, AQP2, EFNA1, MLPH and RPL30, that are involved in the “extracellular exosome” GO term. In boars, seminal plasma exosomes prolong the effective motility time of sperm by directly binding to the membrane of the sperm head; this action could improve sperm plasma membrane integrity<sup>26</sup>. Parts of differentially expressed genes are also reported being involved in spermatogenesis, sperm motility, and fertilization. For example, previous studies have revealed that RBMX which acted as the role of potential splicing factors could regulate spermatogenesis<sup>27</sup>; AQP2 is expressed in seminiferous tubules of the testis, localized to Leydig cells and elongated and round spermatids, which may be pivotal for establishment of male fertility<sup>28,29</sup>. We also found that differentially expressed gene SHD and 52 lncRNAs are located on cattle QTLs related to sperm motility. Moreover, approximately 46.2% (24/52) of lncRNAs



are mapped to a QTL (QTL\_ID: 9927) on chromosome 7. The colocation of these lncRNAs indicates that these differentially expressed genes may influence sperm motility through their complex interactions.

Interestingly, our finding showed that the number of differentially expressed lncRNAs is much greater than that of mRNAs in mature sperm of bull. Increasing studies have shown that specific lncRNAs expressed in mature sperm and testis may be related to sperm motility. Recent research revealed that there was an association between lncRNA expression and sperm motility by analyzing the lncRNA expression profiles in human mature sperm<sup>30</sup>. Previous study has also identified several spermatogenic lncRNA that are differentially expressed in testis of sheep which were associated with sperm motility<sup>31</sup>. Analogously, compared with mRNAs, the number of differentially expressed lncRNAs is much greater in sperm by exposure to cadmium<sup>32</sup>. Therefore, the above data suggest that lncRNA may be as a promising candidate marker for assessing bull sperm motility. Continued studies should be carried out to explore other utilities of bull sperm lncRNAs, such as fertility prediction and assisted reproductive technologies<sup>33</sup>.

A major function of lncRNAs is the regulation of the expression of neighboring protein-coding genes via transcriptional coactivation/repression<sup>34–36</sup>. Therefore, we searched and identified differentially expressed mRNAs that are located within the 100 kb upstream and downstream regions of differentially expressed lncRNAs as cis target genes. We found that the identified differentially expressed lncRNA TCONS\_00041733 may regulate differentially expressed coding gene EFNA1 (*ephrin A1*). EFNA1 is a cell surface protein and is involved in vascular development<sup>37</sup>, inflammatory response and cell proliferation and angiogenesis<sup>38</sup> as well as sperm morphology and membrane functionality<sup>39</sup>. EFNA1 was higher expressed in sperm, seminal vesicle fluid and seminal plasma of bulls<sup>39–41</sup>. Higher levels of EFNA1 may possible result in an attack of immune cells over the sperm and therefore damage the cell, but the precise role of EFNA1 in male reproductive physiology needs to be confirmed<sup>39</sup>. Taken together, the EFNA1 gene can be as a candidate gene affecting the sperm motility based on our data. Therefore, further study should be carried out in the next experiment.

In this study, we aimed to provide a global expression profile of mRNAs and lncRNAs in bull semen. However, as a limitation in the data analysis, specially these expressed genes with the low value of FPKM were not excluded in the process of gene differential expression analysis, which may cause false positive results. Therefore, these lowly expressed mRNAs and lncRNAs should be carefully verified in future experiments. The other limitation was that only a small number of lncRNAs and targets were highlighted due to the limited available annotation and unknown functions of gene.

## Conclusion

We used the RNA-Seq method to reveal for the first time that the expression patterns of lncRNAs and mRNAs in semen from bulls with high sperm motility are different from those of lncRNAs and mRNAs in semen from bulls with low sperm motility. We identified differentially expressed mRNAs and lncRNAs that may potentially influence bull sperm motility. Additionally, aberrantly expressed lncRNAs could regulate the expression of protein-coding genes and pathways that may contribute to bull sperm motility. Although our findings provide a valuable foundation for further research on the role of lncRNAs in sperm motility, further studies are also required to fully elucidate the detailed molecular mechanisms that underlie the function and contribution of lncRNAs to the characteristics of sperm quality.

## Materials and Methods

**Ethics statement.** All experiments involving animals were conducted in accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals published by the Ministry of Science and Technology of China in 2004 ([http://www.most.gov.cn/fggw/zfwj/zfwj2006/200609/t20060930\\_54389.htm](http://www.most.gov.cn/fggw/zfwj/zfwj2006/200609/t20060930_54389.htm)). Our studies were approved by the Animal Care and Use Committee of the Dairy Cattle Research Center, Shandong Academy of Agricultural Sciences (Shandong, China).

**Animals and sample collection.** Three paired full-sibling Holstein bulls (4.3–5.8 years old) produced through embryo transfer were selected from a bull station (Jinan City, Shandong Province, China). Each pair of bulls had approximately equal genomic estimated breeding values (GISU) ([http://idele.fr/fileadmin/medias/Documents/INTRO\\_NO\\_133\\_va.pdf](http://idele.fr/fileadmin/medias/Documents/INTRO_NO_133_va.pdf)), which were calculated by the Animal Genetics Department of INRA (National Institute for Agronomic Research) and genomic evaluation data were provided by Genes Diffusion (Douai Cedex, France). However, their sperm motility phenotypes, as determined with a sperm analysis system (AndroVision, Minitube, Germany), were different ( $p < 0.05$ , Fig. 1). According to the average sperm motility from the production record from 2011 to 2015, three bulls were allocated to the high sperm motility group (H: H1, H2, and H3), and the three remaining bulls were allocated to the low sperm motility group (L: L1, L2, and L3). Prior to the sampling of semen for RNA-seq, the collected semen were assessed three times in a row to confirm the grouping. The sperm motility of the H and L groups were  $68.67\% \pm 1.15\%$  and  $60.00\% \pm 1.63\%$  (mean  $\pm$  SE, Fig. 1), respectively. Fresh semen were collected individually from each bull and immediately frozen in liquid nitrogen for storage until RNA extraction.

**RNA extraction, library construction, and sequencing.** Total RNA was extracted from each fresh semen sample by using Animal Tissue RNA Purification Kit TRK-1002 (LC Sciences, Houston, Texas) in accordance with the manufacturer's instructions. RNA quality was evaluated using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and Nano Drop ND-2000 Spectrophotometer (Nano-Drop, Wilmington, DE, USA). The value of RNA integrity number exceeded 7.0, and samples with  $OD_{260/280}$  values of more than 1.8 were used for the following steps.

Total RNA was randomly fragmented into small pieces and purified into fragments with lengths of 200–600 bp. Cleaved RNA fragments were subsequently reverse-transcribed to create the final RNA-seq library, which

was then sequenced by using the Illumina HiSeq™ 4000 platform with paired-end sequencing (2 × 150 bp read length) following the vendor's recommended protocol. Raw sequencing data in the fastq format were submitted to the NCBI SRA database (<http://www.ncbi.nlm.nih.gov/Traces/sra>), with accession number SRP158901.

**Transcriptome assembly.** Raw image data from sequencing machines were transformed through base calling into sequence reads, which were designated as raw reads. The obtained sequence raw reads in FASTAQ file format were first checked by FastQC software. Then, the raw reads were filtered through following process: (1) Reads with adapter contamination were first eliminated. (2) Reads with unknown bases larger than 5% were discarded. (3) Reads with more than 20% of bases with a quality score of  $Q \leq 10$  were removed. After the above steps, clean reads were produced from raw reads for subsequent analysis.

All clean reads were mapped to the cow reference genome database ([ftp://ftp.ensembl.org/pub/release-86/fasta/bos\\_taurus/dna/](ftp://ftp.ensembl.org/pub/release-86/fasta/bos_taurus/dna/)) by Tophat2 software (v2.0.9) with a mismatch tolerance of three nts<sup>42</sup>. To construct the transcriptome, the mapped reads of each sample were assembled using Cufflinks (v2.1.1) package<sup>43</sup>. Then, Cuffmerge program was used to merge all transcriptomes from six samples to reconstruct a comprehensive transcriptome.

**Identification of lncRNAs.** Assembled transcripts were annotated using the Cuffcompare program in the Cufflinks package<sup>44</sup>. Transcripts that overlap with known mRNAs were first excluded. Then, the coding potential of remaining transcripts with length  $\geq 200$  bp, mapped read coverage  $\geq 3$ , and exon number  $\geq 1$  was calculated by Coding Potential Calculator (CPC) and Coding-Non-Coding-Index (CNCI). CPC assesses the protein-coding potential of transcripts on the basis of six biologically meaningful sequence features by using a classifier based on the support vector machine<sup>45</sup>. CNCI effectively classifies protein-coding or noncoding transcripts by profiling adjoining nucleotide triplets independent of known annotations<sup>46</sup>. If the predicted results were CPC score  $< -1$  and CNCI score  $< 0$ , the transcript was considered as a candidate lncRNA.

**Differential expression analysis.** To eliminate the influence of different transcript lengths and sequencing levels, the expression levels of mRNAs and lncRNAs in each sample were calculated as fragments per kilobase of exon per million mapped fragments (FPKM) by using the Cuffdiff program in Cufflinks package<sup>44</sup>. The formula used was  $FPKM = (106 C)/(NL/103)$ , where C is the number of fragments aligned with unigene exons, N is the number of total fragments aligned with all unigenes, and L is the length of exons in base pairs. Here, the levels of differential mRNA and lncRNA expression were analyzed through Cuffdiff on the basis of negative binomial distribution. The statistical significance of the differences in expression levels between two groups of semen was analyzed using Student's *t*-test.

**Target gene prediction.** Increasing evidence has indicated that lncRNAs may affect the expression of neighboring coding genes by playing a cis regulating role<sup>47,48</sup>. To explore the potential function of lncRNAs, differentially expressed lncRNAs were selected for target gene prediction. To reduce the rate of false positives, we did not analyze the trans role of lncRNAs. We only selected differentially expressed mRNAs as potential target genes of lncRNAs. Moreover, we searched mRNAs within the 100 kb upstream and downstream regions of differentially expressed lncRNAs as the cis target genes.

**Real-time PCR validation.** Real-time PCR (qRT-PCR) was performed on an LightCycler 480II Real-Time PCR System using SYBR® Green PCR Master Mix (TaKaRa, Dalian, China) with specific primers. Gene expression levels were normalized to beta-actin gene expression through the  $2^{-\Delta\Delta C_t}$  method. Each reaction was performed with a 25  $\mu$ L reaction mixture that contained 12.5  $\mu$ L of 2 × SYBR Green Real-time PCR Master Mix, 1  $\mu$ L of each primer, 2  $\mu$ L of cDNA, and 8.5  $\mu$ L of H<sub>2</sub>O. PCR cycling conditions were as follows: an initial single cycle at 95 °C for 3 min followed by 40 cycles at 95 °C for 15 s, 57 °C for 15 s, and 72 °C for 20 s. Three independent biological replicates were used for each reaction.

## References

- Walsh, S. W., Williams, E. J. & Evans, A. C. O. A review of the causes of poor fertility in high milk producing dairy cows. *Anim Reprod Sci* **123**, 127–138 (2011).
- Guzick, D. S. *et al.* Sperm morphology, motility, and concentration in fertile and infertile men. *New Engl J Med* **345**, 1388–1393 (2001).
- Picard-Hagen, N. *et al.* Effect of precocious collection on semen output and quality in young Holstein bulls. *Theriogenology* **57**, 1511–1522 (2002).
- Druet, T. *et al.* Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. *J Anim Breed Genet* **126**, 269–277 (2009).
- Peñagaricano, F., Weigel, K. A. & Khatib, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Anim Genet* **43**(Supp1), 65–71 (2012).
- Hering, D. M., Olenski, K. & Kaminski, S. Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. *Anim Reprod Sci* **146**, 89–97 (2014).
- Suchocki, T. & Szyda, J. Genome-wide association study for semen production traits in Holstein-Friesian bulls. *J Dairy Sci* **98**, 5774–5780 (2015).
- Qin, C. *et al.* Genome-wide association study for semen traits of the bulls in Chinese Holstein. *Anim Genet* **48**, 80–84 (2016).
- Taylor, J. F., Schnabel, R. D. & Sutovsky, P. Identification of genomic variants causing sperm abnormalities and reduced male fertility. *Anim Reprod Sci* **194**, 57–62 (2018).
- Peddinti, D. *et al.* Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Syst Biol* **2**, 19 (2008).
- Selvaraju, S. *et al.* Occurrence and functional significance of the transcriptome in bovine (*Bos taurus*) spermatozoa. *Sci Rep* **7**, 42392 (2017).
- Card, C. J., Krieger, K. E., Kaproth, M. & Sartini, B. L. Oligo-dT selected spermatozoal transcript profiles differ among higher and lower fertility dairy sires. *Anim Reprod Sci* **177**, 105–123 (2017).

13. Liu, W. M. *et al.* Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc Natl Acad Sci USA* **109**, 490–494 (2012).
14. Gapp, K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* **17**, 667–669 (2014).
15. Sendler, E. *et al.* Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res* **41**, 4104–4117 (2013).
16. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145–166 (2012).
17. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
18. Vance, K. W. & Ponting, C. P. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* **30**, 348–355 (2014).
19. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol* **10**, 924–933 (2013).
20. Arun, G., Akhade, V. S., Donakonda, S. & Rao, M. R. mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol Cell Biol* **32**, 3140–3152 (2012).
21. Ni, M. J. *et al.* Identification and characterization of a novel non-coding RNA involved in sperm maturation. *PLoS One* **6**, e26053 (2011).
22. Anguera, M. C. *et al.* Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet* **7**, e1002248 (2011).
23. Zhang, L., Lu, H., Xin, D., Cheng, H. & Zhou, R. A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. *Biochem Biophys Res Commun* **400**, 696–700 (2010).
24. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
25. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol Cell* **58**, 586–597 (2015).
26. Du, J. *et al.* Boar seminal plasma exosomes maintain sperm function by infiltrating into the sperm membrane. *Oncotarget* **7**, 58832 (2016).
27. Delbridge, M. L., Lingenfelter, P. A., Distech, C. M. & Graves, J. M. The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nat Genet* **22**, 223–224 (1999).
28. Nelson, R. D. *et al.* Expression of an AQP2 Cre recombinase transgene in kidney and male reproductive system of transgenic mice. *AM J Physiol-Cell Physiol* **275**, C216–C226 (1998).
29. Klein, C., Troedsson, M. H. T. & Rutllant, J. Region-specific expression of aquaporin subtypes in equine testis, epididymis, and ductus deferens. *The Anatomical Record* **296**, 1115–1126 (2013).
30. Zhang, X. *et al.* Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm. *Biol Reprod*, <https://doi.org/10.1093/biolre/iy253> (2018).
31. Zhang, Y. *et al.* Long noncoding RNA expression profile changes associated with dietary energy in the sheep testis during sexual maturation. *Sci Rep* **7**, 5180 (2017).
32. Gao, F. *et al.* Dysregulation of long noncoding RNAs in mouse testes and spermatozoa after exposure to cadmium. *Biochem Biophys Res Commun* **484**, 8–14 (2017).
33. Selvaraju, S. *et al.* Current status of sperm functional genomics and its diagnostic potential of fertility in bovine (*Bos taurus*). *Syst Biol Reprod Med* **64**, 484–501 (2018).
34. Davila, M. P. *et al.* Mitochondrial ATP is required for the maintenance of membrane integrity in stallion spermatozoa, whereas motility requires both glycolysis and oxidative phosphorylation. *Reproduction* **152**, 683–694 (2016).
35. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
36. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
37. McBride, J. L. & Ruiz, J. C. Ephrin-A1 is expressed at sites of vascular development in the mouse. *Mech Dev* **77**, 201–204 (1998).
38. Coulthard, M. G. *et al.* Eph/ephrin signaling in injury and inflammation. *Am J Pathol* **181**, 1493–1503 (2012).
39. Rego, J. P. *et al.* Proteomic analysis of seminal plasma and sperm cells and their associations with semen freezability in Guzerat bulls. *J Anim Sci* **94**, 5308–5320 (2016).
40. Westfalewicz, B. *et al.* Analysis of bull (*Bos taurus*) seminal vesicle fluid proteome in relation to seminal plasma proteome. *J Dairy Sci* **100**, 2282–2298 (2017).
41. Kasimanickam, R. K., Kasimanickam, V. R., Arangasamy, A. & Kastelic, J. P. Sperm and seminal plasma proteomics of high-versus low-fertility Holstein bulls. *Theriogenology* **126**, 41–48 (2019).
42. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomics in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
43. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578 (2012).
44. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat methods* **5**, 621 (2008).
45. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* **35**, W345–W349 (2007).
46. Sun, L. *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* **41**, e166–e166 (2013).
47. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155–159 (2009).
48. Guil, S. & Esteller, M. Cis-acting noncoding RNAs: Friends and foes. *Nat Struct Mol Biol* **19**, 1068–1075 (2012).

## Acknowledgements

This study was supported by the Shandong Provincial Science Foundation for Distinguished Young Scholars of China (JQ201709), the National Natural Science Foundation of China (31872984, 31500998, 31672397, 31771374), the Major Project of National Transgene in China (2018ZX08007001-002), the Shandong Provincial Key Research and Development Program of China (2017GNC10120), and the Agricultural Scientific and Technological Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2016A04, CXGC2018E14).

## Author Contributions

J.H. conceived and designed the experiments. X.W., C.Y., F.G., Y.L., H.Z., and H.Z. performed experiments and analyzed data. Z.J., Y.Z., Q.J., X.Z., J.W., Y.S., C.W. and H.B.Z. collected samples and prepared materials. X.W. and J.H. wrote the manuscript. All the authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-38462-x>.

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



**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