Check for updates

OPEN

Optimization of sperm RNA processing for developmental research

Won-Ki Pang, Saehan Kang, Do-Yeal Ryu, Md Saidur Rahman, Yoo-Jin Park & Myung-Geol Pang✉

Recent studies have demonstrated the significance of sperm RNA function as a transporter of important information directing the course of life. To determine the message contained in sperm RNA, it is necessary to optimize transcriptomic research tools. The current study was performed to optimize the processing of sperm RNA from sample storage to quantitative real-time PCR and assess the corresponding method with to evaluate male fertility and its representative markers, equatorin (EQTN) and peroxiredoxin (PRDX). Following successive steps of the Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines, several options were compared using boar spermatozoa. To evaluate the optimized procedures, the relationship between mRNA expression of EQTN and PRDX in spermatozoa and the fertility (litter size) of 20 individual boars was assessed. Unexpectedly, DNase treatment during RNA isolation had the deleterious effect by decreasing the RNA concentration by 56% and eliminating the correlation between EQTN and PRDX4 mRNA expression and male fertility. Moreover, when sperm RNA was processed using the corresponding method, the results showed the highest exon sequence expression, male fertility prediction power, and consistency. This optimized protocol for predicting male fertility can be used to study the transport of messages directing the life course from spermatozoon to offspring.

A spermatozoon is a haploid gamete that conveys paternal genetic information to offspring. During the differentiation to spermatozoon, the transcriptional activity of the cell shuts down[1]. Although the condensed nucleus of spermatozoa is thought to be in a transcriptionally inert state, sperm RNA was identified in many mammalian species, such as human[2], bovine[3], porcine[4], equine[5], and rodent[6]. Several plausible hypothesis and experimentation/observations on sperm RNA function, *i.e.* contribution during fertilization by delivering novel paternal RNA to the oocyte[7–9] and signaling in mitochondrial translation[10], are widely known. Moreover, the possibility of using sperm RNA as a marker for diagnosis of male fertility and infertility has recently gained attention[11–13]. Recent studies suggested that many phenotypes from parent generation are transferred to the progeny through sperm RNA[14,15]. Additionally, sperm RNAs were suggested to relay the message to one's life course, providing the opportunity to study health at key stages of life from preconception to offspring[16]. However, despite abundant evidences on the functional significance of sperm RNAs, the contribution on sperm biology remains controversial[17,18] and efforts are needed to optimize transcriptomic research tools. This may lead to improvements in the understanding of the underlying mechanisms of how spermatozoon affects the course of life.

The main transcriptomic research tools currently being used are RNA sequencing (RNA-seq) and reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR)[19,20]. To facilitate sperm RNA research, these protocols require some modifications because of the complexity of sperm biology, including the lower RNA levels compared to other cell types[21] and disulfide bond-rich protamine-packed chromatin structure that resists lysis during RNA isolation[22]. Moreover, most RNAs except for a portion of the nucleus exists in a highly fragmented state in spermatozoa[23]. Thus, isolation of intact sperm RNA is important and has been widely studied[6,22,24]. As a result, various transcriptomic studies of spermatozoa related to sperm function, male fertility, idiopathic infertility, sperm-oocyte fusion, and inheritance of the paternal phenotype have been conducted[16,25,26].

Despite the advantages of RNA-seq, such as its high resolution on total aspects of biological sample, it has several limitations including low specificity, sensitivity, and reproducibility when used for the prognosis and diagnosis of specific phenotypes[19]. Compare to RNA-seq, RT-qPCR is useful for screening specific target genes

Department of Animal Science and Technology and BET Research Institute, Chung-Ang University, Anseong, Gyeonggi-do 17546, Republic of Korea. ✉email: mgpang@cau.ac.kr
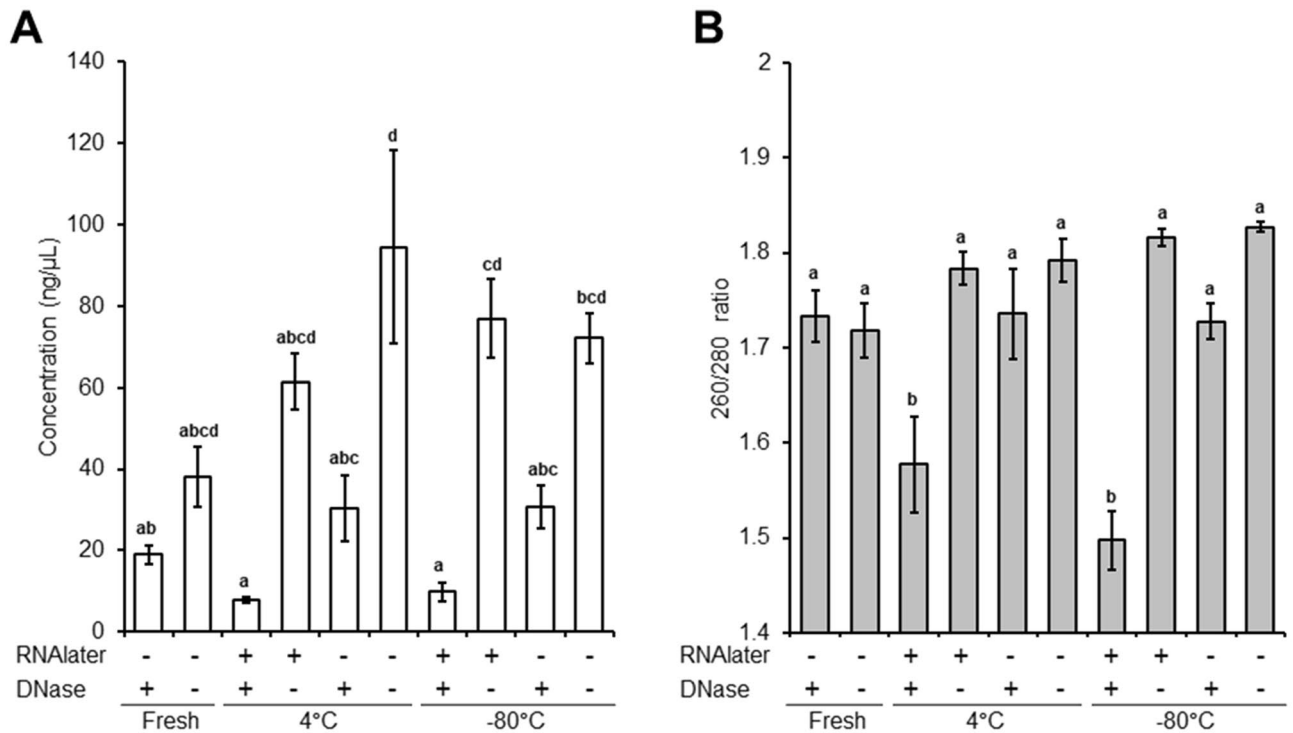
Figure 1. Effect of storage conditions and DNase treatment on isolated RNA quantity and quality. (**A**) RNA quantity is shown as the concentration (ng/μL) of isolated RNA in each treatment. (**B**) RNA quality is shown as 260/280 ratio of RNA. a–d $P < 0.05$ tested by one-way analysis of variance. Post-hoc test was Duncan in RNA quantity and Tukey's method in RNA quality. All data are expressed as the mean ± SE.

for predicting prognosis and diagnosing disease, phenotypes, and other biological aspects in various samples[27]. According to Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines, the use of RNA stabilizing solution and DNase is essential[28]. Many researchers perform DNase treatment to increase the RNA ratio in the sample. Additionally, an RNA-stabilizing solution (*e.g.* RNAlater) is widely used to stabilize and protect cellular RNA because of its immediate RNase inactivation characteristics[29]. However, it is not examined that whether these treatments could be used to improve RT-qPCR results in spermatozoa or not.

Compare to proteomic markers in sperm[30,31], only a few studies have reported that sperm mRNA expression is correlated with male fertility[32–34]. One of the reasons is the protocols for accurate prognosis and diagnosis of male fertility have not been optimized. Therefore, we selected several steps that may affect sperm RNA and RT-qPCR data according to MIQE guidelines[28]. Sperm sample storage, RNA isolation, cDNA synthesis, and RT-qPCR conditions were compared to maximize the power of predicting male fertility using mRNA markers. We used a porcine model to test these conditions because of its large available amount of fertility data from artificial insemination (AI)[35] and well-established genetic traits related to fertility, particularly in large white pigs[36]. The correlation between mRNA expression in spermatozoa and the phenotype related with the life course from spermatozoon through offspring was evaluated using RNA markers related to litter size, equatorin (EQTN)[37,38] and peroxiredoxin (PRDX)[39,40] after improving sperm RNA preparation and RT-qPCR. The EQTN and PRDX4 are related to sperm acrosome function[38,40] and ample evidence suggests the levels of EQTN and PRDX4 protein in spermatozoa are correlated with male fertility[37,39]. Therefore, in this study, we try to establish whether transcript levels of an EQTN/PRDX4 can be used as proxies for the corresponding protein levels.

Results

Effect of sperm processing condition on isolated sperm RNA.
To test the effect of storage conditions on isolated sperm RNA, washed spermatozoa were snap-frozen in liquid nitrogen (LN$_2$) or treated with RNAlater and stored in 4 °C or−80 °C for 7 days. Subsequently, sperm RNAs were isolated with DNase treated or non-treated protocol. Neither RNAlater nor DNase treatment alone affected RNA quantity (Fig. 1A) and quality (260/280 ratio, Fig. 1B). However, when RNA was isolated with DNase from RNAlater-treated samples, the 260/280 ratio was significantly decreased compared to the other groups (Fig. 1B) and its concentration was lower compared to isolated RNA without DNase (Fig. 1A). Furthermore, the average RNA concentration of 20 individual sperm samples was 56% lower when treated with DNase (44.28 ± 8.12 ng/μL) during isolation compared to samples not treated with DNase (101.7 ± 13.5 ng/μL, Table 1). Despite the deleterious effect of DNase treatment on the concentration of RNA, the average quality was same in both RNA isolated with (1.79 ± 0.03) or without DNase (1.80 ± 0.01, Table 1).

Sample name	Litter size	Isolated RNA without DNase		Isolated RNA with DNase	
		Concentration (ng/µL)	260/280 ratio	Concentration (ng/µL)	260/280 ratio
1	14.15 ± 0.79	41.25	1.74	67.72	1.86
2	13.44 ± 0.68	153.97	1.86	35.80	1.79
3	13.30 ± 0.95	75.89	1.74	74.12	1.93
4	13.15 ± 0.79	140.45	1.87	123.00	1.85
5	13.14 ± 0.55	146.45	1.87	58.04	1.86
6	13.10 ± 0.99	88.29	1.83	17.00	1.54
7	12.90 ± 1.06	152.45	1.85	28.12	1.81
8	12.82 ± 0.80	25.81	1.70	129.56	1.93
9	12.79 ± 0.94	109.57	1.79	96.12	1.91
10	12.75 ± 1.28	197.57	1.82	21.40	1.89
11	12.67 ± 0.72	255.65	1.83	13.40	1.61
12	12.65 ± 0.64	29.33	1.77	21.08	1.84
13	12.54 ± 0.75	58.93	1.83	42.84	1.84
14	12.40 ± 1.72	93.89	1.82	33.64	1.88
15	12.38 ± 0.85	33.89	1.58	10.12	1.80
16	12.25 ± 0.77	83.89	1.82	33.96	1.84
17	12.08 ± 0.51	108.29	1.80	37.32	1.77
18	11.81 ± 0.71	54.45	1.80	18.28	1.55
19	11.75 ± 1.16	128.37	1.85	13.64	1.55
20	11.35 ± 0.92	56.37	1.78	10.36	1.72
Average	12.67 ± 0.14	101.7 ± 13.5*	1.80 ± 0.01	44.28 ± 8.12	1.79 ± 0.03

Table 1. Isolated RNA from randomly selected 20 boar spermatozoa. The male fertility data (total litter size) and RNA quantity and quality of isolated RNA with or without DNase in 20 randomly selected boar spermatozoa. *$P < 0.05$, calculated using two-tailed Student's t-test.

Whole sperm RNA processing method on RT-qPCR outcome. To compare each sperm RNA processing method, cDNA was synthesized by using oligo dT or random hexamer primers and RT-qPCR was performed with exon, intron, and junction primers of the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene. The cDNA synthesis efficiency for oligo dT and random hexamer showed no significant difference (Supplementary Fig. S1 online). No amplification was observed in both oligo dT and random hexamers no template cDNA (Supplementary Fig. S2 online). The average Cq value difference of EQTN, PRDX4, and GAPDH between no reverse transcriptase and positive control was $7.0 ± 0.4$, $11.0 ± 0.3$, and $10.6 ± 0.3$, respectively (Supplementary Fig. S3A online). The average Cq value difference of EQTN, PRDX4, and GAPDH between no reverse transcriptase and cDNA from sperm RNA was $4.5 ± 0.4$, $4.2 ± 0.3$, and $6.4 ± 1.3$, respectively (Supplementary Fig. S3B online). The RT-qPCR data of isolated RNA with DNase from −80 °C stored after RNAlater treatment was excluded from further statistical analysis because its melting temperature differed in melting curve analysis and showed a decreased amplification peak compared to the other samples (Fig. 2A). The average Cq value of exon primers between oligo dT and random hexamers showed a significant difference (Fig. 2B). The GAPDH RT-qPCR data showed that the three methods (1. isolated RNA from fresh spermatozoa with DNase. 2. isolated RNA from 4 °C stored spermatozoa without RNAlater and DNase treatment. 3. isolated RNA from −80 °C stored spermatozoa without RNAlater and DNase treatment.) in oligo dT cDNA and only one method (isolated RNA from 4 °C stored spermatozoa without RNAlater and with DNase treatment) in random hexamer cDNA showed a lower Cq value in the exon primer compared to using both intron and junction primers (Fig. 2C,D). The sizes of the PCR products were 110, 112, and 194 base pairs in the exon, intron, and junction, respectively (Fig. 2E and Supplementary Fig. S4 online).

Correlation between RT-qPCR data and male fertility. To evaluate the correlation of RT-qPCR data for each genes and male fertility, the relative expression of genes in 20 individual boar spermatozoa were examined using optimized procedures. According to cDNA synthesis and RT-qPCR data for the GAPDH primer, the oligo dT primer in cDNA synthesis and exon primer of the gene in RT-qPCR were adequate for measuring the expression of mRNA. To rule out the effects of genomic DNA (gDNA), intron primers for each marker were also checked with RT-qPCR. Total cDNAs of 20 individual boar spermatozoa were prepared and the mRNA expression of the EQTN and PRDX4 genes was evaluated. EQTN and PRDX4 expression levels were significantly correlated with litter size when RNA was isolated without DNase and when the Cq value from exon primer RT-qPCR was considered as mRNA expression (rEQTN = −0.341 and rPRDX4 = 0.321; Fig. 3A,B). When the Cq value of the intron primer was considered to rule out the gDNA factor, only EQTN was significantly correlated with litter size (rEQTN = −0.374 and rPRDX = 0.267; Fig. 3C,D). In contrast to the data obtained from isolated RNAs without DNase, the relative expression of both EQTN and PRDX4 in DNase-treated RNA showed no correlation with litter size (rEQTN = 0.013 and rPRDX = 0.124; Fig. 3E,F). Additionally, the Cq value from the
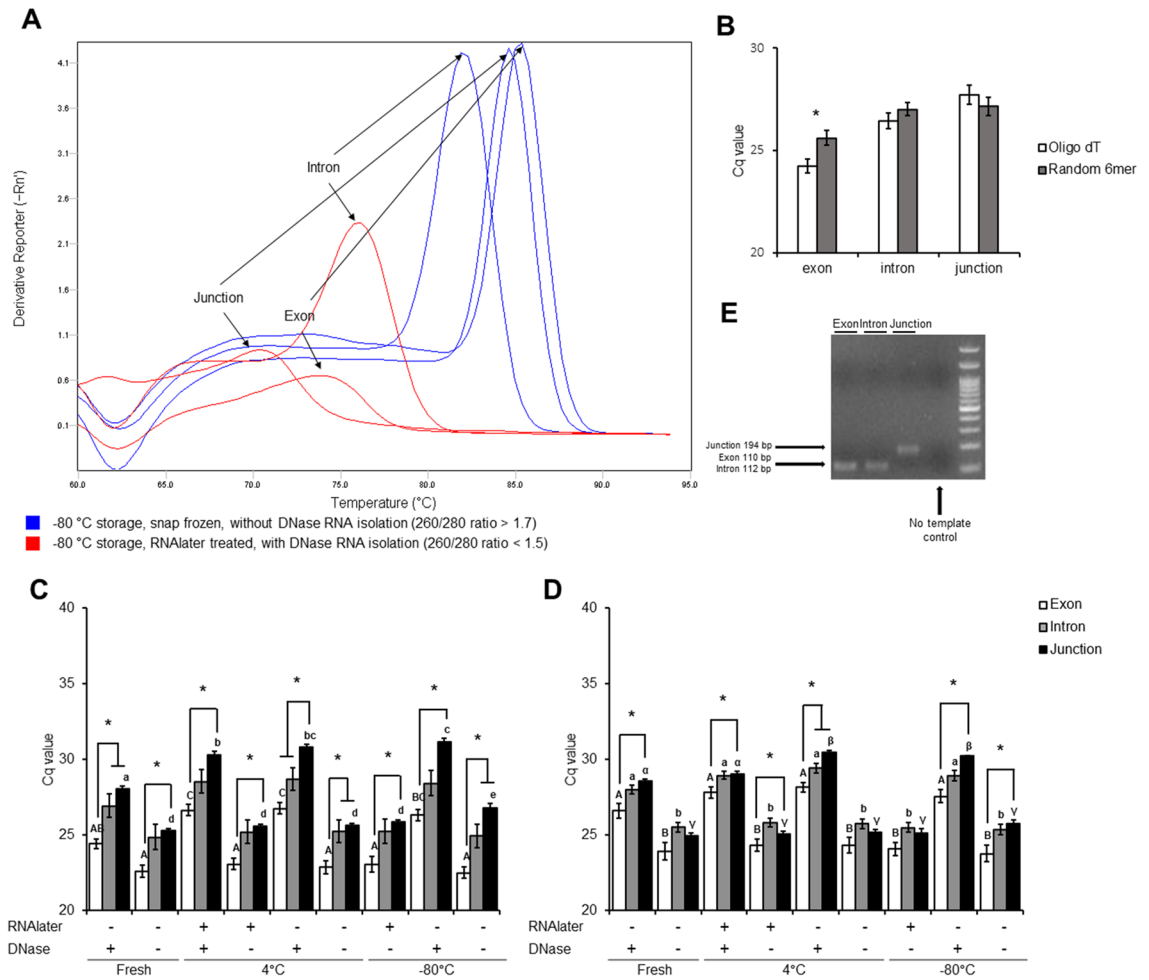
Figure 2. Effect of total sperm RNA processing method on RT-qPCR. (**A**) Melting curve of −80 °C storage, snap-frozen, without DNase RNA isolation as a representative of normal RT-qPCR data from good quality RNA (260/280 ratio > 1.7) and −80 °C storage, RNAlater-treated, with DNase RNA isolation as a representative of unnormal RT-qPCR data from bad quality RNA (260/280 ratio < 1.5). (**B**) Average Cq value of GAPDH exon, intron, and junction primers in synthesized cDNA with oligo dT and random hexamer primers. *$P < 0.05$ tested by Student's t-test (**C**) Cq value of GAPDH exon, intron, and junction primers in synthesized cDNA with oligo dT in each method. *$P < 0.05$ tested by one-way analysis of variance among Cq value of each primers. A-C, a-d $P < 0.05$ tested by one-way analysis of variance among treatments (**D**) Cq value of GAPDH exon, intron, and junction primers in synthesized cDNA with random hexamer in each method. *$P < 0.05$ tested by one-way analysis of variance among Cq value of each primers. A–B, a–b, α–γ $P < 0.05$ tested by one-way analysis of variance among treatments (**E**) Representative gel electrophoresis image of RT-qPCR product in −80 °C storage, snap-frozen, without DNase RNA isolation method with no template control. This image was cropped from Supplementary Figure S1A online. (**B**–**D**) All data are expressed as the mean ± SE.

intron primer of DNase-treated RNA could not be considered to rule out the gDNA factor because there was no difference from the Cq value of the exon primer.

Assessment of male fertility predicting power of mRNA markers.

To assess the male fertility predicting power of mRNA markers when processed with optimized protocols, ROC curve analysis was performed. The overall accuracies of EQTN and PRDX4 in non-DNase-treated samples were the same (60%; Fig. 4A,B). When the gDNA factor was considered in EQTN relative expression, all predictive values were decreased (sensitivity, specificity, negative predictive value, positive predictive value, and overall accuracy were 59.52%, 61.11%, 39.29%, 78.13%, and 60.00% before gDNA consideration and 48.48%, 50.00%, 26.09%, 72.73%, and 48.89% after gDNA consideration, respectively; Fig. 4A). Interestingly, although the cDNA templates were from same samples, the coefficient of variation and quartile deviation showed different patterns according to DNase treatment and gDNA factor consideration (Fig. 4C).
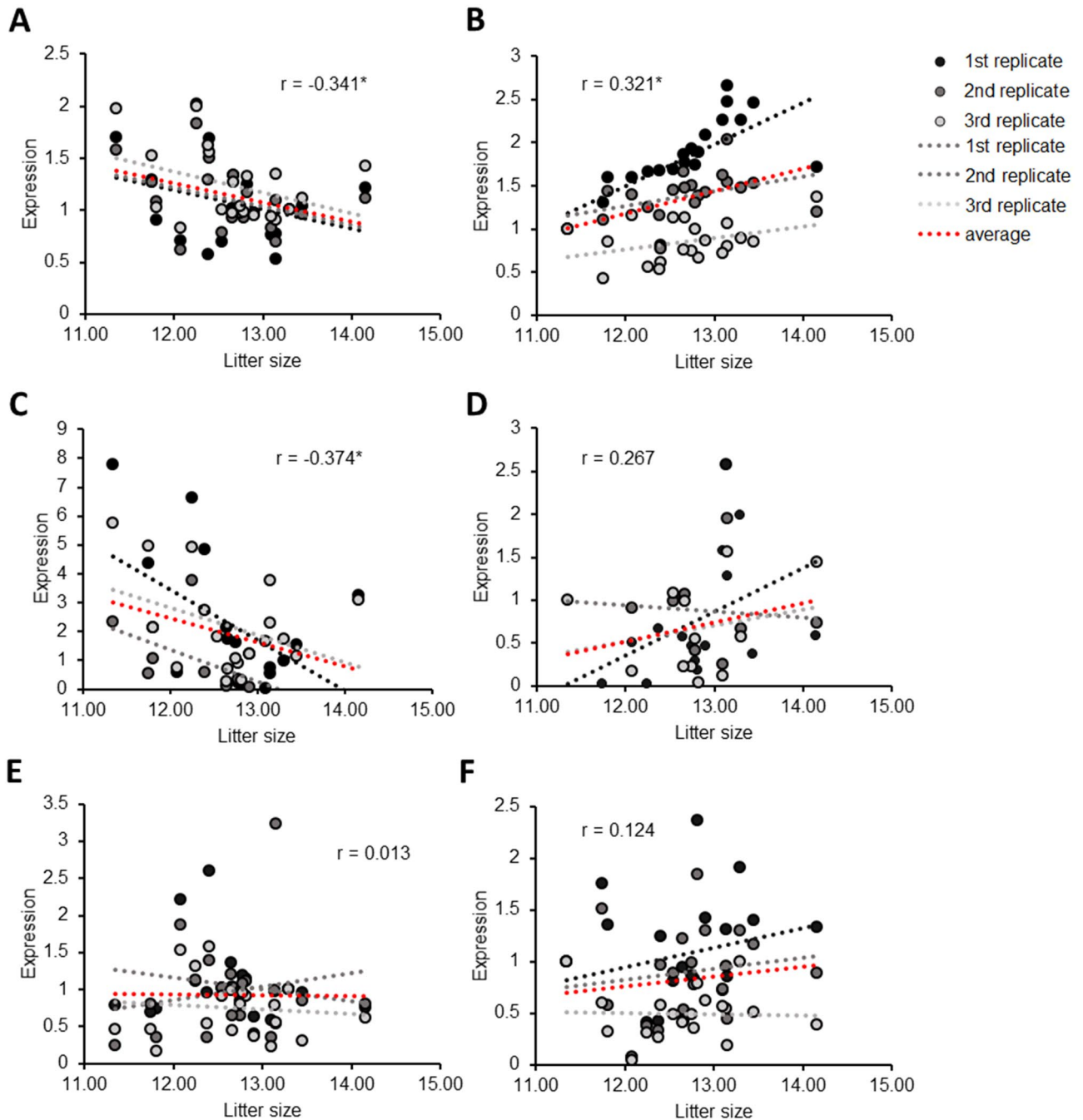
Figure 3. Correlation between EQTN and PRDX4 expression with male fertility. (**A**) linear regression test of EQTN mRNA expression and litter size in isolated RNA without DNase in 20 randomly selected boar spermatozoa. (**B**) Linear regression test of PRDX4 mRNA expression and litter size in isolated RNA without DNase in randomly selected 20 boar spermatozoa. (**C**) Linear regression test of EQTN mRNA expression when gDNA considered with intron primer expression and litter size in isolated RNA without DNase in 20 randomly selected boar spermatozoa. (**D**) Linear regression test of PRDX4 mRNA expression when gDNA was considered with intron primer expression and litter size in isolated RNA without DNase in 20 randomly selected boar spermatozoa. (**E**) Linear regression test of EQTN mRNA expression and litter size in isolated RNA with DNase in 20 randomly selected boar spermatozoa. (**F**) Linear regression test of PRDX4 mRNA expression and litter size in isolated RNA with DNase in 20 randomly selected boar spermatozoa. r, Pearson correlation coefficient; *$P < 0.05$, calculated using linear regression test.
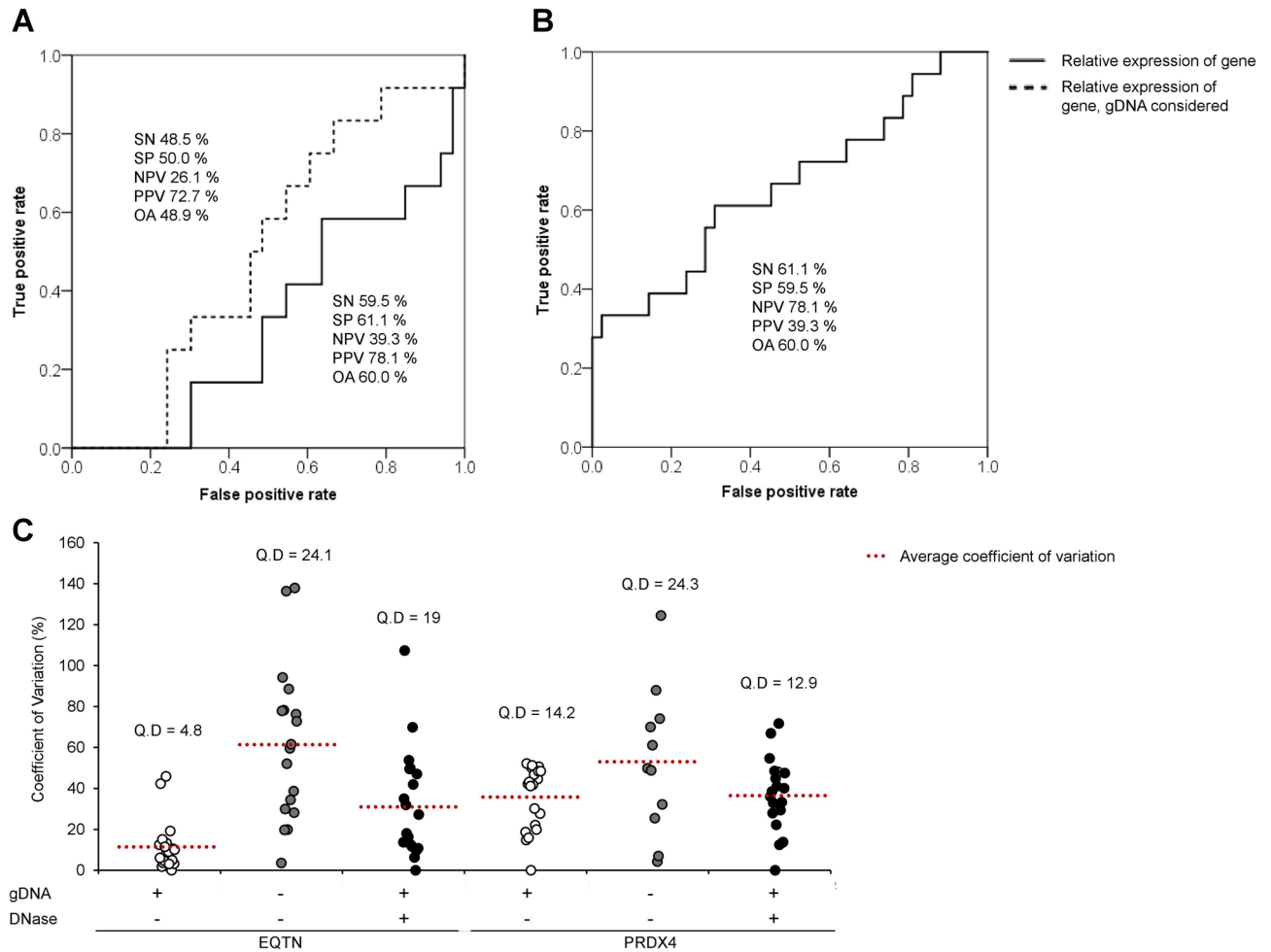
Figure 4. ROC curve of EQTN and PRDX4 expression comparing litter size. (**A**) ROC curve for EQTN mRNA expression when gDNA was considered or not and litter size in isolated RNA without DNase in 20 randomly selected boar spermatozoa. (**B**) ROC curve for PRDX4 mRNA expression and litter size in isolated RNA without DNase in 20 randomly selected boar spermatozoa. (**C**) The coefficient of variation (%) and quartile deviation (Q.D) of EQTN and PRDX4 mRNA expression in each processing method. The cut-off values of relative expression in all treatments were determined based on fixed litter size (13) for comparison among each treatment. Sensitivity (SN) is the percentage of boars showing true-positive results when tested with mRNA expression. Specificity (SP) is the percentage of boars showing true-negative results. The positive predictive value (PPV) is the percentage of boars that tested as positive and simultaneously has a true-positive litter size. The negative predictive value (NPV) is the percentage of boars that tested as negative or simultaneously had a true-negative litter size. OA, Overall accuracy.

Discussion

Despite the critical contribution of sperm RNA to male fertility[26,41], embryo development[42], epigenetic inheritance for acquired traits included in the paternal genome[43], and health[44], the lack of optimization of transcriptomic research tools prevents a thorough understanding of sperm biology. Therefore, we optimized the processing of sperm RNA from sample storage to RT-qPCR and evaluated the male fertility predicting power of the optimized protocol.

The significance of the porcine model on biomedical research studying a spectrum of human diseases, including obesity, arthritis, cardiovascular disease, and skin and eye conditions are already well known[45,46]. In addition, Several studies emphasize the particular importance of the porcine model in xenotransplantation[47]. In the same vein, the porcine model can also be useful in the study of human male infertility. What makes this possible is the similarity between porcine genome and the human genome and well-organized breeding data from AI[35,48]. Furthermore, many researh groups are conducting comprehensive trascriptomic analysis in boar spermatozoa[49–51]. In the present study, a porcine model was used to evaluate the fertility prediction of sperm RNA.

First, we compared RNAlater treatment to the snap-freezing method in LN_2 for sample storage step. Many recent studies reported that the effect of RNAlater on isolated RNA depends on the cell type, storage period, and research application[52–55]. To the best of our knowledge, this is the first study to compare the use of RNAlater and snap-freezing for spermatozoa. RNAlater treatment was time-consuming and costly, without beneficially affecting sperm RNA. Thus, treatment with RNAlater was not used in further steps.

For RNA isolation of spermatozoa, a two-step lysis was used. Spermatozoa chromatin is consists of the enriched disulfide bonds, which can only be broken down by detergents (Sodium dodecyl dulfate or sodium lauryl sulfate) and reducing agents (e.g., dithiothreitol or β-mercaptoethanol). To successfully isolate the large number of sperm RNAs in sperm nuclei[21,22], the sperm head should be fully lysed. Trizol, which contains neither detergents nor reducing agents, cannot lysis sperm head. For the separation of sperm RNA from a cell lysate, chloroform was used. Chloroform is a common and key reagent used for isolation of sperm RNA, both by Trizol-modified methods but also for commercially available RNA extraction kits during phase separation[5,56].

Generally, DNase is used to rule out gDNA contamination in isolated RNA. However, DNase has harmful effect not only in gDNA but also in concentrations of RNA and small RNAs[57,58]. In this study, DNase affected RNA concentration and quality in RNAlater-treated samples. Although the DNase alone had no significant effect on RNA quality during the optimization procedure, the exon Cq value of DNase-treated samples was similar with the intron value. Based on these studies, we suggested that DNase may induce the degradation of small RNA in spermatozoa during RNA isolation. This could lead to biased interpretation of sperm transcriptomic study. Therefore, later researchers need to reconsider the use of DNase during sperm RNA study.

For the cDNA synthesis step, the effects of oligo dT and random hexamer primers were compared. The size of the PCR product with the junction primer for GAPDH was 194 base pairs, suggesting that most amplicons were from gDNA. Because of this, the Cq value of the exon primer should be significantly lower than that of the intron primer and junction primer for detecting mRNA. In case of fresh sperm sample, DNase treated RNA showed this corresponding expression pattern. The oligo dT primer cDNA, from 4 to − 80 °C snap-frozen storage and without DNase-treated isolated sperm RNA samples also showed this corresponding expression pattern. Thus, snap-frozen samples in LN$_2$, with or without DNase RNA isolation and Oligo dT cDNA synthesis, were selected as the optimized sperm processing conditions for predicting male fertility.

To assess the male fertility predicting power of mRNA markers using the optimized protocols, a correlation test was conducted to evaluate the mRNA expression of EQTN and PRDX4 could represent the male fertility data or not. EQTN and PRDX4 exon expression levels were negatively and positively correlated with litter size in isolated RNA without DNase, respectively. In contrast, no correlation between the exon expression of EQTN and PRDX4 from isolated RNA with DNA was observed. As previously noticed, in the isolated RNA with DNase treatment, the expression of the exon sequence was similar or lower than the expression of the intron sequence. In the other word, the portion of RNA in the sample was too low to indicate the correlation between gene expression and male fertility. Laurell et al. described a method for ruling out the gDNA factor in cDNA samples using an optimized gDNA-specific ValidPrime assay and gDNA reference sample[59]. Similarly, we directly evaluated the expression of gene intron sequences and attempted to exclude gDNA contamination. Only EQTN was correlated with male fertility when the gDNA factor was considered. To clarify the effect of gDNA factor consideration in male fertility prediction value the cut-off value of EQTN relative expression was settled at fixed litter size (13). The decreased male fertility prediction value in EQTN expression was a major limitation when considering the gDNA factor. Moreover, the losing correlation on PRDX4 and decrease in predictive value of EQTN was corresponded with coefficient of variation and quartile deviation.

In summary, we optimized the processing of sperm RNA for screening male fertility from sample storage to RT-qPCR. The optimized condition involved snap-freezing in LN$_2$, without DNase RNA isolation, oligo dT cDNA synthesis, and RT-qPCR with exon primers. The highest exon sequence expression, male fertility prediction power, and consistency were obtained using this method. The mRNA level obtained using the optimized method showed better male fertility predicting power and consistency of data. This study will shed light on medical and industrial use of sperm RNA to evaluate infertility, animal health, and offspring phenotypes. Moreover, the optimized protocol can be used to elucidate the underlying mechanisms of life course research from spermatozoa to the next generation.

Materials and methods

All procedures involving animals were approved by the Institutional Animal Care and Use Committee of Chung-Ang University (Approval No. 2017-00018) and performed according to the corresponding guidelines. All methods were performed in accordance with the relevant guidelines and regulations.

Sperm sample preparation. Yorkshire boar semen was provided by Sunjin Co. grand-grand parents farm (Danyang, Korea). The sexually mature (11–23 months) and body weight over 90 kg boars were used to collect semen sample. For the laboratory work, the semen samples were collected once in summer season (June–July). The ejaculates of boar were obtained by gloved-hand technique[60] and collected semen were diluted to a density of 30×10^6 sperm cells/mL in 100 mL of Beltsville thawing solution[37,61]. During sample transport, semen was maintained at 17 °C in an ice box 61. After arrival in the laboratory, each semen sample was washed with discontinuous 35 and 70% Percoll (Sigma-Aldrich Co, St. Louis, MO, USA) gradient to eliminate non-sperm contents from the semen[62]. Thereafter, sperm pellets were snap-frozen in LN$_2$ (− 196 °C) or incubated in RNAlater (12×10^7 sperm cells/mL; Invitrogen, Carlsbad, CA, USA) overnight in 4 °C. Both samples were stored in 4 °C or− 80 °C for 7 days before RNA isolation. To minimize the effect of individual variation, semen samples from three boars were combined and used for method optimization with GAPDH gene. Moreover, to determine whether the optimized method can be used for accurately predicting fertility phenotype of boar spermatozoa, the mRNA expression level of semen from 20 different boars were evaluated.

Computer-assisted sperm analysis. Sperm motility and motion kinematics of Percoll separated spermatozoa was checked by Computer-assisted sperm analysis system (SAIS-Plus ver.10.1; Medical Supply, Seoul, Korea). Percoll separated spermatozoa was resuspended in 1 mL of mTCM199 media and 10 μL of resuspended

Primer name	Upper primer (5′–3′)	Tm (°C)	Lower primer (5′–3′)	Tm (°C)	Amplicon size (bp)
GAPDH exon	AAG AGC ACG CGA GGA GGA G	67.1	GGG GTC TGG GAT GGA AAC T	65.1	109
GAPDH Intron	TTC AAG CCC CAG CCA GAT T	66.9	CCG GAA ACA ACC CAA GAC C	66.5	111
GAPDH junction	TCC TGG GCT ACA CTG AGG AC	64.3	CTT GAC GAA GTG GTC GTT GA	64	193
EQTN exon	AAA CCC TGC AAA TGA AGA CAA C	64	CTG CCA AAA TGA TGA CAA AAA G	63.2	106
EQTN intron	GCA GAA CCC CAG TCT CTG TC	63.9	GGG CTC CTT ATC CAA AAT GG	64.2	95
PRDX4 exon	GTG TCC AAC TGA AAT TAT CG	57.1	AGA TGG GTA AAC TGT GAA TC	55.4	101
PRDX4 intron	CGG GCA GAC AAC TCT TAA CAT	63	TGC ACC TTC GAT GAA CTA GC	63	94

Table 2. Designed primers for RT-qPCR. All primers were designed genes from pig reference genome Sscrofa11.1 Primary Assembly. Gene accession number: GAPDH (ENSSSCG00000000694); EQTN (ENSSSCG00000005121); PRDX4 (ENSSSCG00000012171).

sample was placed on pre-heated (37 °C) Makler counting chamber (Sefi Medical Instruments, Haifa, Israel). Under a phase contrast microscope, the motility, hyperactivation, curvilinear velocity, straight line velocity, average path velocity, linearity, beat cross frequency, wobble, and mean amplitude of head lateral displacement of spermatozoa were measured (Supplementary Table S1 online)[32].

RNA isolation and cDNA synthesis. Sperm RNAs were isolated by using a PureLink RNA Mini Kit (Invitrogen) according to the manufacturer's protocol with some modifications. Briefly, spermatozoa pellets were resuspended in PBS and counted under microscope with a Makler counting chamber to adjust the concentration to 40–50×10 6 cells/mL. Pellets (40–50×10 6 cells) were suspended in non-toxic guanidine-isothiocyanate lysis buffer in PureLink RNA Mini Kit containing β-mercaptoethanol (40 μL/mL; Sigma-Aldrich Co) and homogenized. TRIzol (Invitrogen) and chloroform were added and centrifuged at 12,000× g . The upper layer was transferred into a fresh tube and mixed with pure ethanol (supernatant : pure ethanol=1:1, v/v; Sigma-Aldrich Co). Sperm RNA was attached to a spin cartridge. For DNase-treated RNA, the spin cartridge membrane was incubated in PureLink DNase mixture (80 μL; Invitrogen) for 15 min at room temperature during washing. Both isolated RNAs with and without DNase were immersed in 20 μL of nuclease-free water. The quality (260/280 ratio) and quantity of sperm RNA were checked with an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA). After matching of total RNA concentration among samples (total 400 ng), cDNA was synthesized with the PrimeScript 1st strand cDNA Synthesis Kit (Takara Bio, Inc., Shiga, Japan) according to the manufacturer's protocol. Oligo dT and random hexamer primers were compared for cDNA synthesis.

RT-qPCR. The cDNA was mixed with SYBR Green PCR master mix (Applied Biosystems, Foster city, CA, USA) and amplified on a 7,500 fast real-time PCR system (Applied Biosystems). As a positive control and negative control, cDNA from porcine testis RNA and no template and reverse transcriptase cDNA was used, respectively. All primers were designed based on Reference genome Sscrofa11.1 Primary Assembly. For method optimization, we compared the exon, intron, and exon-exon junction primers of GAPDH, a reference gene for RT-qPCR in boar spermatozoa (Table 2)[63]. The EQTN and PRDX4 genes were selected as target markers of male fertility[37,39]. Exon and intron primers of selected markers were designed; and for reference gene, GAPDH was used (Table 2). The amplification efficiency of primers was tested with calibration curve[64]. Briefly, the cDNA was diluted into 1, 2.5, 5, 10, and 25X (250, 100, 50, 25, and 10 ng) in each well and Cq value of studied genes were analyzed with amount of cDNA (Supplementary Fig. S5 online). The total reaction volume was 20 μL (100 ng of cDNA). The cycling parameters were 95 °C (10 min) followed by 40 cycles at 95 °C (15 s) and 60 °C (60 s). After cycling, continuous melt curve analysis was performed at 95 °C (15 s), 60 °C (60 s), and a progressive increase up to 95 °C (0.5 °C/30 s). Data were analyzed with the 7,500 Software v2.3 (Thermo Fisher Scientific, Waltham, MA, USA). Relative expression of each gene was calculated with delta delta Cq method and to consider the gDNA factor with intron expression, a calculation from Laurell et al. was used[59].

$$Cq\ mRNA\ =\ -\log_2(2^{-Cqexon}-2^{-Cqintron})$$

The size of the PCR products was checked with running gel electrophoresis on 2% agarose gel for 30 min. The gel image was gathered with a PowerShot A640 (Canon, Tokyo, Japan) under ultraviolet light. The image brightness was controlled with PSRemote v1.6.4 software (Breeze Systems, Camberley, UK).

Fertility data acquisition. Fertility data from randomly selected 20 Yorkshire boars were gathered by AI using their semen. Semen samples were collected after sexual maturation and diluted as described above. AI was performed by professional technicians from Sunjin Co. twice per estrus[32,65,66]. The inseminated sows were managed at 20±5 °C with air circulation, 2:1 light/dark cycle, and adequate feed for pregnant sow during the experi-

mental period. AI was performed for 1 year to avoid the seasonal fluctuation and boar semen was inseminated to average 17.25 ± 1.98 sows (total 360 trial). The average number of born piglets from AI results using boar semen was considered as the fertility outcome.

Assessment of male fertility predicting power of mRNA. The correlation between fertility and mRNA expression of EQTN and PRDX4 from 20 individual boar semen sample were analyzed with linear regression in all treatments. To estimate the male fertility predicting power of mRNA markers, receiver operating characteristic (ROC) curve analysis was utilized to set the cut-off value and assess the four key parameters *i.e.* sensitivity, specificity, positive predictive value, and negative predictive value of each marker[67]. The cut-off values of relative expression in all treatments were decided based on a fixed litter size (13) to compare each treatment. Sensitivity is the percentage of boars that were true-positive when tested for mRNA expression. Specificity is the percentage of boars that tested as true-negative. The positive predictive value is the percentage of boars that tested as positive and had a true-positive litter size. The negative predictive value is the percentage of boars that tested as negative or had a true-negative litter size.

Statistics. The Student's *t*-test, one-way analysis of variance, Pearson correlation test, linear regression test, and ROC curve test were performed with SPSS v1.8 software (SPSS, Inc., Chicago, IL, USA). All data were subjected to a normality test (Shapiro–Wilk test) and homogeneity of variances test (Levene's test). $P < 0.05$ was considered to indicate significantly different results. All data are expressed as the mean ± SE.

References

1. Dadoune, J. P., Siffroi, J. P. & Alfonsi, M. F. Transcription in haploid male germ cells. *Int. Rev. Cytol.* **237**, 1–56. https://doi.org/10.1016/S0074-7696(04)37001-4 (2004).
2. Dadoune, J. P., Pawlak, A., Alfonsi, M. F. & Siffroi, J. P. Identification of transcripts by macroarrays, RT-PCR and in situ hybridization in human ejaculate spermatozoa. *Mol. Hum. Reprod.* **11**, 133–140. https://doi.org/10.1093/molehr/gah137 (2005).
3. Gilbert, I., Bissonnette, N., Boissonneault, G., Vallee, M. & Robert, C. A molecular analysis of the population of mRNA in bovine spermatozoa. *Reproduction* **133**, 1073–1086. https://doi.org/10.1530/REP-06-0292 (2007).
4. Yang, C. C. *et al.* Identification and sequencing of remnant messenger RNAs found in domestic swine (Sus scrofa) fresh ejaculated spermatozoa. *Anim. Reprod. Sci.* **113**, 143–155. https://doi.org/10.1016/j.anireprosci.2008.08.012 (2009).
5. Das, P. J. *et al.* Total RNA isolation from stallion sperm and testis biopsies. *Theriogenology* **74**, 1099–1106. https://doi.org/10.1016/j.theriogenology.2010.04.023 (2010).
6. Schuster, A. *et al.* SpermBase: a database for sperm-borne RNA contents. *Biol. Reprod.* **95**, 99. https://doi.org/10.1095/biolreprod.116.142190 (2016).
7. Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P. & Krawetz, S. A. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429**, 154. https://doi.org/10.1038/429154a (2004).
8. Hayashi, S., Yang, J., Christenson, L., Yanagimachi, R. & Hecht, N. B. Mouse preimplantation embryos developed from oocytes injected with round spermatids or spermatozoa have similar but distinct patterns of early messenger RNA expression. *Biol. Reprod.* **69**, 1170–1176. https://doi.org/10.1095/biolreprod.103.016832 (2003).
9. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474. https://doi.org/10.1038/nature04674 (2006).
10. Gur, Y. & Breitbart, H. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. *Genes Dev.* **20**, 411–416. https://doi.org/10.1101/gad.367606 (2006).
11. Lambard, S. *et al.* Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. *Mol. Hum. Reprod.* **10**, 535–541. https://doi.org/10.1093/molehr/gah064 (2004).
12. Ostermeier, G. C., Dix, D. J., Miller, D., Khatri, P. & Krawetz, S. A. Spermatozoal RNA profiles of normal fertile men. *Lancet* **360**, 772–777. https://doi.org/10.1016/S0140-6736(02)09899-9 (2002).
13. Wang, H. *et al.* A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. *J. Mol. Med. (Berl.)* **82**, 317–324. https://doi.org/10.1007/s00109-004-0526-3 (2004).
14. Sarker, G. *et al.* Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. *Proc Natl Acad Sci USA* **116**, 10547–10556. https://doi.org/10.1073/pnas.1820810116 (2019).
15. Gapp, K. *et al.* Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. *Mol. Psychiatry* https://doi.org/10.1038/s41380-018-0271-6 (2018).
16. Godia, M., Swanson, G. & Krawetz, S. A. A history of why fathers' RNA matters. *Biol. Reprod.* **99**, 147–159. https://doi.org/10.1093/biolre/ioy007 (2018).
17. Miller, D., Ostermeier, G. C. & Krawetz, S. A. The controversy, potential and roles of spermatozoal RNA. *Trends Mol. Med.* **11**, 156–163. https://doi.org/10.1016/j.molmed.2005.02.006 (2005).
18. Dadoune, J. P. Spermatozoal RNAs: what about their functions?. *Microsc. Res. Tech.* **72**, 536–551. https://doi.org/10.1002/jemt.20697 (2009).
19. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet* **10**, 57–63. https://doi.org/10.1038/nrg2484 (2009).
20. Freeman, W. M., Walker, S. J. & Vrana, K. E. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**(112–122), 124–125. https://doi.org/10.2144/99261rv01 (1999).
21. Jodar, M. *et al.* The presence, role and clinical use of spermatozoal RNAs. *Hum. Reprod. Update* **19**, 604–624. https://doi.org/10.1093/humupd/dmt031 (2013).
22. Goodrich, R. J., Anton, E. & Krawetz, S. A. Isolating mRNA and small noncoding RNAs from human sperm. *Methods Mol. Biol.* **927**, 385–396. https://doi.org/10.1007/978-1-62703-038-0_33 (2013).
23. Zhao, Y. *et al.* Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. *Hum. Reprod.* **21**, 1583–1590. https://doi.org/10.1093/humrep/del027 (2006).
24. Georgiadis, A. P. *et al.* High quality RNA in semen and sperm: isolation, analysis and potential application in clinical testing. *J. Urol.* **193**, 352–359. https://doi.org/10.1016/j.juro.2014.07.107 (2015).

25. Sharma, U. & Rando, O. J. Father-son chats: inheriting stress through sperm RNA. *Cell Metab* **19**, 894–895. https://doi.org/10.1016/j.cmet.2014.05.015 (2014).
26. Krawetz, S. A. Paternal contribution: new insights and future challenges. *Nat. Rev. Genet.* **6**, 633–642. https://doi.org/10.1038/nrg1654 (2005).
27. Bustin, S. A. & Mueller, R. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin. Sci. (Lond.)* **109**, 365–379. https://doi.org/10.1042/CS20050086 (2005).
28. Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622. https://doi.org/10.1373/clinchem.2008.112797 (2009).
29. Bachoon, D. S., Chen, F. & Hodson, R. E. RNA recovery and detection of mRNA by RT-PCR from preserved prokaryotic samples. *FEMS Microbiol. Lett.* **201**, 127–132. https://doi.org/10.1111/j.1574-6968.2001.tb10745.x (2001).
30. Park, Y. J., Kim, J., You, Y. A. & Pang, M. G. Proteomic revolution to improve tools for evaluating male fertility in animals. *J. Proteome Res.* **12**, 4738–4747. https://doi.org/10.1021/pr400639x (2013).
31. Kwon, W. S., Rahman, M. S. & Pang, M. G. Diagnosis and prognosis of male infertility in mammal: the focusing of tyrosine phosphorylation and phosphotyrosine proteins. *J. Proteome Res.* **13**, 4505–4517. https://doi.org/10.1021/pr500524p (2014).
32. Kim, K. U. *et al.* Sperm solute carrier family 9 regulator 1 is correlated with boar fertility. *Theriogenology* **126**, 254–260. https://doi.org/10.1016/j.theriogenology.2018.12.023 (2019).
33. Bansal, S. K., Gupta, N., Sankhwar, S. N. & Rajender, S. Differential genes expression between fertile and infertile spermatozoa revealed by transcriptome analysis. *PLoS ONE* **10**, e0127007. https://doi.org/10.1371/journal.pone.0127007 (2015).
34. Alvarez-Rodriguez, M. *et al.* The transcriptome of pig spermatozoa, and its role in fertility. *Int J Mol Sci* https://doi.org/10.3390/ijms21051572 (2020).
35. Broekhuijse, M. L., Feitsma, H. & Gadella, B. M. Artificial insemination in pigs: predicting male fertility. *Vet. Q.* **32**, 151–157. https://doi.org/10.1080/01652176.2012.735126 (2012).
36. Roehe, R. & Kennedy, B. W. Estimation of genetic parameters for litter size in Canadian Yorkshire and Landrace swine with each parity of farrowing treated as a different trait. *J. Anim. Sci.* **73**, 2959–2970. https://doi.org/10.2527/1995.73102959x (1995).
37. Kwon, W. S. *et al.* Discovery of predictive biomarkers for litter size in boar spermatozoa. *Mol. Cell Proteomics* **14**, 1230–1240. https://doi.org/10.1074/mcp.M114.045369 (2015).
38. Ito, C. *et al.* Deletion of Eqtn in mice reduces male fertility and sperm-egg adhesion. *Reproduction* **156**, 579–590. https://doi.org/10.1530/REP-18-0394 (2018).
39. Ryu, D. Y. *et al.* Peroxiredoxin activity is a major landmark of male fertility. *Sci. Rep.* **7**, 17174. https://doi.org/10.1038/s41598-017-17488-7 (2017).
40. Sasagawa, I. *et al.* Possible involvement of the membrane-bound form of peroxiredoxin 4 in acrosome formation during spermiogenesis of rats. *Eur. J. Biochem.* **268**, 3053–3061. https://doi.org/10.1046/j.1432-1327.2001.02200.x (2001).
41. Jodar, M. *et al.* Absence of sperm RNA elements correlates with idiopathic male infertility. *Sci. Transl. Med.* **7**, 295re296. https://doi.org/10.1126/scitranslmed.aab1287 (2015).
42. Johnson, G. D. *et al.* The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction* **141**, 21–36. https://doi.org/10.1530/REP-10-0322 (2011).
43. Chen, Q., Yan, W. & Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat. Rev. Genet.* **17**, 733–743. https://doi.org/10.1038/nrg.2016.106 (2016).
44. Burl, R. B., Clough, S., Sendler, E., Estill, M. & Krawetz, S. A. Sperm RNA elements as markers of health. *Syst. Biol. Reprod. Med.* **64**, 25–38. https://doi.org/10.1080/19396368.2017.1393583 (2018).
45. Tumbleson, M. E. & Schook, L. B. Advances in swine in biomedical research. *Advances in Swine in Biomedical Research, Vols 1 and 2*, 1–4 (1996).
46. Schook, L. B. *et al.* Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. *Comput. Funct. Genom.* **6**, 251–255. https://doi.org/10.1002/cfg.479 (2005).
47. Fishman, J. A. & Patience, C. Xenotransplantation: infectious risk revisited. *Am. J. Transpl.* **4**, 1383–1390. https://doi.org/10.1111/j.1600-6143.2004.00542.x (2004).
48. Humphray, S. J. *et al.* A high utility integrated map of the pig genome. *Genome Biol.* **8**, R139. https://doi.org/10.1186/gb-2007-8-7-r139 (2007).
49. Yang, C. C. *et al.* Seasonal effect on sperm messenger RNA profile of domestic swine (Sus Scrofa). *Anim. Reprod. Sci.* **119**, 76–84. https://doi.org/10.1016/j.anireprosci.2009.12.002 (2010).
50. Li, Y. *et al.* High throughput small RNA and transcriptome sequencing reveal capacitation-related microRNAs and mRNA in boar sperm. *BMC Genom.* **19**, 736. https://doi.org/10.1186/s12864-018-5132-9 (2018).
51. Fraser, L. *et al.* Transcriptome analysis of boar spermatozoa with different freezability using RNA-Seq. *Theriogenology* **142**, 400–413. https://doi.org/10.1016/j.theriogenology.2019.11.001 (2020).
52. Pedersen, K. B., Williams, A., Watt, J. & Ronis, M. J. Improved method for isolating high-quality RNA from mouse bone with RNAlater at room temperature. *Bone Rep.* **11**, 100211. https://doi.org/10.1016/j.bonr.2019.100211 (2019).
53. Passow, C. N. *et al.* Nonrandom RNAseq gene expression associated with RNAlater and flash freezing storage methods. *Mol. Ecol. Resour.* **19**, 456–464. https://doi.org/10.1111/1755-0998.12965 (2019).
54. Wang, M., Ji, X., Wang, B., Li, Q. & Zhou, J. Simultaneous evaluation of the preservative effect of RNAlater on different tissues by biomolecular and histological analysis. *Biopreserv. Biobank.* **16**, 426–433. https://doi.org/10.1089/bio.2018.0055 (2018).
55. Zhu, Y. *et al.* Assessment of RNAlater((R)) as a potential method to preserve bovine muscle proteins compared with dry ice in a proteomic study. *Foods* https://doi.org/10.3390/foods8020060 (2019).
56. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159. https://doi.org/10.1006/abio.1987.9999 (1987).
57. Norhazlin, J. *et al.* Effect of DNase treatment on RNA extraction from preimplantation murine embryos. *Genet Mol. Res.* **14**, 10172–10184. https://doi.org/10.4238/2015.August.28.1 (2015).
58. Huang, Z., Fasco, M. J. & Kaminsky, L. S. Optimization of Dnase I removal of contaminating DNA from RNA for use in quantitative RNA-PCR. *Biotechniques* **20**, 1012–1014. https://doi.org/10.2144/96206st02 (1996).
59. Laurell, H. *et al.* Correction of RT-qPCR data for genomic DNA-derived signals with ValidPrime. *Nucl. Acids Res* **40**, e51. https://doi.org/10.1093/nar/gkr1259 (2012).
60. Almond, G. *The Swine AI Book: A Field and Laboratory Technicians' Guide to Artificial Insemination in Swine* (Morgan Morrow, New York, 1998).
61. Dube, C., Beaulieu, M., Reyes-Moreno, C., Guillemette, C. & Bailey, J. L. Boar sperm storage capacity of BTS and Androhep Plus: viability, motility, capacitation, and tyrosine phosphorylation. *Theriogenology* **62**, 874–886. https://doi.org/10.1016/j.theriogenology.2003.12.006 (2004).
62. Harrison, R. A., Jacques, M. L., Minguez, M. L. & Miller, N. G. Behaviour of ejaculated spermatozoa from bull, boar and ram during thin-layer countercurrent partition in aqueous two-phase systems. *J. Cell Sci.* **102**(Pt 1), 123–132 (1992).
63. Zeng, C. *et al.* Selection of optimal reference genes for quantitative RT-PCR studies of boar spermatozoa cryopreservation. *Cryobiology* **68**, 113–121. https://doi.org/10.1016/j.cryobiol.2014.01.004 (2014).
64. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.* **29**, e45. https://doi.org/10.1093/nar/29.9.e45 (2001).

65. Oh, S. A., Park, Y. J., You, Y. A., Mohamed, E. A. & Pang, M. G. Capacitation status of stored boar spermatozoa is related to litter size of sows. *Anim. Reprod. Sci.* **121**, 131–138. https://doi.org/10.1016/j.anireprosci.2010.05.019 (2010).
66. Kwon, W. S. *et al.* Applications of capacitation status for litter size enhancement in various pig breeds. *Asian Aust. J. Anim. Sci.* **31**, 842–850. https://doi.org/10.5713/ajas.17.0760 (2018).
67. Evans, M. I., Krivchenia, E. L., Wapner, R. J. & Depp, R. III. Principles of screening. *Clin Obstet Gynecol* **45**, 657–660 (2002) ((**Discussion 730–652**)).

Acknowledgements

Author contributions

W.K.P. designed and performed most experiments and analyzed the data; S.K. participated in sampling and RT-qPCR of optimization procedure; D.Y.R., M.S.R., and Y.J.P. conceived the design of experiment and data analysis; and M.G.P. supervized this study and wrote the manuscript with contribution from W.K.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-68486-1.

Correspondence and requests for materials should be addressed to M.-G.P.

Reprints and permissions information is available at www.nature.com/reprints.

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