Journal of Animal Science and Biotechnology

RESEARCH Open Access

Establishment of a male fertility prediction model with sperm RNA markers in pigs as a translational animal model

Won-Ki Pang, Shehreen Amjad, Do-Yeal Ryu, Elikanah Olusayo Adegoke, Md Saidur Rahman, Yoo-Jin Park and Myung-Geol Pang*

Abstract

Background: Male infertility is an important issue that causes low production in the animal industry. To solve the male fertility crisis in the animal industry, the prediction of sperm quality is the most important step. Sperm RNA is the potential marker for male fertility prediction. We hypothesized that the expression of functional genes related to fertilization will be the best target for male fertility prediction markers. To investigate optimum male fertility prediction marker, we compared target genes expression level and a wide range of field data acquired from artificial insemination of boar semen.

Results: Among the genes related to acrosomal vesicle exocytosis and sperm–oocyte fusion, equatorin (*EQTN*), zona pellucida sperm-binding protein 4 (*ZP4*), and sperm acrosome membrane-associated protein 3 exhibited high accuracy (70%, 90%, and 70%, respectively) as markers to evaluate male fertility. Combinations of *EQTN-ZP4*, *ZP4*-protein unc-13 homolog B, and *ZP4*-regulating synaptic membrane exocytosis protein 1 (*RIMS1*) showed the highest prediction value, and all these markers are involved in the acrosome reaction.

Conclusion: The *EQTN-ZP4* model was efficient in clustering the high-fertility group and may be useful for selection of animal that has superior fertility in the livestock industry. Compared to the *EQTN-ZP4* model, the *ZP4-RIMS1* model was more efficient in clustering the low-fertility group and may be useful in the diagnosis of male infertility in humans and other animals. The appointed translational animal model and established biomarker combination can be widely used in various scientific fields such as biomedical science.

Keywords: Male fertility, Multiple markers, Pig model, Prediction model, Sperm RNA

Introduction

Male infertility is a severe problem that extends across all animals. In humans, increasing male infertility threatens the deterioration of next-generation and eventually, the entire human race. The livestock industry comprising cow, pig, and chicken also suffers from male fertility issues. In cows and pigs, male infertility is directly correlated with the economics of the livestock industry and world food production. Concerning this issue, many scientists point out that a decrease in sperm quality causes male infertility. Therefore, the precise assessment of sperm quality is the first step to treat male infertility. Traditionally, sperm quality is evaluated employing laboratory techniques assessing sperm motility and viability [1, 2]. However, the clinical acceptance of these methods is still uncertain in many species [3, 4].

There is an emphasis on the importance of an omics approach to predict male fertility, instead of conventional laboratory assays [5–7]. Many studies indicate the

* Correspondence: mgpang@cau.ac.kr

Department of Animal Science & Technology and BET Research Institute, Chung-Ang University, Anseong, Gyeonggi-do 17546, Republic of Korea

potential of sperm RNA as a diagnostic tool for assessing male infertility [8–10]. Spermatozoa contain mRNAs, fragmented tRNAs, tsRNAs, miRNAs, lncRNAs, piRNAs, and siRNAs [11, 12]. Sperm RNAs are delivered to the oocyte [13] and serve critical roles like transportation of paternal phenotype [14–17], epigenetic modification of offspring genome [18], and aid the detection of transcripts to guide normal embryogenesis and implantation [19]. Moreover, comprehensive omics studies have revealed that the differential expression of sperm RNA is related to different fertility statuses in humans, porcine, bovine, ovine, and equine species [20–24]. Despite the functional and clinical importance of sperm RNA, studies that use RNA to evaluate male fertility are still in the experimental stage due to the lack of an accurate fertility model for clinical trials.

Pig, as an animal model, contributes greatly to biomedical research, particularly studies on obesity, arthritis, cardiovascular disease, skin and eye conditions, and xenotransplantation [25, 26]. Compared to rodent models, pigs are more physiologically related to humans [27]. In addition, pig fertility is digitized accurately through artificial insemination (AI) results [28]. Considering such diverse characteristics, pigs are an adequate animal model to study the clinical acceptance of sperm RNA on male fertility in animals, including humans.

The sperm-oocyte interaction is the starting point of embryonic development, and for successful interaction, all molecular mechanisms must work in harmony [29]. Before sperm-oocyte interaction, the acrosome of the sperm undergoes several changes for fusion with the zona pellucida. This biological process, called the acrosome reaction, involves acrosomal exocytosis, substructure remodeling, and other biochemical modifications [30]. These processes are inevitable for the successful fusion of sperm to the oocyte plasma membrane in single fertilization [31]. The normal functioning of genes related to the aforementioned mechanisms leads to fertility success, and the expression pattern of genes is closely related to the corresponding mechanisms. While such genes provide early clues regarding male fertility in a rodent model [13], little is known about their function in humans and other large animals. Considering their important roles in rodents, we postulated that they may play a key role in male fertility in other large animals. Therefore, in this study, we aimed to identify male fertility markers by screening the mRNA expression levels of genes that encode proteins crucial for acrosomal vesicle exocytosis and sperm–oocyte fusion in pig spermatozoa. To understand the physiological roles of these genes in male fertility in other large animals, we selected pig as the translational animal model of male fertility. We also concluded the accuracy and sensitivity of our protocol by employing large samples of boar semen representing a wide range of levels of field data.

Materials and methods

Experimental design

Firstly, 'acrosomal vesicle exocytosis' and 'sperm-egg fusion' related genes were listed to investigate relevant boar fertility prediction marker models.

Secondly, to narrow down target genes, high (average litter size = 13.97 ± 0.07)- and low (average litter size = 11.17 ± 0.13)-fertility representative samples ($n = 3$) were chosen based on total litter size data after AI trials. The differentially expressed genes in high- and low-fertility groups were progressed to further process.

Thirdly, the expression of genes from the second step was tested in 20 randomly selected Yorkshire boars and then compared with litter size. The genes that were significantly correlated with litter size were subjected to marker evaluation with the receiver operating characteristic (ROC) curve analysis.

Finally, each evaluated markers were combined to improve boar fertility prediction value. Multiple marker model predictability of litter size (over and under cut-off value) was evaluated with ROC curve analysis. Multiple marker models which exhibited outstanding prediction value in the previous analysis were subjected to further statistical analysis. The predictability of marker models to distinguish boars into three groups (high, medium, and low) was evaluated to develop an optimum combination.

Animal care

All procedures involving animals were approved by the Institutional Animal Care and Use Committee of Chung-Ang University (approval number: 2017–00018). All boars and sows used in the experiment were reared at 20 ± 5 °C under a 2:1 light/dark cycle, continuous air circulation, and appropriate feed and water.

Sperm sample preparation

A total of 26 Yorkshire boar (body weight over 90 kg and age between 9 to 24 months) semen was collected using the gloved-hand technique [32] from Sunjin Co. (Danyang, Korea). The motile spermatozoa were separated from semen with discontinuous 35% and 70% Percoll gradient (Sigma-Aldrich Co., St. Louis, MO, USA) [33]. The spermatozoa were cultured using modified tissue culture media 199 (containing 10% fetal bovine serum [v/v], 0.91 mmol/L sodium pyruvate, 3.05 mmol/L D-glucose, 2.92 mmol/L calcium lactate, and 2.2 g/L sodium bicarbonate) (Sigma-Aldrich Co.) under 5% CO_2 and at 37 °C for 30 min [34]. The sperm samples were pelleted and then deep-frozen with liquid nitrogen (– 196 °C) and stored at – 80 °C before use.

Sperm motility and motion kinematics assessment

Sperm motility and motion kinematics were analyzed by a computer-assisted sperm analysis system (SAIS-Plus ver.10.1; Medical Supply, Seoul, Korea) [35]. Washed spermatozoa were resuspended in modified tissue culture media 199. Then, under a microscope, the motility (%), hyperactivation (HYP; %), curvilinear velocity (VCL; μm/s), straight-line velocity (VSL; μm/s), average path velocity (VAP; μm/s), linearity (LIN; %), beat cross frequency (BCF; Hz), wobble (WOB; %), and mean amplitude of head lateral displacement (ALH; %) of the spermatozoa were measured on a pre-heated (37 °C) Makler counting chamber (Sefi Medical Instruments, Haifa, Israel). For a single analysis, 200–300 spermatozoa were subjected, and all samples were analyzed three times.

Capacitation status assessment

The spermatozoa were stained with chlortetracycline (Sigma-Aldrich Co.) and Hoechst 33258 (Sigma-Aldrich Co.) and their capacitation status were measured in approximately 400 spermatozoa per slide [36, 37]. Each sample was smeared into three slides. Briefly, the incubated spermatozoa are stained with Hoechst 33258 solution. Then, the excess dye was inactivated with a 2% polyvinylpyrrolidone (Sigma-Aldrich) solution. Stained spermatozoa were pelleted and then resuspended in 600 μL of PBS and 600 μL of chlortetracycline (CTC) fluorescence solution (750 mmol/L CTC in 5 μL buffer; 20 mmol/L Tris, 130 mmol/L sodium chloride, and 5 mmol/L cysteine, pH 7.4) (Sigma-Aldrich). The stained samples were counted under epifluorescence illumination with UV BP 340–380/LP 425 and BP 450–490/LP 515 excitation/emission filters for Hoechst 33258 and CTC, respectively, using a Microphot-FXA microscope (Nikon, Tokyo, Japan). Sperm capacitation status was further classified into four categories: live non-capacitated (F; yellow fluorescence distributed evenly throughout the sperm head), live capacitated (B; yellow fluorescence over the acrosome region and a dark post-acrosome region), acrosome-reacted (AR; showing no fluorescence over the head), and dead (D; nuclei with blue fluorescence within the sperm head).

Sperm RNA isolation

Sperm RNA was isolated according to a previous study [38]. Briefly, sperm numbers were adjusted to 40–50 × 10^6 cells/mL. Sperm pellets were suspended in non-toxic guanidine-isothiocyanate lysis buffer containing β-mercaptoethanol (40 μL/mL; Sigma-Aldrich Co.) and homogenized using a 20G syringe. TRIzol (500 μL) (Invitrogen, Carlsbad, CA, USA) and chloroform (200 μL) were added to the homogenized sample and centrifuged at 12,000 × *g* for 25 min. After centrifugation, the upper layer (500 μL) was moved to a fresh tube and mixed with pure ethanol (500 μL) (Sigma-Aldrich Co.). Sperm RNA was attached to a spin cartridge and washed with the PureLink RNA Mini Kit (Invitrogen) wash buffer 1 and 2. The isolated RNA was immersed in 20 μL of nuclease-free water (60 °C). The quality (260/280 ratio) and quantity of isolated RNA were measured using an Epoch Microplate Spectrophotometer (BioTek, Winooski, VT, USA).

Reverse transcription-quantitative real-time polymerase chain reaction

Reverse transcription was performed using PrimeScript 1st strand cDNA Synthesis Kit (Takara Bio, Inc., Shiga, Japan) according to the manufacturer's protocol. Total 400 ng of sperm RNA from each sample was reverse transcribed and 8–11 μg of cDNAs were yielded. The relative expression of target genes was quantified using the 7500 fast real-time PCR system (Applied Biosystems, Foster City, CA, USA). SYBR Green PCR master mix (Applied Biosystems) was used. The total reaction volume was 20 μL (100 ng of cDNA), and the cycling conditions were set to initial denaturation at 95 °C (10 min) followed by 40 cycles at 95 °C for denaturation (15 s) and 60 °C for annealing (60 s). The results were analyzed using the 7500 Software v2.3 (Thermo Fisher Scientific, Waltham, MA, USA) using the $2^{-\Delta\Delta Cq}$ method. All primers used for qPCR were designed based on Sscrofa11.1 Genome Assembly from The Swine Genome Sequencing Consortium (Additional file 1: Table S1).

Artificial insemination

Male fertility status of randomly selected 20 Yorkshire boars was evaluated from the results obtained after AI using their semen. Boar semen was diluted with 100 mL of the Beltsville thawing solution (30×10^6 sperm cells/mL) [39]. AI was performed twice per estrus [40] in an average of 19.3 ± 1.2 sows per boar (total 386 trials). The average number of piglets born was considered as the male fertility status of the corresponding boar.

Bioinformatics

The Gene Ontology database was used to categorize target genes related with the functional classes "acrosomal vesicle exocytosis" and "sperm–egg fusion." The listed genes were applied to the Pathway Studio program (Elsevier) to visualize their biological functions.

Statistics

All data obtained were tested for normality and homogeneity of variance. The normality test (Shapiro–Wilk test), homogeneity of variances test (Levene's test), Student's *t*-test, linear and multiple regression test, and ROC curve analysis were performed using the SPSS v1.8

software (SPSS, Inc., Chicago, IL, USA). principal component analysis (PCA), *k*-medoids clustering, silhouette plotting, and heatmap analysis were conducted using the R software (RStudio, Boston, MA, USA). $P < 0.05$ was considered to indicate significantly different results and was correlated with fertility. Standard error of the mean was used to show sampling distribution.

Results

Sperm motility, motion kinematics, and capacitation status of boar spermatozoa

The functional parameters of spermatozoa, including sperm motility, hypermotility, motion kinetics, and capacitation status were analyzed from all boar spermatozoa. No functional parameter showed a significant difference between the high- and low-fertility groups (Table 1). The sperm motility, motion kinematics, and capacitation status of 20 randomly selected Yorkshire boars were shown in Additional file 1: Table S2.

Expression of fertility-related genes in the high- and low-fertility groups

We found that equatorin (*EQTN*), zona pellucida sperm-binding protein 4 (*ZP4*), regulating synaptic membrane exocytosis protein 1 (*RIMS1*), Ras-related protein Rab-3A (*RAB3A*), protein unc-13 homolog B (*UNC13B*), and synaptotagmin-6 (*SYT6*) are involved in acrosomal vesicle exocytosis. Moreover, *EQTN*, Izumo sperm–egg fusion protein 1 (*IZUMO1*), sperm acrosome membrane-associated protein 3 (*SPACA3*), *CD9*, lysozyme like protein 6 (*LYZL6*), spermatogenesis-associated protein 46 (*SPATA46*), and IZUMO1 receptor (*IZUMO1R*) are involved in sperm-oocyte fusion. All genes associated with acrosomal vesicle exocytosis are differentially expressed in the high and low litter size groups (Fig. 1A–E; $P < 0.05$). Of the genes categorized in "sperm–egg fusion" *EQTN*, *IZUMO1*, *SPACA3*, *CD9*, and *IZUMO1R* were differentially expressed in the high and low litter size groups (Fig. 1E–I and L; $P < 0.05$). *LYZL6* and *SPATA46* were not differentially expressed according to litter size (Fig. 1J and K, $P < 0.05$). Of the twelve genes investigated, ten differentially expressed genes were selected to be further evaluated as candidates for male fertility markers in validation trials.

Table 1 Functional parameters of spermatozoa and litter size of the high- and low-fertility groups

	High fertility	Low fertility
Litter size	13.97 ± 0.07*	11.17 ± 0.13
MOT, %	72.94 ± 18.26	70.19 ± 14.68
HYP, %	13.19 ± 6.80	12.67 ± 7.57
VCL, μm/s	134.59 ± 21.57	134.21 ± 25.00
VSL, μm/s	58.66 ± 9.32	65.51 ± 16.53
VAP, μm/s	69.96 ± 11.02	74.29 ± 14.75
LIN, %	43.87 ± 3.14	49.25 ± 7.96
BCF, Hz	13.75 ± 1.10	14.09 ± 1.28
WOB, %	52.10 ± 1.90	55.94 ± 5.31
ALH, μm	5.96 ± 0.92	6.01 ± 1.01
AR, %	2.68 ± 0.88	8.20 ± 6.95
F, %	65.24 ± 10.42	76.92 ± 9.32
B, %	32.08 ± 10.58	14.88 ± 2.80

MOT motility, *HYP* hyperactivation, *VCL* curvilinear velocity, *VSL* straight line velocity, *VAP* average path velocity, *LIN* linearity, *BCF* beat cross frequency, *WOB* wobble, *ALH* mean amplitude of head lateral displacement, *AR* acrosome reacted, *F* normal, *B* capacitated. Data are expressed as mean ± SEM; *$P < 0.05$

Assessment of single fertility marker

The expression of genes encoding *ZP4, RIMS1, RAB3A, UNC13B, SYT6, EQTN, IZUMO1, SPACA3, CD9*, and *IZUMO1R* were screened in 20 randomly selected boars. The fertility of boars was significantly correlated with the expression of *EQTN*, *ZP4*, *UNC13B*, *RIMS1*, and *SPACA3* (Fig. 2B–F; $P < 0.05$). The expression pattern of the genes encoding these five proteins in 20 randomly selected boars was the same as the expression pattern of the high- and low-fertility groups. The male fertility prediction value was assessed with ROC curve analysis [41]. The area under the curve ranged from 0.6 to 0.9, and overall accuracy (OA) ranged from 45% to 90% (Fig. 2A and Table 2). *ZP4* showed the highest values of sensitivity (SN), specificity (SP), negative predictive value (NPV), positive predictive value (PPV), and OA, all of which were 90% (Table 2). The OA of *UNC13B* and *RIMS1* were comparatively low (Table 2).

Assessment and optimization of multiple fertility marker model

To improve the clinical validity of the fertility marker candidate, we performed multiple regression tests using all possible marker combinations with *EQTN*, *ZP4*, *UNC13B*, *RIMS1*, and *SPACA3*. A total of 14 combinations were chosen to identify a significant marker model to predict male fertility (Table 3, Fig. 3). The combination of *EQTN-ZP4* (SN = 100%, SP = 100%, NPV = 100%, PPV = 100%, OA = 100%), *ZP4-UNC13B* (SN = 100%, SP = 100%, NPV = 100%, PPV = 100%, OA = 100%), and *ZP4- RIMS1* (SN = 100%, SP = 100%, NPV = 100%, PPV = 100%, OA = 100%) showed outstanding predictive value (Table 3, Fig. 3A; $P < 0.05$). The combinations of *UNC13B-SPACA3-CD9* (SN = 80%, SP = 80%, NPV = 80%, PPV = 80%, OA = 80%) and *RIMS1-SPACA3-CD9* (SN = 80%, SP = 80%, NPV = 80%, PPV = 80%, OA = 80%) showed the highest predictive values among three-marker predictive models (Table 3, Fig. 3B; $P < 0.05$). The multiple regression model was established (Fig. 4A–E). In the correlation heatmap, the multiple-marker models were more clustered and showed higher

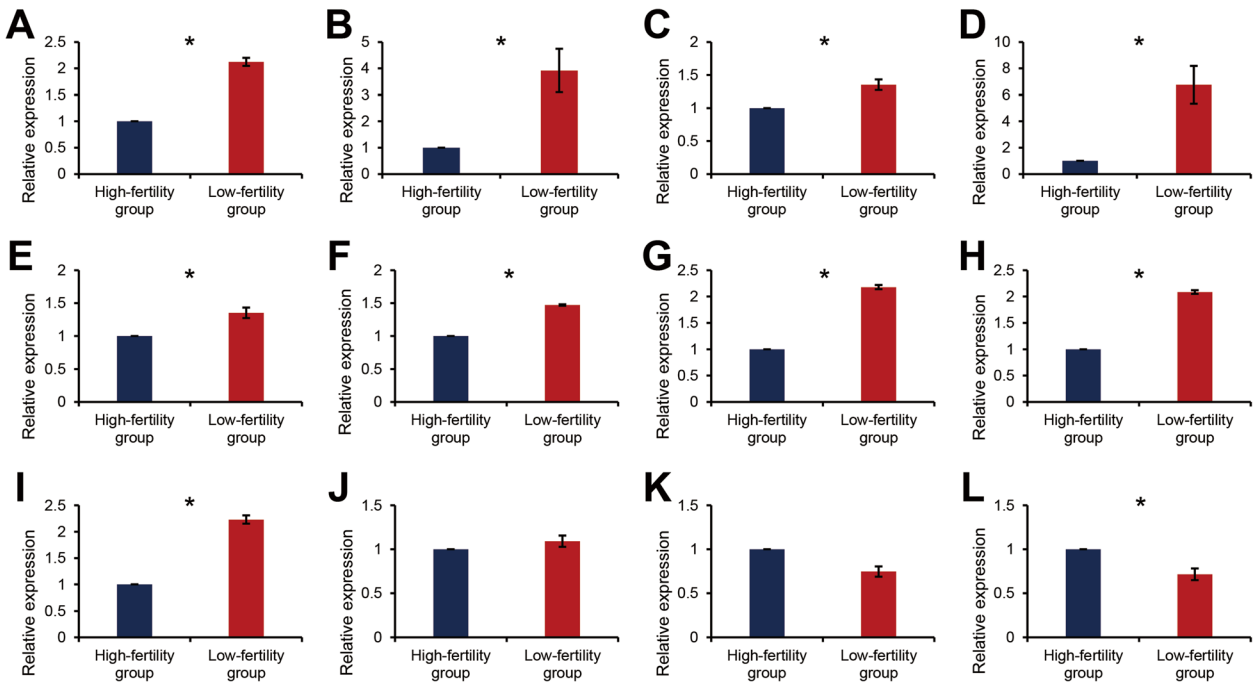

Fig. 1 The expression pattern of target genes in the high- and low-fertility groups. Relative mRNA expression of (**A**) *ZP4*, (**B**) *RIMS1*, (**C**) *RAB3A*, (**D**) *UNC13B*, (**E**) *SYT6*, (**F**) *EQTN*, (**G**) *IZUMO1*, (**H**) *SPACA3*, (**I**) *CD9*, (**J**) *LYZL6*, (**K**) *SPATA46*, and (**L**) *IZUMO1R* in the high- and low-fertility groups. ($n = 3$). * $P < 0.05$

correlation value than the single-marker models (Fig. 5A, $P < 0.05$). To identify the predictive model most efficient in diagnosing male fertility in the high-, medium-, and low-fertility groups, PCA, k-medoids clustering, and silhouette plotting were employed. The average litter size of high-, medium-, and low-fertility groups after clustering with each prediction model is shown in Additional file 1: Table S3. The PCA plot showed that the *EQTN-ZP4* model is the most efficient in diagnosing the high-fertility group (Fig. 4F) compared to other multiple-marker models (Fig. 4G–J). Moreover, in clustering analysis, the *EQTN-ZP4* model was efficient in clustering the high-fertility group (Fig. 4F, K, and P) compared to other multiple-marker models (Fig. 4G–J, L–O, and Q–T). Compared to *EQTN-ZP4*, the *ZP4-RIMS1* and *RIMS1-SPACA3-CD9* models were more effective in

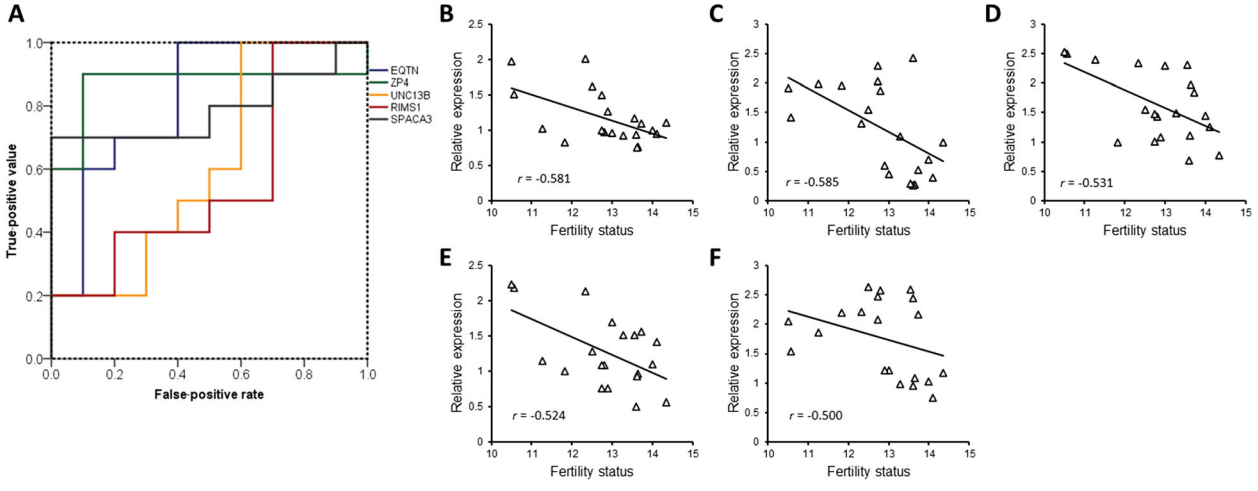

Fig. 2 Assessment of single fertility markers. **A** The receiver operating characteristic curve of *EQTN, ZP4, UNC13B, RIMS1*, and *SPACA3* for single marker evaluation. Linear regression of (**B**) *EQTN*, (**C**) *ZP4*, (**D**) *UNC13B*, (**E**) *RIMS1*, and (**F**) *SPACA3* and fertility status of 20 randomly selected pigs. *r*, Pearson coefficient; $P < 0.05$

Table 2 Male fertility prediction value of single markers

Gene	Cut-off value	AUC	Sensitivity, %	Specificity, %	NPV, %	PPV, %	OA, %
EQTN	1.0	0.8	70.0	70.0	70.0	70.0	70.0
ZP4	1.2	0.9	90.0	90.0	90.0	90.0	90.0
UNC13B	1.5	0.6	60.0	50.0	55.6	54.5	55.0
RIMS1	1.1	0.6	50.0	40.0	44.4	45.5	45.0
SPACA3	2.0	0.8	70.0	70.0	70.0	70.0	70.0

AUC Area under the curve. Sensitivity is the percentage of pigs showing true-positive results when tested with mRNA expression. Specificity is the percentage of pigs showing true-negative results. The positive predictive value (PPV) is the percentage of pigs that tested as positive and simultaneously showed true-positive litter size. The negative predictive value (NPV) is the percentage of pigs that tested as negative and simultaneously showed true-negative litter size. *OA* overall accuracy

clustering the low-fertility group by excluding the high- and medium-fertility groups (Fig. 4G, J, L, O, Q, and T).

Marker-related diseases and cell processes

To understand the functional significance of the genes contained in male fertility predictive models, the Pathway Studio program (v 12.0, Elsevier, Amsterdam, NL) was used, and the following gene sets were applied to investigate the changing patterns of marker-related diseases and cell processes: all 12 genes related to "acrosomal vesicle exocytosis" and "sperm–egg fusion," 10 differentially expressed genes in the high- and low-fertility groups, and 5 genes passed the validation trial. The percentage of fertility-related diseases and cell processes changed according to the validation level of each gene set (Fig. 5B and C). It was the highest when the five genes that passed validation trials were applied to Pathway Studio to find related diseases (11.1%; Fig. 5B) and cell processes (8.7%; Fig. 5C). In fertility-related diseases, terms like "infertility," "germ cell neoplasm," "male infertility," and "male sterility" were contained in the program (Additional file 2: Table S4). For the fertility-related cell processes, the terms like "acrosome exocytosis," "sperm cell function," "spermatogenesis," "sperm-oocyte fusion," and "sperm cell adhesion" were found (Additional file 2: Table S4). Specifically, the genes part of the male fertility predictive model was analyzed to find the related cell processes. We found that acrosome reaction is involved in all the two-gene male fertility predictive models (Fig. 5D–F). We also found that the genes in the *UNC13B-SPACA3-CD9* and *RIMS1-SPACA3-CD9* male fertility predictive models played a common role in membrane fusion (Fig. 5G and H).

Discussion

To the best of our knowledge, no systemic empirical research exists addressing the question of how to translate basic research findings into new diagnoses in male

Table 3 Male fertility prediction value of multiple marker models

Prediction model	AUC	Sensitivity, %	Specificity, %	NPV, %	PPV, %	OA, %
EQTN-ZP4	1.0	100.0	100.0	100.0	100.0	100.0
ZP4-UNC13B	1.0	100.0	100.0	100.0	100.0	100.0
ZP4-RIMS1	1.0	100.0	100.0	100.0	100.0	100.0
UNC13B-CD9	0.6	50.0	50.0	50.0	50.0	50.0
IZUMO1R-SPACA3	0.8	70.0	60.0	66.7	63.6	65.0
IZUMO1-SYT6	0.7	70.0	60.0	66.7	63.6	65.0
SPACA3-CD9	1.0	90.0	80.0	88.9	81.8	85.0
CD9-RAB3A	0.8	80.0	80.0	80.0	80.0	80.0
CD9-SYT6	0.8	70.0	70.0	70.0	70.0	70.0
UNC13B-IZUMO1-SYT6	0.7	70.0	60.0	66.7	63.6	65.0
UNC13B-SPACA3-CD9	0.9	80.0	80.0	80.0	80.0	80.0
UNC13B-CD9-SYT6	0.7	70.0	70.0	70.0	70.0	70.0
RIMS1-SPACA3-CD9	0.9	80.0	80.0	80.0	80.0	80.0
RIMS1-CD9-SYT6	0.7	70.0	60.0	66.7	63.6	65.0

AUC Area under the curve. Sensitivity is the percentage of pigs showing true-positive results when tested with mRNA expression. Specificity is the percentage of pigs showing true-negative results. The positive predictive value (PPV) is the percentage of pigs that tested as positive and simultaneously has a true-positive litter size. The negative predictive value (NPV) is the percentage of pigs that tested as negative or simultaneously had a true-negative litter size. *OA* overall accuracy

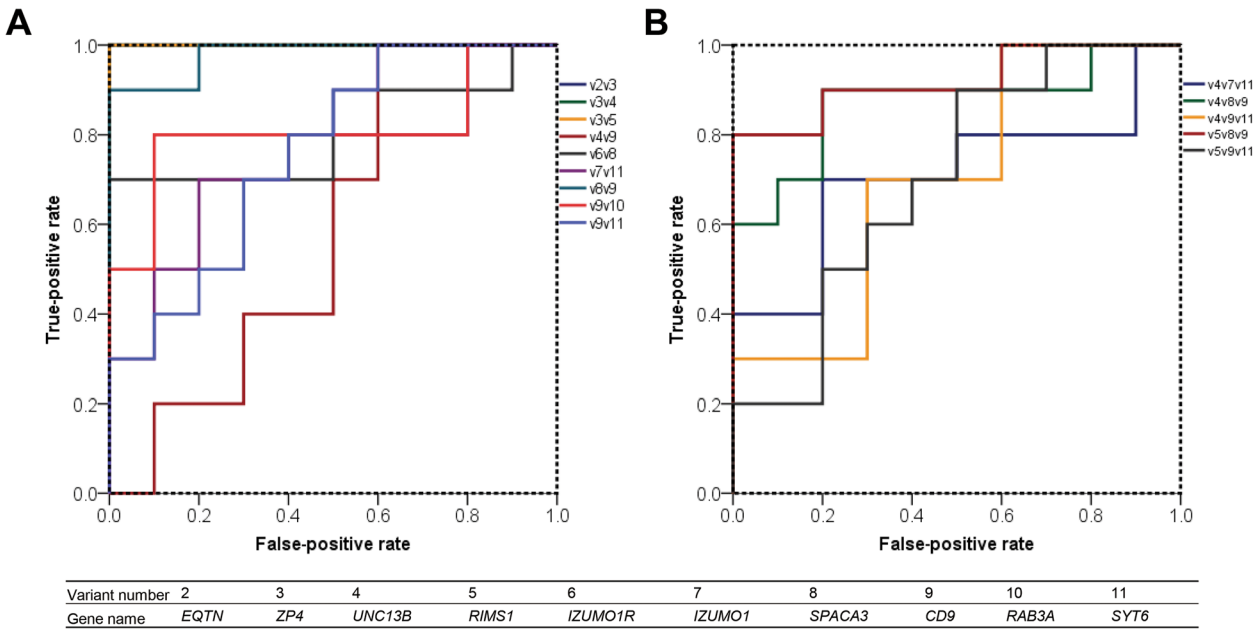

Fig. 3 Assessment of multiple fertility markers. **A** The receiver operating characteristic curve of the predictive ability of the two-gene marker models for male fertility prediction. **B** The receiver operating characteristic curve of the predictive ability of the three-gene marker models for male fertility prediction

infertility. This is the first translational study to predict male infertility using pigs with sperm RNA markers. In this study, we found that *EQTN*, *ZP4*, and *SPACA3* exhibited high accuracy as markers to evaluate male fertility. Combinations of *EQTN-ZP4*, *ZP4-UNC13B*, and *ZP4-RIMS1* showed the highest prediction value. The *EQTN-ZP4* model was efficient for clustering the high-fertility group and may be useful for selection of animal that has superior fertility. Compared to the *EQTN-ZP4* model, the *ZP4-RIMS1* model was more efficient in clustering the low-fertility group.

Evaluation of sperm quality is critical in the timely diagnosis and prevention of male infertility. Most diagnostic tools analyze the phenotype and morphology of spermatozoa. However, the clinical value of such current sperm evaluation methods is still debatable [3] because of the complexity of processes involved in achieving male fertility. Rodent models are widely used to understand the highly structured mechanism of male fertility [42, 43]. The short gestation period and well-established gene regulation technique are the powerful benefits of using a rodent model for investigating male fertility. However, it is hard to apply the findings from studies on rodents to humans or other large animals directly for the following reasons: (i) insufficient clinical/phenotypic information; (ii) lack of field/clinical trial; (iii) the high genetic diversity of human and other large animals; and (iv) the genetic differences between rodents and human/large animals. The pig model has high genetic and physiological similarities with humans [27, 44] and is a better candidate for reproductive research than rodent model. Recently, Yue et al. presented a pig germline genome engineering technique [45]. Moreover, huge and delicate fertility data acquired from AI help to get qualified field data. These factors make pigs more accessible in studying male fertility. Therefore, we selected the pig as a translational animal model to determine a prominent marker to evaluate male fertility.

Large-scale transcriptomic analysis in many mammalian species has revealed fertility-related differential gene expressions [20–24]. Although the huge dataset available is valuable for future male fertility studies, the clinical application of differentially expressed genes as markers is uncertain. Recently, our previous studies found fertility markers to predict male fertility [46–48]. However, because of the polygenicity of male fertility, a single marker is not enough to predict male fertility [49, 50]. Thus, we used a multiple-marker approach to establish an optimum male fertility prediction model. We made the following efforts: (i) listed the genes related to two critical biological events that determine fertilization success; (ii) found a single marker to predict male fertility by analyzing field data; (iii) tested combinations of single markers to establish multiple-marker prediction models, and (iv) evaluated the multiple-marker prediction models.

The marker genes *ZP4*, *EQTN*, and *SPACA3* showed high accuracy in male fertility prediction. ZP4 is known

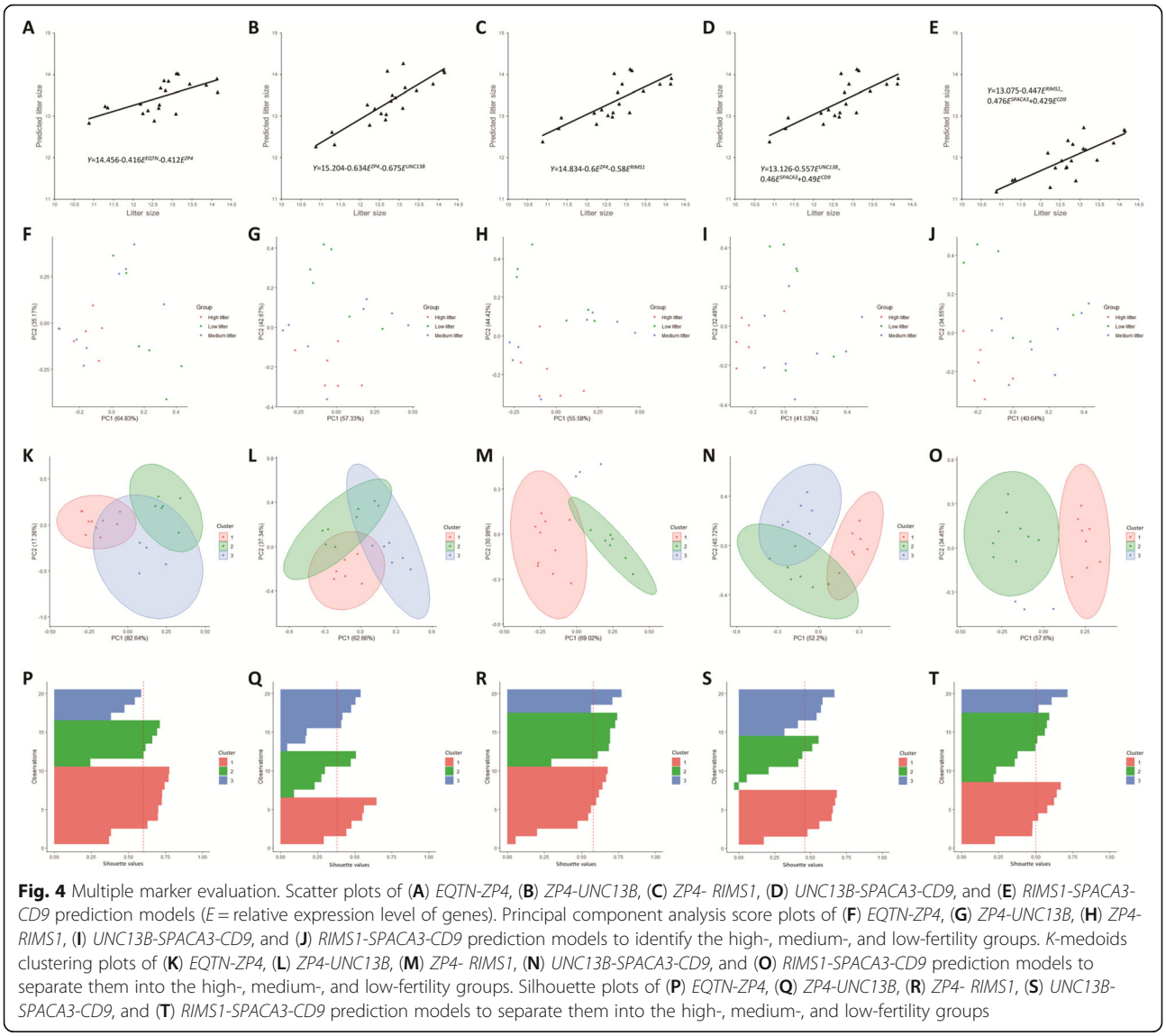

Fig. 4 Multiple marker evaluation. Scatter plots of (**A**) *EQTN-ZP4*, (**B**) *ZP4-UNC13B*, (**C**) *ZP4- RIMS1*, (**D**) *UNC13B-SPACA3-CD9*, and (**E**) *RIMS1-SPACA3-CD9* prediction models (*E* = relative expression level of genes). Principal component analysis score plots of (**F**) *EQTN-ZP4*, (**G**) *ZP4-UNC13B*, (**H**) *ZP4-RIMS1*, (**I**) *UNC13B-SPACA3-CD9*, and (**J**) *RIMS1-SPACA3-CD9* prediction models to identify the high-, medium-, and low-fertility groups. *K*-medoids clustering plots of (**K**) *EQTN-ZP4*, (**L**) *ZP4-UNC13B*, (**M**) *ZP4- RIMS1*, (**N**) *UNC13B-SPACA3-CD9*, and (**O**) *RIMS1-SPACA3-CD9* prediction models to separate them into the high-, medium-, and low-fertility groups. Silhouette plots of (**P**) *EQTN-ZP4*, (**Q**) *ZP4-UNC13B*, (**R**) *ZP4- RIMS1*, (**S**) *UNC13B-SPACA3-CD9*, and (**T**) *RIMS1-SPACA3-CD9* prediction models to separate them into the high-, medium-, and low-fertility groups

to play a critical role in the taxon-specific binding of the spermatozoa to the oocyte [51] and induce the acrosomal reaction in human spermatozoa in a dose-dependent manner in humans [52]. Although the function of the ZP4 protein in spermatozoa is unknown, its mRNA expression is found in rodents and humans [53]. Evaluation from an evolutionary perspective can help us understand how *ZP4* mRNA expression is associated with fertilization [54]. In *Arabidopsis*, short suspensor transcripts are expressed during male gametogenesis and conveyed to female gamete for zygotic translation. Paternally derived *ZP4* transcripts might affect fertilization by aiding in detecting transcripts in the zygote.

EQTN is expressed in the equatorial segment of spermatozoa and induces an acrosome reaction to help sperm–oocyte fusion [55–57]. EQTN protein was differentially expressed according to the fertility status of the pig [4]. SPACA3 protein is expressed in cow [58] and human spermatozoa [59], and it acts by interacting with other sperm proteins to facilitate fertilization [60]. In spermatozoa, mRNA-protein interaction is still unclear. Therefore, it is suggested that sperm mRNAs have roles separate from the functional roles of sperm proteins. In another view, a study by Gur and Breitbart suggested the presence of mitochondrial translation in spermatozoa [61] and provided insights into the physiological roles of *EQTN* and *SPACA3* transcripts in spermatozoa. mRNA levels cannot directly be correlated with protein expression as post-transcriptional processes are crucial in the final synthesis of the native protein [62]. Therefore, *EQTN* and *SPACA3* transcripts may be associated with male fertility through the regulation of complex processes including acrosome reaction, sperm–oocyte fusion, fertilization, and beyond. Although, the mRNA

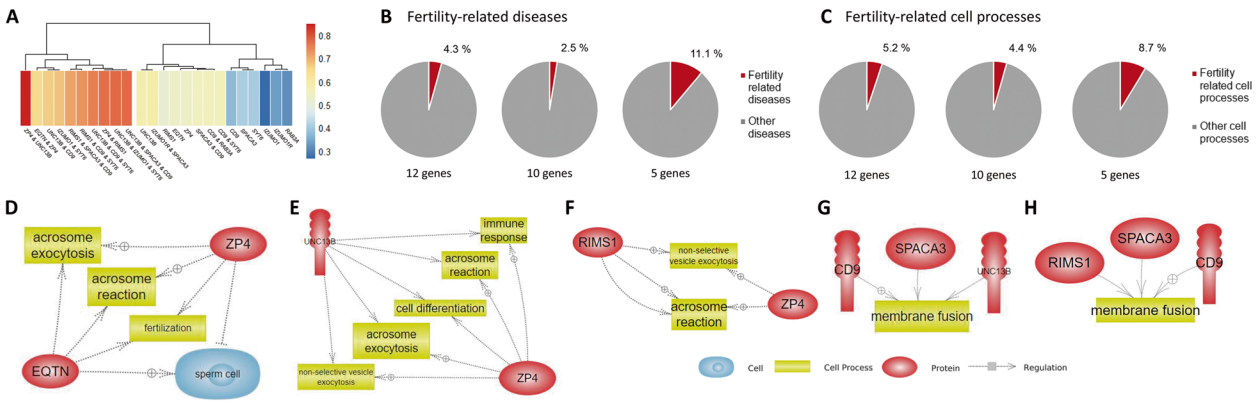

Fig. 5 Heatmap and bioinformatics analysis of marker models. **A** Heatmap showing the coefficient values of the multiple-marker prediction models and single markers. **B** Pie chart denoting the percentage of fertility-related diseases associated with the different number of genes obtained by Pathway Studio. **C** Pie chart denoting the percentage of fertility-related cell processes associated with the different number of genes obtained by Pathway Studio. The common cell processes associated with genes in the (**D**) *EQTN-ZP4*, (**E**) *ZP4-UNC13B*, (**F**) *ZP4- RIMS1*, (**G**) *UNC13B-SPACA3-CD9*, and (**H**) *RIMS1-SPACA3-CD9* prediction models

expression of *EQTN* and *SPACA3* are prominent markers for determining male fertility, the elucidation of the functional roles of mRNA in sperm warrants further research.

After identifying single fertility markers, we conducted multiple regression tests with all possible gene combinations to establish an optimum prediction model. Multiple markers for the diagnosis of various diseases, such as cancer and neurological diseases have been developed [63, 64]. We established the *EQTN-ZP4*, *ZP4-UNC13B*, and *ZP4- RIMS1* prediction models that showed 100% accuracy in the prediction of the fertility status in the validation trial data. All genes in the three-gene prediction models were related to acrosome reaction. This indicates the importance of acrosome reaction in fertilization success. Interestingly, *ZP4* and *UNC13B* showed "immune response" and "cell differentiation" as related cell processes. Thus, *ZP4* and *UNC13B* might be potential targets for future studies on unraveling the connection between immune response and fertility.

For further evaluation of the multiple-marker prediction models, PCA and clustering were conducted. PCA is a powerful tool to combine multiple factors that affect the main variable and has been used to identify biomarkers of various diseases [65–68]. In the livestock industry, identifying animals with superior genetics is crucial for increasing productivity. Therefore, the *EQTN-ZP4* model will be useful in this field because of its efficiency in clustering high-fertility groups. Although we only focused on the reproductive traits of pigs, other traits such as immune resistance, heat stress resistance, and growth rate could be improved by adopting the same approach. Compared to the *EQTN-ZP4* model, the *ZP4- RIMS1* and *RIMS1-SPACA3-CD9* models were more efficient in clustering the low-fertility group by excluding the high- and medium-fertility groups. Therefore, we suggest that these models may be useful in biomedical studies to detect male infertility/sterility.

Conclusions

In conclusion, our study established an efficient male fertility prediction model using the pig model and multiple gene combinations. These genes were involved in biological events critical to male fertility. The *ZP4* gene is highly predictive of male fertility. *ZP4* in combination with *EQTN* has a higher predictive value and may be useful in the livestock industry. The *EQTN-ZP4* male fertility prediction model demonstrated outstanding predictive value in detecting the low-fertility group. Therefore, we postulate that this pig as a translational model and biomarker combination can be utilized to diagnose male infertility/sterility. However, as our study was conducted in a single pig stud, validation of our proposed prediction model requires further investigation and trans-species evaluation.

Abbreviations

AI: Artificial insemination; ALH: Mean amplitude of head lateral displacement; BCF: Beat cross frequency; cDNA: Complementary DNA; CTC: Chlortetracycline; EQTN: Equatorin; HYP: Hyperactivation; IZUMO1: Izumo sperm–egg fusion protein 1; IZUMO1R: IZUMO1 receptor; LIN: Linearity; lncRNA: Long non-coding RNA; LYZL6: Lysozyme like protein 6; miRNA: MicroRNA; mRNA: Messenger RNA; NPV: Negative predictive value; OA: Overall accuracy; PCA: Principal component analysis; piRNA: Piwi-interacting RNA; PPV: Positive predictive value; qPCR: Quantitative polymerase chain reaction; RAB3A: Ras-related protein Rab-3A; RIMS1: Regulating synaptic membrane exocytosis protein 1; ROC: Receiver operating characteristic; siRNA: Short interfering RNA; SN: Sensitivity; SP: Specificity; SPACA3: Sperm acrosome membrane-associated protein 3; SPATA46: Spermatogenesis-associated protein 46; SYT6: Synaptotagmin-6; tRNA: Transfer RNA; tsRNA: Transfer RNA-derived small RNA; UNC13B: Protein unc-13 homolog B; VAP: Average path velocity; VCL: Curvilinear velocity; VSL: Straight-line velocity; WOB: Wobble; ZP4: Zona pellucida sperm-binding protein 4.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40104-022-00729-9.

Additional file 1: Table S1. Primers designed for RT-qPCR. **Table S2.** Sperm motility, motion kinematics, and capacitation status of 20 randomly selected boar spermatozoa. **Table S3.** Average litter size of high-, medium-, and low-fertility groups after clustering.

Additional file 2: Table S4. All fertility-related diseases and cell processes related with markers.

Acknowledgements

Not applicable.

Authors' contributions

This study was conceived and designed by WKP and MGP. WKP, JHS, DYR, and YJP performed the experiments. The data were analyzed by SA, EOA, and WKP. The manuscript was mainly written by WKP and MGP, with the assistance of EOA, DYR, MSR, and YJP. All authors read and approved the final manuscript.

Funding

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2018R1A6A1A03025159).

Availability of data and materials

The datasets from the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Experimental procedures involving animals were performed in accordance with Institutional Animal Care and Use Committee of Chung-Ang University guidelines and were approved by the corresponding committee (Approval No. 2017–00018).
Consent to participate is not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 January 2022 Accepted: 8 May 2022
Published online: 07 July 2022

References

1. Rodriguez-Martinez H. Laboratory semen assessment and prediction of fertility: still utopia? Reprod Domest Anim. 2003;38(4):312–8. https://doi.org/10.1046/j.1439-0531.2003.00436.x.
2. Khatun A, Rahman MS, Pang MG. Clinical assessment of the male fertility. Obstet Gynecol Sci. 2018;61(2):179–91. https://doi.org/10.5468/ogs.2018.61.2.179.
3. Lewis SE. Is sperm evaluation useful in predicting human fertility? Reproduction. 2007;134(1):31–40. https://doi.org/10.1530/REP-07-0152.
4. Kwon WS, Rahman MS, Lee JS, Yoon SJ, Park YJ, Pang MG. Discovery of predictive biomarkers for litter size in boar spermatozoa. Mol Cell Proteomics. 2015;14(5):1230–40. https://doi.org/10.1074/mcp.M114.045369.
5. Vijayalakshmy K, Kumar D, Virmani M, Jacob N, Kumar P. Sperm transcriptomics: an emerging technique to assess male fertility. Int J Curr Microbiol App Sci. 2018;7(09):1188–200. https://doi.org/10.20546/ijcmas.2018.709.141.
6. Burl RB, Clough S, Sendler E, Estill M, Krawetz SA. Sperm RNA elements as markers of health. Syst Biol Reprod Med. 2018;64(1):25–38. https://doi.org/10.1080/19396368.2017.1393583.
7. Alves MBR, Celeghini ECC, Belleannee C. From sperm motility to sperm-borne microRNA signatures: new approaches to predict male fertility potential. Front Cell Dev Biol. 2020;8:791. https://doi.org/10.3389/fcell.2020.00791.
8. Montjean D, De La Grange P, Gentien D, Rapinat A, Belloc S, Cohen-Bacrie P, et al. Sperm transcriptome profiling in oligozoospermia. J Assist Reprod Genet. 2012;29(1):3–10. https://doi.org/10.1007/s10815-011-9644-3.
9. Jodar M, Sendler E, Moskovtsev SI, Librach CL, Goodrich R, Swanson S, et al. Absence of sperm RNA elements correlates with idiopathic male infertility. Sci Transl Med. 2015;7(295):295re6.
10. Pereira N, Cheung S, Parrella A, O'Neill C, Nikprelevic N, Rosenwaks Z, et al. Investigating the role of sperm-specific RNA to screen men with unexplained infertility. Fertil Steril. 2017;108(3):e46. https://doi.org/10.1016/j.fertnstert.2017.07.151.
11. Zhang Y, Shi J, Rassoulzadegan M, Tuorto F, Chen Q. Sperm RNA code programmes the metabolic health of offspring. Nat Rev Endocrinol. 2019;15(8):489–98. https://doi.org/10.1038/s41574-019-0226-2.
12. Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA, Network RM. The presence, role and clinical use of spermatozoal RNAs. Hum Reprod Update. 2013;19(6):604–24. https://doi.org/10.1093/humupd/dmt031.
13. Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature. 2004;429(6988):154. https://doi.org/10.1038/429154a.
14. Bedi Y, Chang RC, Gibbs R, Clement TM, Golding MC. Alterations in sperm-inherited noncoding RNAs associate with late-term fetal growth restriction induced by preconception paternal alcohol use. Reprod Toxicol. 2019;87:11–20. https://doi.org/10.1016/j.reprotox.2019.04.006.
15. Hughes V. Sperm RNA carries marks of trauma. Nature. 2014;508(7496):296–7. https://doi.org/10.1038/508296a.
16. Sarker G, Sun W, Rosenkranz D, Pelczar P, Opitz L, Efthymiou V, et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. Proc Natl Acad Sci U S A. 2019;116(21):10547–56. https://doi.org/10.1073/pnas.1820810116.
17. Gapp K, van Steenwyk G, Germain PL, Matsushima W, Rudolph KLM, Manuella F, et al. Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. Mol Psychiatry. 2020;25(9):2162–74. https://doi.org/10.1038/s41380-018-0271-6.
18. Sendler E, Johnson GD, Mao S, Goodrich RJ, Diamond MP, Hauser R, et al. Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. 2013;41(7):4104–17. https://doi.org/10.1093/nar/gkt132.
19. Johnson GD, Lalancette C, Linnemann AK, Leduc F, Boissonneault G, Krawetz SA. The sperm nucleus: chromatin, RNA, and the nuclear matrix. Reproduction. 2011;141(1):21–36. https://doi.org/10.1530/REP-10-0322.
20. Zhang X, Zhang P, Song D, Xiong S, Zhang H, Fu J, Gao F, Chen H, Zeng X. Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm. Biol Reprod. 2019;100(4):982–93. https://doi.org/10.1093/biolre/ioy253.
21. Alvarez-Rodriguez M, Martinez C, Wright D, Barranco I, Roca J, Rodriguez-Martinez H, et al. The transcriptome of pig spermatozoa, and its role in fertility. Int J Mol Sci. 2020;21(5):1572. https://doi.org/10.3390/ijms21051572.
22. Selvaraju S, Parthipan S, Somashekar L, Kolte AP, Krishnan Binsila B, Arangasamy A, et al. Occurrence and functional significance of the transcriptome in bovine (Bos taurus) spermatozoa. Sci Rep. 2017;7(1):42392. https://doi.org/10.1038/srep42392.
23. Hodge MJ, de Las H-SS, Rindfleish SJ, Stephen CP, Pant SD. Characterization of breed specific differences in Spermatozoal transcriptomes of sheep in Australia. Genes. 2021;12(2):203. https://doi.org/10.3390/genes12020203.
24. Suliman Y, Becker F, Wimmers K. Implication of transcriptome profiling of spermatozoa for stallion fertility. Reprod Fertil Dev. 2018;30(8):1087–98. https://doi.org/10.1071/RD17188.
25. Tumbleson ME, Schook LB. Advances in swine in biomedical research, vol. 1 and 2. New York: Plenum Press; 1996. p. 1–4.
26. Fishman JA, Patience C. Xenotransplantation: infectious risk revisited. Am J Transplant. 2004;4(9):1383–90. https://doi.org/10.1111/j.1600-6143.2004.00542.x.
27. Humphray SJ, Scott CE, Clark R, Marron B, Bender C, Camm N, et al. A high utility integrated map of the pig genome. Genome Biol. 2007;8(7):R139. https://doi.org/10.1186/gb-2007-8-7-r139.
28. Broekhuijse ML, Feitsma H, Gadella BM. Artificial insemination in pigs: predicting male fertility. Vet Q. 2012;32(3–4):151–7. https://doi.org/10.1080/01652176.2012.735126.
29. Evans JP. Sperm-egg interaction. Annu Rev Physiol. 2012;74(1):477–502. https://doi.org/10.1146/annurev-physiol-020911-153339.

30. Sawada H, Inoue N, Iwano M. Sexual reproduction in animals and plants. Tokyo: Springer Tokyo; 2014. https://doi.org/10.1007/978-4-431-54589-7.
31. Knobil E, Neill JD. The physiology of reproduction. 2nd ed. New York: Raven Press; 1994.
32. Almond G. The Swine AI Book: A Field and Laboratory Technicians' Guide to Artificial Insemination in Swine. New York: Morgan Morrow; 1998.
33. Harrison RA, Jacques ML, Minguez ML, Miller NG. Behaviour of ejaculated spermatozoa from bull, boar and ram during thin-layer countercurrent partition in aqueous two-phase systems. J Cell Sci. 1992;102(Pt 1):123–32. https://doi.org/10.1242/jcs.102.1.123.
34. Kwon WS, Rahman MS, Lee JS, Kim J, Yoon SJ, Park YJ, et al. A comprehensive proteomic approach to identifying capacitation related proteins in boar spermatozoa. BMC Genomics. 2014;15(1):897. https://doi.org/10.1186/1471-2164-15-897.
35. Kwon WS, Rahman MS, Ryu DY, Park YJ, Pang MG. Increased male fertility using fertility-related biomarkers. Sci Rep. 2015;5(1):1–11. https://doi.org/10.1038/srep15654.
36. Kwon WS, Park YJ, Kim YH, You YA, Kim IC, Pang MG. Vasopressin effectively suppresses male fertility. PLoS One. 2013;8(1):e54192. https://doi.org/10.1371/journal.pone.0054192.
37. Perez L, Valcarcel A, de Las HM, Moses D, Baldassarre H. *In vitro* capacitation and induction of acrosomal exocytosis in ram spermatozoa as assessed by the chlortetracycline assay. Theriogenology. 1996;45(5):1037–46. https://doi.org/10.1016/0093-691X(96)00031-3.
38. Pang WK, Kang S, Ryu DY, Rahman MS, Park YJ, Pang MG. Optimization of sperm RNA processing for developmental research. Sci Rep. 2020;10(1):11606. https://doi.org/10.1038/s41598-020-68486-1.
39. Dube C, Beaulieu M, Reyes-Moreno C, Guillemette C, Bailey JL. Boar sperm storage capacity of BTS and Androhep plus: viability, motility, capacitation, and tyrosine phosphorylation. Theriogenology. 2004;62(5):874–86. https://doi.org/10.1016/j.theriogenology.2003.12.006.
40. Oh S, Park Y, You Y, Mohamed E, Pang M. Capacitation status of stored boar spermatozoa is related to litter size of sows. Anim Reprod Sci. 2010;121(1–2):131–8. https://doi.org/10.1016/j.anireprosci.2010.05.019.
41. Raveney BJE, Sato W, Takewaki D, Zhang C, Kanazawa T, Lin Y, et al. Involvement of cytotoxic Eomes-expressing CD4(+) T cells in secondary progressive multiple sclerosis. Proc Natl Acad Sci U S A. 2021;118(11):e2021818118. https://doi.org/10.1073/pnas.2021818118.
42. Cooke HJ, Saunders PT. Mouse models of male infertility. Nat Rev Genet. 2002;3(10):790–801. https://doi.org/10.1038/nrg911.
43. Ehmcke J, Schlatt S. Animal models for fertility preservation in the male. Reproduction. 2008;136(6):717–23. https://doi.org/10.1530/REP-08-0093.
44. Alberio R, Wolf E. 25th ANNIVERSARY OF CLONING BY SOMATIC-CELL NUCLEAR TRANSFER: Nuclear transfer and the development of genetically modified/gene edited livestock. Reproduction. 2021;162(1):F59–68. https://doi.org/10.1530/REP-21-0078.
45. Yue Y, Xu W, Kan Y, Zhao HY, Zhou Y, Song X, et al. Extensive germline genome engineering in pigs. Nat Biomed Eng. 2021;5(2):134–43. https://doi.org/10.1038/s41551-020-00613-9.
46. Ryu DY, Pang WK, Rahman MS, Park YJ, Pang MG. Peroxiredoxin 4 directly affects the male fertility outcome in porcine. Theriogenology. 2021;171:85–93. https://doi.org/10.1016/j.theriogenology.2021.05.020.
47. Kang S, Pang WK, Ryu DY, Song WH, Rahman MS, Park YJ, et al. Porcine seminal protein-I and II mRNA expression in boar spermatozoa is significantly correlated with fertility. Theriogenology. 2019;138:31–8. https://doi.org/10.1016/j.theriogenology.2019.06.043.
48. Kim KU, Pang WK, Kang S, Ryu DY, Song WH, Rahman MS, et al. Sperm solute carrier family 9 regulator 1 is correlated with boar fertility. Theriogenology. 2019;126:254–60. https://doi.org/10.1016/j.theriogenology.2018.12.023.
49. Xu T, Fang Y, Rong A, Wang J. Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. BMC Med Res Methodol. 2015;15(1):94. https://doi.org/10.1186/s12874-015-0085-z.
50. Park YJ, Pang WK, Ryu DY, Song WH, Rahman MS, Pang MG. Optimized combination of multiple biomarkers to improve diagnostic accuracy in male fertility. Theriogenology. 2019;139:106–12. https://doi.org/10.1016/j.theriogenology.2019.07.029.
51. Gupta SK. Human zona pellucida glycoproteins: Binding characteristics with human spermatozoa and induction of acrosome reaction. Front Cell Dev Biol. 2021;9:619868. https://doi.org/10.3389/fcell.2021.619868.
52. Gupta S, Bansal P, Ganguly A, Bhandari B. Delineation of the role of zona pellucida glycoproteins during fertilization in humans. J Reprod Immunol. 2009;2(81):117–8. https://doi.org/10.1016/j.jri.2009.06.137.
53. Schuster A, Tang C, Xie Y, Ortogero N, Yuan S, Yan W. SpermBase: a database for sperm-borne RNA contents. Biol Reprod. 2016;95(5):99. https://doi.org/10.1095/biolreprod.116.142190.
54. Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, Lukowitz W. Paternal control of embryonic patterning in Arabidopsis thaliana. Science. 2009;323(5920):1485–8. https://doi.org/10.1126/science.1167784.
55. Ito C, Yamatoya K, Yoshida K, Fujimura L, Sugiyama H, Suganami A, et al. Deletion of Eqtn in mice reduces male fertility and sperm-egg adhesion. Reproduction. 2018;156(6):579–90. https://doi.org/10.1530/REP-18-0394.
56. Hao J, Chen M, Ji S, Wang X, Wang Y, Huang X, et al. *Equatorin* is not essential for acrosome biogenesis but is required for the acrosome reaction. Biochem Bioph Res Co. 2014;444(4):537–42. https://doi.org/10.1016/j.bbrc.2014.01.080.
57. Toshimori K, Saxena D, Tanii I, Yoshinaga K. An MN9 antigenic molecule, equatorin, is required for successful sperm-oocyte fusion in mice. Biol Reprod. 1998;59(1):22–9. https://doi.org/10.1095/biolreprod59.1.22.
58. Nagdas SK, Smith L, Medina-Ortiz I, Hernandez-Encarnacion L, Raychoudhury S. Identification of bovine sperm acrosomal proteins that interact with a 32-kDa acrosomal matrix protein. Mol Cell Biochem. 2016;414(1–2):153–69. https://doi.org/10.1007/s11010-016-2668-3.
59. Mandal A, Klotz KL, Shetty J, Jayes FL, Wolkowicz MJ, Bolling LC, et al. SLLP1, a unique, intra-acrosomal, non-bacteriolytic, c lysozyme-like protein of human spermatozoa. Biol Reprod. 2003;68(5):1525–37. https://doi.org/10.1095/biolreprod.102.010108.
60. Ulrich AS, Otter M, Glabe CG, Hoekstra D. Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. Biol Chem. 1998;273(27):16748–55. https://doi.org/10.1074/jbc.273.27.16748.
61. Gur Y, Breitbart H. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. Genes Dev. 2006;20(4):411–6. https://doi.org/10.1101/gad.367606.
62. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mRNA abundance. Cell. 2016;65(3):535–50. https://doi.org/10.1016/j.cell.2016.03.014.
63. Freydanck MK, Laubender RP, Rack B, Schuhmacher L, Jeschke U, Scholz C. Two-marker combinations for preoperative discrimination of benign and malignant ovarian masses. Anticancer Res. 2012;32(5):2003–8.
64. Chan MK, Krebs M, Cox D, Guest P, Yolken RH, Rahmoune H, et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. Transl Psychiatry. 2015;5(7):e601. https://doi.org/10.1038/tp.2015.91.
65. Zhang P, West NP, Chen PY, Thang MWC, Price G, Cripps AW, et al. Selection of microbial biomarkers with genetic algorithm and principal component analysis. BMC Bioinformatics. 2019;20(Suppl 6):413. https://doi.org/10.1186/s12859-019-3001-4.
66. Taguchi YH, Murakami Y. Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers. PLoS One. 2013;8(6):e66714. https://doi.org/10.1371/journal.pone.0066714.
67. Sell SL, Widen SG, Prough DS, Hellmich HL. Principal component analysis of blood microRNA datasets facilitates diagnosis of diverse diseases. PLoS One. 2020;15(6):e0234185. https://doi.org/10.1371/journal.pone.0234185.
68. Fortino V, Wisgrill L, Werner P, Suomela S, Linder N, Jalonen E, et al. Machine-learning–driven biomarker discovery for the discrimination between allergic and irritant contact dermatitis. Proc Nat Acad Sci U S A. 2020;117(52):33474–85. https://doi.org/10.1073/pnas.2009192117.