

---

This is the **accepted version** of the article:

Ablondi, Michela; Gòdia, Marta; Rodríguez Gil, Joan Enric; [et al.]. «Characterisation of sperm piRNAs and their correlation with semen quality traits in swine». *Animal Genetics*, Vol. 52, Issue 1 (February 2021), p. 114-120. DOI 10.1111/age.13022

---

This version is available at <https://ddd.uab.cat/record/236258>

under the terms of the  <sup>IN</sup> COPYRIGHT license

**Title:**

Characterisation of sperm piRNAs and their correlation with semen quality traits in swine

**Running Title:**

piRNAs in porcine sperm

**Authors with affiliations:**

M. Ablondi<sup>1\*</sup>, M. Gòdia<sup>2\*</sup>, J. E. Rodriguez-Gil<sup>3</sup>, A. Sánchez<sup>2,4</sup>, A. Clop<sup>2,5†</sup>

<sup>1</sup> Department of Veterinary Science, University of Parma, 43126 Parma, Italy

<sup>2</sup> Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193, Cerdanyola del Vallès, Catalonia, Spain

<sup>3</sup> Department of Animal Medicine and Surgery, School of Veterinary Sciences, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès, Catalonia, Spain

<sup>4</sup> Departament de Ciència Animal i dels Aliments, School of Veterinary Sciences, Universitat Autònoma de Barcelona, 08193, Cerdanyola del Vallès, Catalonia, Spain

<sup>5</sup> Consejo Superior de Investigaciones Científicas (CSIC), 08003, Barcelona, Catalonia, Spain

\*These authors contributed equally to this work.

†Corresponding author:

[alex.clop@cragenomica.es](mailto:alex.clop@cragenomica.es); Tel. +34 935636600 ext 3353

**Summary (max 250 words)**

Piwi-interacting RNAs (piRNAs) are a class of non-coding RNAs that are essential in the transcriptional silencing of transposable elements and warrant genome stability in mammalian germline. In this study we have identified piRNAs in porcine sperm using male germline and zygote datasets from human, mice, cow and pig, and evaluated the relation between their abundances and sperm quality traits. In our analysis, we identified 283,382 piRNAs, 1,355 of which correlated with  $P \leq 0.01$  to at least one semen quality trait. Fifty-seven percent of the correlated piRNAs mapped less than 50kb apart from any other piRNA in the pig genome. Furthermore, piRNA location was significantly enriched near Long Interspersed Nuclear Elements (LINEs). Moreover, some of the significant piRNAs mapped within or close to genes relevant for fertility or spermatogenesis such as *CSNK1G2* and *PSMF1*.

**Key words: (3 to 10)**

Piwi-interacting RNAs, Repeat Elements, RNA-seq, swine, sperm, semen quality

## **Main text**

Fertility is of paramount importance in livestock production as it influences its efficiency and thus the sustainability of the sector. Pig production highly depends on the genetic merit of elite boars used for artificial insemination (AI) and on their sperm quality, which is ultimately used to spread the pigs' genetic material (Zak *et al.* 2017). Several independent studies have shown the functional relevance of small non-coding RNAs (sncRNAs) on the molecular mechanisms behind spermatogenesis, testis development (He *et al.* 2009; Gebert *et al.* 2015; Goh *et al.* 2015) and fertility (Salas-Huetos *et al.* 2015). Piwi-interacting RNAs (piRNAs) are a class of sncRNAs that are highly abundant in sperm and that have a size that ranges between 26 and 32 nucleotides long with slight differences reported between studies (Aravin *et al.* 2006; Gebert *et al.*, 2015). piRNAs are crucial for transposon silencing and genome stability during spermatogenesis (Aravin *et al.* 2008; Yan 2017). Research in mutant Piwi proteins has confirmed their role in spermatogenesis and fertility (Fu and Wang 2014; Gòdia *et al.* 2018a). Few studies have evaluated the relationship between piRNA abundance and semen quality in human (Cui *et al.* 2018,) and cattle (Capra *et al.* 2017) but this remains unchecked in pigs. To shed light into the relevance of

piRNAs on sperm quality in swine, we have characterised the piRNA composition of the porcine sperm and evaluated the correlation between piRNA abundance and semen quality parameters.

Thirty-four fresh sperm ejaculates, each from a different boar, were obtained by specialised professionals and spermatozoa were directly purified from the ejaculate by density gradient centrifugation (Gòdia *et al.* 2018b). RNA extraction and short RNA-seq analysis were performed as described by Gòdia *et al.* (2018b) and Gòdia *et al.* (2019). The fastq files can be found at the SRA (accession number PRJNA520978). This dataset has been previously analysed under the scope of other studies (Gòdia *et al.* 2020a; Gòdia *et al.* 2020b) that allowed us quantifying that 15.6% of this short RNA fraction corresponded to putative piRNAs. We also detected other RNA types including mitochondrial transfer RNAs (46.9%), nuclear transfer RNAs (9.4%), micro-RNAs (8.6%), small nuclear RNAs (8.1%), mitochondrial ribosomal RNAs (8.1%), nuclear ribosomal RNAs (3.1%) and small nucleolar RNAs (0.1%) (Table S1).

Raw reads quality was evaluated with FastQC 0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Low-quality reads, sequence adaptors and reads with length  $< 26$  and  $> 32$  bp were removed with Cutadapt 2.8 (Martin 2011). Filtered reads were then mapped to the pig genome (Sscrofa11.1) with Bowtie 1 (Langmead 2010), using the “--best--strata” option and allowing for 1 mismatch. Previously annotated piRNAs in relevant tissues (sperm, testis and zygote) in *Sus scrofa*, *Bos taurus*, *Homo sapiens* and *Mus musculus* were retrieved from the piRBase database v.2.0 (Wang *et al.* 2019). piRNA abundance was quantified using FeatureCounts v.1.6.4 (Liao *et al.* 2014) and read counts were normalised by sequencing depth as counts per million (CPM). Only piRNAs with CPM  $> 1$  in at least 75% (N = 26) of the samples were kept. piRNA abundances were then stabilised with a log<sub>2</sub> transformation. Thirteen semen quality parameters were assessed and corrected for the farm, year and season and age effects (Table S2) as detailed by Gòdia *et al.* (2020a) and Gòdia *et al.* (2020b). Pearson correlations between piRNA abundances and these semen quality parameters were calculated and only suggestive correlations with  $P \leq 0.01$  were considered for further analysis. We also applied an FDR correction for multiple testing with q-value  $\leq 0.10$ . Statistical overrepresentation of biological processes of genes mapping within a  $\pm 2.5$ kb window from significant piRNAs was

conducted with PANTHER 14.0 (<http://pantherdb.org>) using Human as reference and by setting the level of significance to  $q\text{-value} \leq 0.01$ . There are some technical aspects that should be mentioned. First, the piRNA size range between 21 and 32 nucleotide with slight variations between studies. One of the landmark articles on piRNA research provided a size range of 26-31 nucleotide long (Aravin *et al.* 2006). Also, research in porcine testis showed that about 97% of the piRNAs fall within the 26-32 nucleotide range (Gebert *et al.*, 2015). Moreover, microRNAs range between 18 and 24 nucleotides long (Fang *et al.* 2013). We chose a minimum piRNA size of 26 nucleotides as this should include most piRNAs while exclude at the same time, miRNAs. Second, we based our analysis on previously identified piRNAs in swine, human, mouse and cattle as piRNAs identified by independent studies and samples are more likely to be real. Human, mouse and cattle were included under the assumption that functional elements of the genome tend to be evolutionarily conserved across species (Hardison, 2003). Third, although there are some tools and databases to search for messenger RNA targets of piRNAs (Wu *et al.* 2018; Wu *et al.* 2019), these are species-specific, or they are still subjected to limited specificity and sensitivity. For this reason, we



did not search for piRNA targets in our piRNA dataset.

On average, 2.8 M reads passed the quality control and length filters and were used for piRNA characterisation. Overall, 283,382 piRNAs were successfully mapped to the swine genome (Table S3) and 6,729 piRNAs had CPM > 1 in more than 75% of the samples. Two piRNA biogenesis mechanisms have been proposed in animals. These are the primary processing resulting in piRNAs with a 5' uracil enrichment signature and the so-called ping pong amplification loop which produce piRNAs with an adenine at the tenth position (Fu and Wang 2014). 51.2% of the piRNAs in pig sperm started with an uracil (Figure S1) but no preference for adenine at position 10 was observed (Figure S1), thus indicating a preponderance for the primary processing pathway (Fu and Wang, 2014). These results are similar to the previously reported results in swine testes (Gebert *et al.* 2015), with a clear enrichment of an uracil in position 1 and no predominance of adenine at position 10. In the spermatogenic lineage, the primary processing is involved specifically in the biogenesis of the so-called pachytene piRNAs, which are thought to play a role in transposon silencing and spermiogenesis (Fu and Wang, 2014).

A total of 1,355 piRNAs showed correlation between their abundance and at least one phenotype at  $P \leq 0.01$  (Table S4). These correlations ranged between -0.72 and 0.64 and the number of associated piRNAs varied considerably across phenotypes and chromosomes (Figure S2; Figure 1a). Two piRNA : phenotype correlations reached a q-value  $\leq 0.10$  (Table S4). These were piR-ssc-1201188 (q-value = 0.08) and piR-ssc-120840 (q-value=0.008). piR-ssc-1201188, located within the *DAB2IP* gene, showed negative correlation with tail abnormalities (TABN), neck abnormalities (NABN) and proximal droplets (PDROP) with corr. = -0.67, -0.50 and -0.48, respectively. *DAB2IP* is involved in multiple processes including cell cycle checkpoint control but no direct link with the piRNA pathway has been suggested. piR-ssc-120840, located within the *EIF4H* gene, showed negative correlation with PDROP (corr.=-0.72). *EIF4H* is expressed in the mouse testis and knock-out mice exhibited reduced fertility (Capossela *et al.*, 2012).

The whole catalogue of piRNAs associated to semen quality with significance level of  $P \leq 0.01$  mapped, on average, less than 500 bp away from a Repetitive Element (RE). Nearly 34%, 25.3% and 21.3% of these piRNAs were less than 500 bp from Short

Interspersed Nuclear Elements (SINEs), simple repeats and Long Interspersed Nuclear Elements (LINEs), respectively (Table S5). Interestingly, the piRNAs associated to semen quality traits tended to be closer to LINEs (Wilcoxon rank sum test,  $P = 0.007$ ) compared to the list of piRNAs not correlated with semen quality (Figure 1b). These findings go in line with the potential role of piRNAs in transposon silencing and germline genome stability as reported in several mammalian species (O'Donnell and Boeke 2007) (Figure S3). Di Giacomo *et al.* (2013) demonstrated that Piwi proteins and the piRNA pathway play a role in the silencing of LINE1 during spermatogenesis. Moreover, several independent studies have reported associations between the disruption of LINE silencing and spermatogenesis aberrations (reviewed by Gòdia *et al.* 2018a). Previous studies have shown that piRNAs tend to cluster in the genome (Brennecke *et al.* 2007; Capra *et al.* 2017). We evaluated the distance between contiguous piRNAs showing significant phenotypic associations and identified that 57% of these were less than 50kb apart (Figure 1c). Five-hundred twenty-seven of the 1,355 significant piRNAs gathered within 50 clusters with a mean length of 18kb and an average of 10.5 piRNAs per cluster. These results are in agreement with the data presented by Liu *et al.* (2012) in

porcine testes and Li *et al.* (2019) in the testes of horse.

The piRNA : trait correlations were also analysed with a hierarchical clustering algorithm. In general, semen quality correlations showed that the abundance of individual piRNAs has a beneficial impact on semen quality (Figure 1d). Noteworthy, this trend was also found when considering the total piRNA abundance, defined as the sum of the RNA abundance of all individual piRNAs, and the semen quality parameters (e.g. tail abnormalities -TABN- with corr.= -0.32 and head abnormalities -HABN-, with corr.= -0.31, both with  $P = 0.07$ ), although none of the correlations reached significance levels (Table S6). Overall, these results indicate that RE silencing and genome stability impacts on spermatogenesis and semen quality as previously suggested (Ernst *et al.* 2017; Ge *et al.* 2017).

A total of 541 protein-coding genes were located within a 5kb region centred at each piRNA displaying significant correlation with semen traits. Gene ontology analysis showed enrichment for 11 biological processes, including reproduction or positive regulation of cell projection organisation (Table S7). Some of these genes, *CATSPER2*, *CATSPERG*, *OAZ3*, *ODF1*, *ODF2*, *PRM1*, *TEX14*, *TSSK2*, *TSSK3* and *TSSK6* are known to play key roles in spermatogenesis, sperm hyperactivation, sperm morphology, and

male fertility (Tokuhiro *et al.* 2009; Wacławska and Kurpisz 2012; Yang *et al.* 2012; Zhang *et al.* 2012; Savadi-Shiraz *et al.* 2015; Cai *et al.* 2017; Sun *et al.* 2017).

Eight of the 10 piRNAs most significantly correlated with semen quality traits, presented negative correlation values (Table 1a). Once more, this suggests that a reduction of piRNA abundance leads to a decrease of semen quality. This could be due to a deficient silencing of REs and the consequent increase of genome instability which could impact on germ cell development (Di Giacomo *et al.* 2013; Gunes and Kulac 2014). Some of these 10 piRNAs are located near genes with functions related to spermatogenesis. A positive correlation was found between piR-ssc-112967 and the Osmotic Resistance Test trait (ORT) (corr. = 0.62). This piRNA maps within the *CSNK1G2* gene. *CSNK1G2* has been linked to sperm surface modifications, sperm maturation and sperm–egg communication in bull sperm (Byrne *et al.* 2012). piR-ssc-113649, which correlated with the percentage of viable cells (VIAB), abnormal acrosomes (ACRO) and TABN (corr. = 0.64, -0.51 and -0.50, respectively), maps within *PSMF1*, a gene suggested to be a key regulator of spermatogenesis in humans (Cui *et al.* 2008).

The majority of these piRNAs were significantly correlated to only

one phenotype (75.7%). On the other side, the abundance of 8 piRNAs correlated with 4 phenotypes each (Table 1b). Among these, piR-bta-16647898 and piR-ssc-109407 were of great interest. piR-bta-16647898 showed negative correlation with average curvilinear velocity (VCL), average path velocity (VAP), PDROP and distal droplets (DDROP) and it is located within the *ACE* gene (Table 1b). *ACE* regulates sperm development and functions in human, mice and bull (Fuchs *et al.* 2005; Fujihara *et al.* 2013; Ojaghi *et al.* 2018). piR-ssc-109407, which was positively correlated with VIAB and concentration (CON) and negatively correlated with TABN and PDROP, is located within the *SILI* gene. Interestingly, previous research by another group found that the levels of *SILI* protein in seminal plasma were negatively correlated with fertility in bulls (Viana *et al.* 2018).

In conclusion, we characterised the piRNA composition of the porcine sperm and identified several piRNAs that could impact semen quality traits. These results indicate that piRNAs may hold potential as markers of sperm quality in swine.

[Figure 1 here]

## **Funding**

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) under grant AGL2013-44978-R and grant AGL2017-86946-R and by the CERCA Programme/Generalitat de Catalunya. AGL2017-86946-R was also funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERDF). We thank the Agency for Management of University and Research Grants (AGAUR) of the Generalitat de Catalunya (Grant Numbers 2014 SGR 1528 and 2017 SGR 1060). We acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the Center of Excellence Severo Ochoa 2016–2019 (Grant Number SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG). MA acknowledges a Ph.D. scholarship from the Department of Veterinary science, University of Parma. MG acknowledges a Ph.D. studentship from MINECO (Grant Number BES-2014-070560).

## **Availability of data**

The datasets generated and analysed are available at NCBI's BioProject PRJNA520978.

## Competing interests

The authors declare that they have no competing interests.

## Bibliography

Aravin, A.A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207.

Aravin, A.A., Sachidanandam, R., Bourchis, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., Hannon, G.J. (2008) A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell* **31**, 785–799.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., Hannon, G.J. (2007) Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089–1103.

Byrne, K., Leahy, T., Mcculloch, R., Colgrave, M.L., Holland, M.K. (2012) Comprehensive mapping of the bull sperm surface proteome. *Proteomics* **12**, 3559–3579.

Cai, X., Yu, S., Mipam, T.D., Yang, F., Zhao, W., Liu, W., Cao,



S.Z., Shen, L., Zhao, F., Sun, L., Xu, C., Wu, S. (2017) Comparative analysis of testis transcriptomes associated with male infertility in cattle yak. *Theriogenology* **88**, 28–42.

Capossela, S., Muzio, L., Bertolo, A., Bianchi, V., Dati, G., Chaabane, L., Godi, C., Politi, L.S., Biffo, S., D'Adamo, P., et al. (2012) Growth Defects and Impaired Cognitive–Behavioral Abilities in Mice with Knockout for Eif4h, a Gene Located in the Mouse Homolog of the Williams-Beuren Syndrome Critical Region. *The American Journal of Pathology* **180**, 1121–1135.

Capra, E., Turri, F., Lazzari, B., Cremonesi, P., Gliozzi, T.M., Fojadelli, I., Stella, A., Pizzi, F. (2017) Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between High- and Low-motile sperm populations. *BMC Genomics* **18**, 1–12.

Cui, Y., Zhu, H., Zhu, Y., Guo, X., Huo, R., Wang, X., Tong, J., Qian, L., Zhou, Z., Jia, Y., Lue, Y.H., Hikim, A.S., Wang, C., Swerdloff, R.S., Sha, J. (2008) Proteomic analysis of testis biopsies in men treated with injectable testosterone undecanoate alone or in combination with oral levonorgestrel as potential male contraceptive. *Journal of Proteome Research* **7**, 3984–3993.

Cui, L., Fang, L., Shi, B., Qiu, S., Ye, Y. (2018) Spermatozoa

Expression of piR-31704, piR-39888, and piR-40349 and Their Correlation to Sperm Concentration and Fertilization Rate After ICSI. *Reproductive Sciences* **25**, 733-739.

Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A.J., OCarroll, D. (2013) Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. *Molecular Cell* **50**, 601–608.

Ernst, C., Odom, D.T., Kutter, C. (2017) The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nature Communications* **8**, 1–9.

Fang, Z., Du, R., Edwards, A., Flemington, E.K., Zhang, K. (2013) The Sequence Structures of Human MicroRNA Molecules and Their Implications. *PLoS ONE*, **8**, e54215.

Fu, Q., Wang, P.J. (2014) Mammalian piRNAs: Biogenesis, function, and mysteries. *Spermatogenesis* **4**, e27889.

Fuchs, S., Frenzel, K., Hubert, C., Lyng, R., Muller, L., Michaud, A., Xiao, H.D., Adams, J.W., Capecchi, M.R., Corvol, P., Shur, B.D., Bernstein, K.E. (2005) Male fertility is dependent on dipeptidase activity of testis ACE. *Nature Medicine* **11**, 1140–1142.

Fujihara, Y., Tokuhira, K., Muro, Y., Kondoh, G., Araki, Y., Ikawa,

M., Okabe, M. (2013) Expression of TEX101, regulated by ACE, is essential for the production of fertile mouse spermatozoa. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8111–8116.

Ge, S.Q., Lin, S.L., Zhao, Z.H., Sun, Q.Y. (2017) Epigenetic dynamics and interplay during spermatogenesis and embryogenesis: Implications for male fertility and offspring health. *Oncotarget* **8**, 53804–53818.

Gebert, D., Ketting, R.F., Zischler, H., Rosenkranz, D. (2015) piRNAs from pig testis provide evidence for a conserved role of the Piwi pathway in post-transcriptional gene regulation in mammals. *PLoS ONE* **10**, 1–22.

Gòdia, M., Castelló, A., Rocco, M. et al. (2020a) Identification of circular RNAs in porcine sperm and evaluation of their relation to sperm motility. *Scientific Reports* **10**, 7985.

Gòdia M., Reverter A., González-Prendes R., Ramayo-Caldas Y., Castelló A., Rodríguez-Gil J.E., Sanchez A., Clop A. (2020b) An integrative systems biology approach to identify the molecular basis of sperm quality in swine. Available at Research Square. doi: 10.21203/rs.3.rs-19366/v1.

Gòdia, M., Estill, M., Castelló, A., Balasch, S., Rodríguez-Gil, J.E.,

- Krawetz, S.A., Sánchez, A., Clop, A. (2019) A RNA-seq analysis to describe the boar sperm transcriptome and its seasonal changes. *Frontiers in Genetics* **10**, 1–14.
- Gòdia, M., Swanson, G., Krawetz, S.A. (2018a) A history of why fathers RNA matters. *Biology of Reproduction* **99**, 147–159.
- Gòdia, M., Mayer, F.Q., Nafissi, J., Castelló, A., Rodríguez-Gil, J.E., Sánchez, A., Clop, A. (2018b) A technical assessment of the swine ejaculated spermatozoa for a sperm-specific RNA-seq analysis. *Systems Biology in Reproductive Medicine* **64**, 291–303.
- Goh, W.S.S., Falciatori, I., Tam, O.H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M., Hannon, G.J. (2015) PiRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes and Development* **29**, 1032–1044.
- Gunes, S., Kulac, T. (2014) The role of epigenetics in spermatogenesis. *Türk Üroloji Dergisi/Turkish Journal of Urology* **39**, 181–187.
- Hardison, R.C. (2003) Comparative Genomics. *PLoS Biology*, **1**, e58.
- He, Z., Kokkinaki, M., Pant, D., Gallicano, G.I., Dym, M. (2009) Small RNA molecules in the regulation of spermatogenesis. *Reproduction* **137**, 901–911.

Langmead, B. (2010) Aligning Short Sequencing Reads with Bowtie. *Current Protocols in Bioinformatics* **32**, 1–24.

Li, B., He, X., Zhao, Y., Bai, D., Bou, G., Zhang, X., Su, S., Dao, L., Liu, R., Wang, Y., Manglai, D. (2019) Identification of piRNAs and piRNA clusters in the testes of the Mongolian horse. *Scientific Reports* **9**, 1–9.

Liao, Y., Smyth, G.K., Shi, W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

Liu, G., Lei, B., Li, Y., Tong, K., Ding, Y., Luo, L., Xia, X., Jiang, S., Deng, C., Xiong, Y., Li, F. (2012) Discovery of potential pirnas from next generation sequences of the sexually mature swine testes. *PLoS ONE* **7**, 1–9.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10.

ODonnell, K.A., Boeke, J.D. (2007) Mighty Piwis Defend the Germline against Genome Intruders. *Cell* **129**, 37–44.

Ojaghi, M., Kastelic, J., Thundathil, J.C. (2018) Testis-specific isoform of angiotensin-converting enzyme (tACE) as a candidate marker for bull fertility. *Reproduction, Fertility and Development* **30**, 1584.

Rosenkranz, D., Han, C.T., Roovers, E.F., Zischler, H., Ketting, R.F. (2015) Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. *Genomics Data* **5**, 309–313.

Salas-Huetos, A., Blanco, J., Vidal, F., Godo, A., Grossmann, M., Pons, M.C., F-Fernández, S., Garrido, N., Anton, E. (2015) Spermatozoa from patients with seminal alterations exhibit a differential micro-ribonucleic acid profile. *Fertility and Sterility* **104**, 591–601.

Savadi-Shiraz, E., Edalatkhah, H., Talebi, S., Heidari-Vala, H., Zandemami, M., Pahlavan, S., Modarressi, M.H., Akhondi, M.M., Paradowska-Dogan, A., Sadeghi, M.R. (2015) Quantification of sperm specific mRNA transcripts (PRM1, PRM2, and TNP2) in teratozoospermia and normozoospermia: New correlations between mRNA content and morphology of sperm. *Molecular Reproduction and Development* **82**, 26–35.

Sun, X. hong, Zhu, Y. ying, Wang, L., Liu, H. ling, Ling, Y., Li, Z. li, Sun, L. bo (2017) The Catsper channel and its roles in male fertility: A systematic review. *Reproductive Biology and Endocrinology* **15**, 1–12.

Tokuhiro, K., Isotani, A., Yokota, S., Yano, Y., Oshio, S., Hirose,

M., Wada, M., Fujita, K., Ogawa, Y., Okabe, M., Nishimune, Y., Tanaka, H. (2009) OAZ-t/OAZ3 is essential for rigid connection of sperm tails to heads in mouse. *PLoS Genetics* **5**, 1–7.

Viana, A.G.A., Martins, A.M.A., Pontes, A.H., Fontes, W., Castro, M.S., Ricart, C.A.O., Sousa, M. V., Kaya, A., Topper, E., Memili, E., Moura, A.A. (2018) Proteomic landscape of seminal plasma associated with dairy bull fertility. *Scientific Reports* **8**, 1–13.

Waclawska, A., Kurpisz, M. (2012) Key functional genes of spermatogenesis identified by microarray analysis. *Systems Biology in Reproductive Medicine* **58**, 229–235.

Wang, J., Zhang, P., Lu, Y., Li, Y., Zheng, Y., Kan, Y., Chen, R., He, S. (2019) PiRBase: A comprehensive database of piRNA sequences. *Nucleic Acids Research* **47**, 175–180.

Wu, W.S., Huang, W.C., Brown, J.S., Zhang, D., Song, X., Chen, H., Tu, S., Weng, Z., Lee, H.C. (2018) PirScan: A webserver to predict piRNA targeting sites and to avoid transgene silencing in *C. elegans*. *Nucleic Acids Research* **46**, 43–48.

Wu, W.S., Brown, J.S., Chen, T.-T., Chu, Y.-H., Huang, W.C., Tu, S., Lee, H.C. (2019) piRTarBase: a database of piRNA targeting sites and their roles in gene regulation. *Nucleic Acids Research* **47**, 181–187.

Yan, W. (2017) piRNA-independent PIWI function in spermatogenesis and male fertility. *Biology of Reproduction* **96**, 1121–1123.

Yang, K., Meinhardt, A., Zhang, B., Grzmil, P., Adham, I.M., Hoyer-Fender, S. (2012) The Small Heat Shock Protein ODF1/HSPB10 Is Essential for Tight Linkage of Sperm Head to Tail and Male Fertility in Mice. *Molecular and Cellular Biology* **32**, 216–225.

Zak, L.J., Gaustad, A.H., Bolarin, A., Broekhuijse, M.L.W.J., Walling, G.A., Knol, E.F. (2017) Genetic control of complex traits, with a focus on reproduction in pigs. *Molecular Reproduction and Development* **84**,1004–1011.

Zhang, Z., Wang, G.L., Li, H.X., Li, L., Cui, Q.W., Wei, C. Bin, Zhou, F. (2012) Regulation of fertilization in male rats by CatSper2 knockdown. *Asian Journal of Andrology* **14** 301–309.

## **Figure and table**

**Figure 1. Characteristics of the sperm piRNAs associated to semen quality.**

**a) Genome wide location of the piRNAs with suggestive association**





**Table 1.** List of piRNAs with most significant (a) or largest number of (b) correlations with semen quality parameters.

	piRNA	piRNA genomic location (chr:start-end)	Phenotype	Corr.	<i>P</i>	Adjusted <i>P</i> (FDR)	5kb genomic window (chr:start-end)	Genomic context	Genes within the 5kb window
<b>(a) piRNAs with most significant associations</b>	piR-ssc-120840	3:11,346,365-11,346,391	PDROP	-0.72	1.33E-06	0.008	3:11,343,865-11,348,891	exonic piRNA	<i>EIF4H</i>
	piR-ssc-1201188	1:261,796,921-261,796,949	TABN	-0.67	1.21E-05	0.081	1:261,794,421-261,799,449	intronic piRNA	<i>DAB2IP</i>
	piR-ssc-117286	5:4,191,151-4,191,180	PDROP	-0.64	4.01E-05	0.135	5:4,188,651-4,193,680	exonic piRNA	<i>KIAA0930</i>
	piR-ssc-127298*	2:76,577,902-76,577,934	TABN	-0.64	4.77E-05	0.140	2:76,575,402-76,580,434	intronic piRNA	<i>CSNK1G2</i>
	piR-ssc-114575	4:99,453,520-99,453,546	DDROP	-0.65	2.89E-05	0.194	4:99,451,020-99,456,046	exonic piRNA	<i>ANKRD35</i>
	piR-ssc-1272746	5:66,796,437-66,796,467	HABN	-0.63	5.45E-05	0.228	5:66,793,937-66,798,967	intronic piRNA	<i>PRMT8</i>

	piR-ssc-887100	2:38,231,083-38,231,114	DDROP	-0.62	7.94E-05	0.267	2:38,228,583-38,233,614	intronic piRNA	<i>NELLI</i>
	piR-ssc-113649	17:34,118,505-34,118,534	VIAB	0.64	4.91E-05	0.330	17:34,116,005-34,121,034	exonic piRNA	<i>PSMF1</i>
	piR-ssc-1133701	MT:8,958-8,990	VLC	-0.64	4.94E-05	0.333	MT:6,458-11,490	exonic piRNA	<i>ND3, ND4, ND4L, ND5, ND6, COX1, COX2, COX3, ATP6, ATP8, CYTB</i>
	piR-ssc-112967	2:5,745,472-5,745,503	ORT	0.62	9.69E-05	0.422	2:5,742,972-5,748,003	exonic piRNA	<i>RBM4B</i>
<b>(b)</b> <b>piRNAs</b>	piR-ssc-96498	2:59,064,952-59,064,981	HABN	-0.58	3.59E-04	0.228	2:59,062,452-59,067,481	exonic piRNA	<i>COMP</i>

		MOT	0.42	1.12E-02	1.000		
		VAP	0.44	8.37E-03	0.557		
		VCL	0.52	1.62E-03	0.545		
		TABN	-0.64	4.77E-05	0.140		
		PDROP	-0.46	6.41E-03	0.331		
piR-ssc- 127298*	2:76,577,902- 76,577,934					2:76,575,402- 76,580,434	intronic piRNA 2
		HABN	-0.43	1.01E-02	0.369		
		VIAB	0.43	1.01E-02	0.712		

		CON	0.45	8.02E-03	0.898			
piR-ssc-109407	2:140,888,219-140,888,247	TABN	-0.50	2.32E-03	0.252	2:140,885,719-140,890,747	exonic piRNA	<i>SILI</i>
		PDROP	-0.46	6.12E-03	0.331			
		VIAB	0.42	1.12E-02	0.712			
		ACRO	-0.51	2.25E-03	0.978			
piR-bta-16289189	4:42,395,010-42,395,037	CON	0.44	9.91E-03	0.898	4:42,392,510-42,397,537	intronic piRNA	<i>GEM</i>
		VCL	0.45	8.14E-03	0.596			

		VIAB	0.50	2.78E-03	0.573			
		ACRO	-0.42	1.08E-02	0.979			
		VAP	0.43	1.07E-02	0.557			
piR-ssc-121829	8:121,868,455-121,868,487					8:121,865,955-121,870,987	exonic piRNA	<i>RAP1GD</i> <i>SI</i>
		VIAB	0.43	1.04E-02	0.712			
		VSL	0.48	4.42E-03	0.754			
		VCL	-0.53	7.32E-03	0.545			
piR-bta-16647898	12:15,398,756-15,398,782					12:15,396,256-15,401,282	exonic piRNA	<i>ACE</i>
		PDROP	-0.51	1.85E-03	0.300			

		DDROP	-0.50	2.59E-03	0.378			
		VAP	-0.45	1.23E-03	0.557			
		TABN	-0.53	1.27E-03	0.221			
		NABN	-0.42	1.07E-02	1.000			
piR-ssc- 1211950	14:42,767,163- 42,767,192					14:42,764,663- 42,769,692	intergenic piRNA	<i>there are no genes in this window</i>
		VAP	0.42	1.01E-02	0.557			
		VSL	0.47	4.60E-03	0.754			
piR-bta- 13353933	14:71,977,746- 71,977,776	DDROP	-0.49	2.91E-03	0.384	14:71,975,246- 71,980,276	exonic piRNA	<i>STOX1</i>





\*: piR-ssc-127298 is present in (a) and in (b). ACRO = Abnormal Acrosomes; CON = Concentration; DDROP = Distal droplets; HABN = Head abnormalities; MT = Motility; NABN = Neck abnormalities; ORT = Osmotic Resistance Test; PDROP = Proximal droplets; TABN = Tail abnormalities; VAP = Average Path Velocity; VCL = Curvilinear Velocity; VIAB = Viability; VSL = Straight Line Velocity. Genes: genes located within  $\pm 2.5$ kb distance from the corresponding piRNA.

## Supporting information

**Table S1.** Percentage of reads mapping to different RNA biotypes. Average and Standard Deviation (SD) for the dataset of 34 ejaculates. The proportion of these biotypes were assessed based on the proportion of mapped reads without prior filtering of read length.

**Table S2.** Phenotypic values of the 13 semen quality parameters in the 34 samples including the mean value and Standard Deviation (SD).

**Table S3.** Number of human, mouse and swine piRNAs from the piRBase database retrieved, mapped, present in the boar sperm and associated to semen quality.

**Table S4.** Genomic coordinates of the phenotype-associated piRNAs, species of origin and correlated phenotypes. Chr = chromosome; Corr = correlation.

**Table S5.** Distance between the phenotype-associated piRNAs and Repetitive Element (RE) classes and number of piRNAs in the vicinity of each RE class. SINE = Short interspersed nuclear element; LINE = Long interspersed nuclear element; DNA/hAT = hAT DNA transposon; LTR = Long terminal repeats.

**Table S6.** Correlation between total piRNA abundance and the 13 phenotypes. Acronym phenotypes can be found in Table S1.

**Table S7.** Biological processes (gene ontology GO term) significantly overrepresented within the set of genes located in the  $\pm 2.5$ kb window centred at the phenotype associated piRNAs.

**Figure S1.** Relative nucleotide composition of the 6,729 piRNAs with CPM > 1 in more than 75% of the samples from position 1 to position 10.

**Figure S2.** Number of significantly associated piRNAs for each of the 13 phenotypes.

**Figure S3.** Violin plots of piRNA distribution within each Repetitive Element class. SINE = Short interspersed nuclear element; LINE = Long interspersed nuclear element; DNA/hAT = hAT DNA transposon; LTR = Long Terminal Repeat.