Title. Stable microRNA pairs: new perspectives in the search for male fertility biomarkers

Running title: Sperm miRNA pairs as fertility biomarkers

Author names and affiliations.

- C Corral-Vazquez (M.Sc.) [a]
  celia.corral@uab.cat
  ORCID: 0000-0002-0621-4590
- A. Salas-Huetos (Ph.D.) [a 1]
  albert.salas@utah.edu
  ORCID: 0000-0001-5914-6862
- J.Blanco (Ph.D.) [a]
  joan.blanco@uab.cat
  ORCID: 0000-0003-0647-3856
- F. Vidal (Ph.D.) [a]
  francesca.vidal@uab.cat
  ORCID: 0000-0002-0985-7348
- Z. Sarrate (Ph.D.) [a]
  zaida.sarrate@uab.cat
  ORCID: 0000-0001-9677-1376
- E Anton (Ph.D.) [a] **(corresponding author)**
  ester.anton@uab.cat
  Tel: +34 935813733
  ORCID: 0000-0002-8914-2667

[a] Genetics of Male Fertility Group, Unitat de Biologia Cel·lular (Facultat de Biociències), Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain.

[1] Present Address: Andrology and IVF Laboratory, Division of Urology, Department of Surgery, University of Utah, Salt Lake City, UT 84180, USA.

Capsule. Expression levels of sperm microRNA pairs can be used as fertility biomarkers in patients with seminal alterations and patients with unexplained male infertility.

Disclosure of Potential Conflicts of Interest. The authors report no conflicts of interest.

STRUCTURED ABSTRACT

Objective: To identify candidates of fertility biomarkers among pairs of human sperm microRNAs.

Design: Expression data of 736 sperm microRNAs from fertile and infertile individuals characterized in previous published studies by TaqMan® qPCR were re-examined. A set of microRNA pairs with the best biomarker potential were selected and validated by qRT-PCR in an independent cohort.

Setting: University research facility.

Patients: Semen samples were obtained from fertile (n=10) and infertile individuals (asthenozoospermia, n=10; teratozoospermia, n=10; oligozoospermia, n=10; unexplained male infertility (UMI), n=8). The validation cohort was composed of 9 fertile donors and 14 infertile patients with different seminal alterations.

Intervention(s): None.

Main Outcome Measure(s): Spearman test was used to select microRNA pairs with a correlated expression in fertile individuals and a non-correlated expression in each infertile group. The biomarker potential of these pairs was determined by Receiver Operating Characteristic curves. The differential relative expression of each pair in fertile and infertile populations was verified (Mann-Whitney test). The pairs that obtained the best results were validated by qRT-PCR.

Results: 48 pairs showed significant correlations in the fertile group. The pairs that were uncorrelated in the infertile populations and displayed the greatest biomarker potential were hsa-miR-942-5p/hsa-miR-1208 (asthenozoospermia), hsa-miR-296-5p/hsa-miR-328-3p (teratozoospermia), hsa-miR-139-5p/hsa-miR-1260a (oligozoospermia), and hsa-miR-34b-3p/hsa-miR-93-3p (UMI). The hsa-miR-942-5p/hsa-miR-1208 pair displayed the greatest potential for detecting seminal alterations in the qRT-PCR validation (85.71% True Positives).

Conclusions: The pairs hsa-miR-942-5p/hsa-miR-1208 and hsa-miR-34b-3p/hsa-miR-93-3p have the potential to become new molecular biomarkers that could help to diagnose male infertility, especially in cases of UMI or when seminal parameters are close to the threshold values.

Keywords: miRNA, sperm, biomarker, stable pairs, male infertility

INTRODUCTION

Approximately 15% of couples trying to conceive are infertile (1). Among this 15%, it has been suggested that males contribute to infertility in 30–40% of cases, and are the only causal factor in one-fifth of them (2). The diagnosis of male infertility is primarily based on the seminogram, which mostly relies on the microscopic analysis of sperm morphology, motility, and concentration (1). Nevertheless, the diagnostic potential of this approach has been called into question from the 1980s (3) to present (4). One of the main arguments against its predictive power is the noteworthy heterogeneity of semen samples, since significant variations are noticeable even between consecutive samples of the same individual. Furthermore, although World Health Organization established threshold for these parameters, based on data obtained from fertile men, these reference values have been modified over the last years (1,5,6). Therefore, a diagnosis fully based on these thresholds is not recommended, especially when the seminal parameters of a certain sample are close to the reference standards. These handicaps suppose a limitation of the test, which only constitutes an overview of fluctuating biological variables that do not reliably indicate the general fertility potential of the patient (7). Moreover, from all diagnosed male infertility cases, approximately 30% have an unknown origin non-detectable by seminogram analyses, which is known as *unexplained male infertility* (UMI) (8).

At present, the need for alternative diagnostic approaches is encouraging the search for new effective male infertility biomarkers. Researchers have delved into the predictive power of chromosome aneuploidies, sperm DNA integrity, epigenetic marks, and RNA profiling as markers (9). Specifically, the assessment of sperm RNAs constitutes a source of non-invasive molecular infertility indicators, which enables the acquisition of in-depth information beyond the seminogram (10).

Although spermatozoa are transcriptionally inactive cells, they carry a small amount of RNAs (10–20 fg) (11), which includes a wide variety of coding (mRNAs) and noncoding (small RNAs and long noncoding RNAs) molecules (12). Among the small RNA subpopulation, microRNAs (miRNAs) are important regulators of gene expression via mRNA degradation or translational repression, which is essential in several biological processes (13).

Sperm RNAs participate in regulatory pathways during spermatogenesis and early embryo development (12). Some studies have revealed an association between male infertility and specific variations of the sperm transcriptome profile. These variations have been associated with sperm

mRNA content (10,14–16) and miRNA expression profiles (17–21). Altogether, these studies have allowed the identification of RNA molecules that are differentially expressed in infertile patients, which follow specific patterns depending on the infertility phenotype.

Furthermore, the presence of pairs of mRNAs whose expression is strongly correlated in fertile donors but disrupted in infertile patients has been reported. A clear example are the protamine 1 and 2 (PRM1, PRM2) transcripts, whose expression ratio can be altered in patients with impaired spermatogenesis (22,23), asthenozoospermia (24–26), or varicocele (27). Other examples of sperm transcripts pairs are SRP54/ACSBG2, SRP54/GRP137, SRP54/TTC7A, SRP54/UBAC1.2, and UBAC1.2/RNF7, which have been shown to exhibit a disrupted expression ratio in infertile patients with Dysplasia of the Fibrous Sheath (28).

Nevertheless, no data regarding pairs of sperm miRNAs with a stable or disrupted expression ratio in relation to the fertility status has been reported so far. Given the strong regulatory role of miRNAs on mRNA expression, it is plausible that the maintenance of this stable correlation could also be reflected in miRNA expression profiles. The identification of 48 miRNA pairs with a highly stable relative expression in fertile patients provided the first hint of the possible role of these molecules as fertility biomarkers (29). However, the expression level of these miRNA pairs in infertile populations has not yet been assessed. In this sense, the discovery of a panel of miRNAs that would allow the classification of individuals according to their fertility potential would be of great interest. This categorization would be especially relevant to asses those cases in which conventional semen analyses do not offer a conclusive diagnosis, for example in normozoospermic infertile individuals with UMI, or also in individuals with seminal parameters close to threshold values.

The primary objective of the present study was to identify an optimal panel of biomarkers of male infertility among stable pairs of human sperm miRNAs. The evaluation was performed using the expression profiles of 736 miRNAs in 48 individuals corresponding to five different subgroups: fertile control individuals, asthenozoospermic, teratozoospermic, oligozoospermic, and normozoospermic infertile patients (UMI). The most suitable biomarker miRNA pairs for the correct diagnosis were determined in each group. The selected biomarker panels were further validated in an independent cohort of fertile and infertile patients by qPCR. Therefore, the ultimate objective was to set the basis for the development of new diagnosis approaches, aiming to provide a more accurate profiling of male fertility status.

MATERIALS AND METHODS

Sperm miRNA profiles

Sperm miRNA profiles from control and infertile patients were compiled from prior published studies (19,29,30). In all cases, studies were performed in ejaculated samples collected after a period of sexual abstinence of 3-5 days. Control data corresponded to a population of 10 fertile individuals with normal seminogram, 46,XY karyotype, and proven fertility. Infertile populations included 38 patients that did not achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse, with a 46,XY karyotype and no urogenital abnormalities detected. These patients were classified into four categories according to their seminal parameters (World Health Organization 2010) (1,31) that were analyzed following the standardized checklist set by Björndahl et al. (32): i) 10 individuals with an altered seminogram in which the sperm motility was the sole parameter affected (i.e. asthenozoospermic group); ii) 10 individuals with an altered seminogram in which the sperm morphology was the sole parameter affected (i.e. teratozoospermic group); iii) 10 individuals with an altered seminogram in which the sperm count was the sole parameter affected (i.e. oligozoospermic group); iv) Eight infertile individuals with 46,XY karyotype and normal seminal parameters from couples in which any female contribution to infertility was discarded (i.e. normozoospermic group with UMI).

Written informed consent was obtained from all patients, and the study was approved by the ethics committees of the collaborative centers and the Ethics Committee on Animal and Human Experimentation of the Universitat Autònoma de Barcelona.

The methodology for RNA isolation and quality control of the samples allocated to this study is described elsewhere (19,29,30). Total sperm RNA fractions were extracted from all samples following the same protocol and applying the same quality controls. Briefly, TRIzol reagent (Thermo Fisher Scientific, Carlsbad, CA, USA) was used to isolate sperm RNA after treating the semen samples with a somatic cell lysis buffer to eliminate any non-sperm cells (33). Proper somatic cell elimination was confirmed by optical microscopic examination. rDNase I was used to ensuring the absence of DNA in the obtained RNA samples (Thermo Fisher Scientific). A qRT-PCR amplification was performed in order to verify the integrity of the obtained RNA and the absence of genomic DNA (*PRM1* and *GAPDH* genes) as well as the absence of somatic RNA (*CD45* gene). As an additional quality control, small RNA distribution and the absence of intact rRNA was assessed by nanoeletrophoretic Small-RNA and Nano-RNA chips (2100 Bioanalyzer; Agilent Technologies).

Afterward, cDNA was preamplified using the TaqMan® PreAmp kit (Thermo Fisher Scientific) and the expression profiles of 736 human miRNAs were evaluated by quantitative real-time PCR (qRT-PCR) using the TaqMan® Array Human MicroRNA A and B Cards Set v3.0 (Thermo Fisher Scientific). Expression data were analyzed using SDS v.2.3 and RQ Manager v.1.2 software (Thermo Fisher Scientific) considering threshold cycle (Ct) values < 35 and ≥ 15. The expression values obtained were normalized against the Mean-Centering Restricted (MCR) method (34).

Statistical and bioinformatics analyses

Statistical analyses were performed using R v2.14.2 (www.r-project.org) (35) and the HTqPCR package v1.13.1 (www.bioconductor.org) (36). For the whole bioinformatics analysis pipeline (see **Figure 1** for a schematic diagram) each infertile subgroup was considered as a separated population (control vs asthenozoospermia, control vs. teratozoospermia, control vs. oligozoospermia and control vs. UMI). For each fertile and infertile population, only the ubiquitous miRNAs (expressed in all the samples) were taken into account.

The presence of ubiquitous miRNA pairs with a constant relative expression in the control population was assessed by Spearman correlation test over normalized Ct (normCt) values of every possible pairwise miRNA–miRNA combination. *P*-values < 0.05 were considered significant following strict post-hoc Bonferroni correction. The pairs significantly correlated in the control population were further evaluated in each infertile population with the Spearman test. The miRNA–miRNA combinations that were detected as correlated in the control population and showed non-correlated normCt values in each infertile group (Spearman *P*-values > 0.05 after Bonferroni correction) were selected for downstream analysis.

The ΔnormCt values ($normCt_{miRNA1} – normCt_{miRNA2}$) of the obtained pairs were calculated. To assess the accuracy of these pairs when discerning infertile individuals, ROC curve analyses were performed employing the R Graphical User Interface Deducer (www.deducer.org). In these analyses, ΔnormCt values for each infertile individual were compared to values of the control samples. The obtained area under the curve (AUC) values were indicative of the discriminatory potential of each evaluated miRNA pair. These scores were classified into *excellent* ($0.90 \leq AUC \leq 1.00$), *good* ($0.80 \leq AUC < 0.90$), *fair* ($0.70 \leq AUC < 0.80$), *poor* ($0.60 \leq AUC < 0.70$), and *failed* ($AUC < 0.60$). The pair that reached the highest AUC score in each population was selected, so a total of four pairs were considered for further analyses.

The ΔnormCt distribution of each pair in every infertile group was compared to control reference values by Mann–Whitney *U* test ($p < 0.05$ were considered as significant). The mean and range of ΔnormCt values (from minimum to maximum) were established for each miRNA pair and population. The presence of outliers was evaluated by Grubbs' test. Besides, percentages of presence (indicating the percentage of samples in which both miRNAs were expressed) were calculated in each infertile group (**Figure 1**).

Finally, in order to check possible relations between the expression patterns of the selected miRNA pairs and their genome position (e.g. pairs that are clustered within a given gene would have interrelated transcription processes), the chromosome and genome location of the sequences were verified using miRBase (www.mirbase.org), and the Genome Browser database (GRCh38/hg38 assembly; www.genome.ucsc.edu).

Validation of the selected miRNA biomarker pairs by qRT-PCR

The selected biomarker pairs were validated in a new population of 9 fertile men and 14 infertile patients (**Supplemental Table 1**). Fertile individuals met the same inclusion criteria as the ones described before for the control population (normal seminal parameters, normal karyotype, and proven fertility). Infertile individuals were selected in the collaborating centers after consulting for infertility. According to World Health Organization (1), all of them showed abnormal seminal parameters.

A semen sample from each individual was compiled after 3-5 days of sexual abstinence. Samples were processed likewise the individuals of the study populations. That is, RNA extractions were performed using the same protocols and the obtained RNA fractions were subjected to the same quality controls stated above for the several populations of study. *TaqMan® microRNA Assays* (Thermo Fisher Scientific) were employed to evaluate the miRNA pairs selected from the previous analysis, plus two additional miRNAs described elsewhere as suitable normalizers for sperm miRNA expression studies (i.e. hsa-miR-100-5p, and hsa-miR-30a-5p) (37). Firstly, a miRNA-specific reverse-transcription (*TaqMan® MicroRNA Reverse Transcription Kit*, Thermo Fisher Scientific) was performed using 1-10 ng of RNA. The resulting cDNA was submitted to qRT-PCR (*TaqMan® Universal PCR Master Mix II, No UNG*, Thermo Fisher Scientific) with three technical replicates, using 384-well plates and an ABI Prism® 7900HT thermocycler (Thermo Fisher Scientific) according to the manufacturer's instructions. Expression data were analyzed using SDS v.2.4 and RQ Manager v1.2 software. Afterwards, ΔnormCt values were calculated for every miRNA pair. Data

from the fertile and infertile individuals were statistically compared by Mann–Whitney *U* test, considering $p < 0.05$ as significant. The mean and range of ΔnormCt values (from minimum to maximum) of each pair was established from the control group of the validation cohort after verifying the absence of outliers by Grubbs' test. The ΔnormCt value of each infertile sample was compared to this range: values that were higher or lower than the control range were categorized as True Positives (TP).

RESULTS

Selection of stable miRNA pairs in the control group with uncorrelated expression in infertile populations

A summary of the number of miRNA pairs that were selected according to their ubiquity and correlation/non-correlation characteristics is displayed in **Supplemental Table 2**. Among the miRNAs that were ubiquitously expressed in the fertile group, 48 miRNA pairs showed statistically significant correlated expression. Only 31 of these pairs were expressed in all the asthenozoospermic patients; the expression of one or both miRNAs of the other 17 pairs was not detected in at least one patient of this group and thus discarded for subsequent tests. Among the 31 pairs, the correlation was maintained by one miRNA pair, thus, no correlation was shown in the expression of the other 30 pairs. In the teratozoospermic group, only 24 of the 48 control stable pairs were ubiquitously detected, from which 23 were non-correlatively expressed. Regarding the oligozoospermic and UMI populations, 20 and 19 pairs were found to be ubiquitous in each population, and 19 and 18 pairs of them respectively appeared to be uncorrelated.

Further analyses were based on these sets of ubiquitous miRNA pairs that presented a significant correlation in the control population but were uncorrelated in the infertile groups (**Table 1**).

Assessment of AUC values as an indicator of biomarker potential and evaluation of the selected miRNA pairs

ROC curve analyses revealed a wide range of AUC values for the established miRNA pairs (ranging from 0.34 to 1.00) (**Table 1**). Analyzing each infertile population, the highest AUC score (AUC = 1.00) was reached by the hsa-miR-139-5p/hsa-miR-1260a and hsa-miR-34b-3p/hsa-miR-93-3p pairs in the oligozoospermic and UMI groups, respectively (**Table 1**). In the asthenozoospermic

and teratozoospermic populations, high AUC scores were also achieved (0.91 and 0.87, respectively, by the pairs hsa-miR-942-5p/hsa-miR-1208 and hsa-miR-296-5p/hsa-miR-328-3p) (**Table 1**).

The ΔnormCt distributions of the above-mentioned pairs, along with their presence percentages, mean values, ranges, and TP percentages are summarized in **Figure 2**. Assessing the data of the miRNA pairs with the highest AUC score in each infertile population, the four pairs presented a significant difference ($p < 0.05$) between fertile individuals and their specific infertile population. Additionally, when ΔnormCt distributions of the pairs were cross-analyzed in the rest of the infertile groups, a high presence (> 80%) of the pairs hsa-miR-942-5p/hsa-miR-1208 (**Figure 2A**), hsa-miR-296-5p/hsa-miR-328-3p (**Figure 2B**) and hsa-miR-139-5p/hsa-miR-1260a (**Figure 2C**) was found in the asthenozoospermic, teratozoospermic and oligozoospermic samples, as well as significant expression differences when compared to control data. When categorizing infertile individuals according to the control ΔnormCt range, the pair hsa-miR-942-5p/hsa-miR-1208 achieved the best results with a 100% of TP in all the studied populations (**Figure 2A**). Regarding UMI population, although the four miRNA pairs achieved a 100% of TP, only the pairs hsa-miR-296-5p/hsa-miR-328-3p (**Figure 2B**), and hsa-miR-34b-3p/hsa-miR-93-3p (**Figure 2D**) showed a high presence in this group (> 80%) and significant differences when compared to fertile ΔnormCt ranges. In fact, one of these two pairs (hsa-miR-34b-3p/hsa-miR-93-3p) displayed a significantly different ΔnormCt range only in the UMI population, but not in the other infertile groups (**Figure 2D**).

Regarding the genome annotation of these miRNAs, although some stem-loop sequences were found to be located within intronic or exonic regions, no coinciding miRNA cluster, host-gene or chromosome was identified in any pair (**Supplemental Table 3**).

Therefore, and according to all compiled results, the four miRNA pairs mentioned above (hsa-miR-942-5p/hsa-miR-1208 for asthenozoospermia, hsa-miR-296-5p/hsa-miR-328-3p for teratozoospermia, hsa-miR-139-5p/hsa-miR-1260a for oligozoospermia, and hsa-miR-34b-3p/hsa-miR-93-3p for normozoospermia) were considered as good potential infertility biomarker candidates and thus further validated in additional cohorts of individuals.

Validation of the selected miRNA pairs

Normalized Ct and ΔnormCt values from the qPCR of the eight selected miRNAs (hsa-miR-942-5p, hsa-miR-1208, hsa-miR-296-5p, hsa-miR-328-3p, hsa-miR-139-5p, hsa-miR-1260a, hsa-miR-34b-3p,

and hsa-miR-93-3p) that were obtained from the fertile and infertile validation cohorts are shown in **Supplemental Table 4**. The analyzed miRNAs, including the normalizer molecules, were expressed in all the samples (100% presence).

Statistical comparisons between ΔnormCt from fertile and infertile cohorts are shown in **Figure 3**. Significant differences were only observed regarding the hsa-miR-942-5p/hsa-miR-1208 pair.

When infertile individuals were compared over the ΔnormCt ranges of the nine control individuals, TP percentages could be determined (**Figure 3)**. The hsa-miR-942-5p/hsa-miR-1208 pair reached the highest TP percentage, allowing the correct classification of the 85.71% of infertile samples. Both the hsa-miR-139-5p/hsa-miR-1260a and hsa-miR-34b-3p/hsa-miR-93-3p obtained a 64.29% of TP, while the hsa-miR-296-5p/hsa-miR-328-3p pair only reached 21.43% TP.

DISCUSSION

The presence of miRNA pairs in the sperm transcriptome that maintain a constant expression ratio in fertile men has been reinforced by the results of the present study. The 48 pairs described by Salas-Huetos and collaborators (29) constituted the starting point to search for an optimal set of biomarkers to classify individuals according to their fertility status. Data compiled in this study constitute a basis for new diagnosis methods in which expression levels of sperm miRNA pairs could be indicative of the male fertility status. This kind of approach brings an additional tool for assessing sperm malfunctioning related to infertility, so current diagnostic methods could be complemented with a molecular insight. This would be especially informative when facing diagnostic cases in which seminogram parameters are close to the threshold values (1), as minimal variations in the measurements of seminal parameters could be critical for the final diagnosis. Also, this type of molecular approach could be a crucial tool for tracing cases of UMI since current methods cannot provide a reliable diagnosis.

Aiming for introducing miRNA biomarkers as an alternative diagnostic tool, employing qRT-PCR single assays constitutes an affordable resource in terms of time-consume and cost-effectiveness. Accordingly, to classify individuals in the fertile category, the ΔnormCt values of a given miRNA pair should fit in a specific control range; on the contrary, an alteration of the relative expression of the pair could be a signal of disruption related to infertility.

The approach of considering the differential expression of two molecules as a biomarker value instead of the normCt level of a single molecule is mainly based on the biological implications of these pairwise fluctuations in male fertility. In one hand, the analysis of the relationship between two molecules (ΔnormCt) is more robust than the analysis of a single molecule because it will provide information about the alteration of biological processes from two different standpoints. Therefore, no matter if the expression level of one or the other molecule is altered (or both in an uncorrelated way), the global ratio will reflect an alteration that will be detected by this pairwise analysis.

In the other hand, in complex cellular process (such as spermatogenesis) in which the expression of multiple genes is involved, the assessment of the relative presence of two molecules (and thus their regulatory pathways) could be more informative than the absolute expression value of two single molecules. In fact, PRM1 and PRM2 are a clear example of this type of interaction (22) and their aberrant expression ratios have been found to be associated with infertility. Other studies have also claimed that, beyond the heterogeneity observed among the sperm transcriptome in fertile individuals, the presence of a conserved set of transcripts involving several stable mRNA pairs is strictly regulated (38). Regarding miRNAs, we had also previously observed the presence of stable correlated miRNA pairs in spermatozoa from fertile individuals (29). In this context, the results obtained in this manuscript reinforce the idea that some of these miRNA pairs may participate in co-regulated transcriptional pathways that should be balanced for normal fertility. The connection between these molecules (and pathways) might not be evident, not even simple, or maybe it has not been described yet. However, what seems clear is that alterations in the relative co-expression of specific miRNA pairs have an association to different infertile conditions and therefore, these molecules have the potential to play a role as biomarkers.

Therefore, a primary goal of this study was searching for a small number of miRNA pairs that fulfilled some indispensable conditions for good potential biomarkers: i) being present in all sperm samples, ii) displaying a correlated expression in fertile individuals; iii) showing no correlation in an infertile population; and iv) exhibiting a strong predictive power to discern infertile from fertile individuals.

Regarding the first requirement, the criteria for selecting only ubiquitous miRNAs ensured that the final selected pair would be constitutively expressed in semen samples of every individual. To ensure the requirements ii) and iii), expression profiles of the miRNA pairs were determined by a

correlation test in where a Bonferroni correction was applied. This method provides a stringent threshold for the selection of strictly significant correlations.

To evaluate the iii) requirement, ROC curve analysis allowed obtaining AUC values that enabled classifying the predictive power of the selected pairs in measurable categories. Moreover, only the pairs that exhibited high AUC values were considered as valid potential biomarkers. This goal was achieved by the pairs hsa-miR-942-5p/hsa-miR-1208 (from asthenozoospermic individuals), hsa-miR-139-5p/hsa-miR-1260a (from oligozoospermic individuals), and hsa-miR-34b-3p/hsa-miR-93-3p (from UMI individuals), which presented an *excellent* predictive power with AUCs ≥ 0.90. Also, the AUC value of the biomarker pair selected from the teratozoospermic patients (hsa-miR-296-5p/hsa-miR-328-3p) did not reach the *excellent* classification but, even so, it achieved a *good* predictive power (AUC = 0.87) (**Table 1**).

To go forward in the evaluation of these four pairs as fertility biomarkers, their expression profiles were also cross-analyzed in all included infertile populations to evaluate their biomarker potential for detecting individuals with other infertility phenotypes. It was observed that three of the pairs (hsa-miR-942-5p/hsa-miR-1208, hsa-miR-296-5p/hsa-miR-328-3p, and hsa-miR-139-5p/hsa-miR-1260a) were expressed in a high percentage of the asthenozoospermia, teratozoospermia and oligozoospermia samples, and showed differentiated expression patterns in these patients when compared to control samples (**Figure 2A-C**). These facts indicate that these pairs could be suitable candidates for detecting infertile individuals with alterations in sperm motility, morphology and count. This versatility is a convenient characteristic since more than one type of these sperm parameters tend to be simultaneously affected in samples from infertile males (e.g. the oligo-astheno-teratozoospermia syndrome has a frequency of ~26% of male infertility cases) (39,40). Among the three above-mentioned miRNA pairs, the hsa-miR-942-5p/hsa-miR-1208 pair showed an especially strong potential since all the infertile patients affected by any seminal alteration could be categorized as infertile by their ΔnormCt value (always located outside the range established from fertile individuals).

Concerning the results derived from the qRT-PCR validation, the ubiquity of the hsa-miR-942-5p/hsa-miR-1208 pair, its differential relative expression between the fertile and infertile cohorts, and its high rate of correct categorization of infertile patients were confirmed (**Figure 3**). Regarding the hsa-miR-296-5p/hsa-miR-328-3p and hsa-miR-139-5p/hsa-miR-1260a pairs, although initial analyses predicted a strong biomarker power, they did not show significant differences between fertile and infertile patients in the validation stage. Although the miR-139-

5p/miR-1260a pair led to a 64.29% of TP, the high dispersion and overlapping of the compiled data suggests that a categorization based on the established control range could probably be unreliable. Therefore, both the initial data analysis and the qRT-PCR validation study indicate that the hsa-miR-942-5p/hsa-miR-1208 pair displays the best potential for the diagnosis of infertile males with seminal alterations. Although the biological relevance of the expression impairment of this pair is still unknown, many studies have suggested that altered miRNA expression patterns could be the origin of seminal defects, since abnormal levels of certain miRNAs could modify the gene expression status of sperm cells (18,19,41). A search in the public databases miRTarBase (www.mirtarbase.mbc.nctu.edu.tw/php/index.php) (42) and CSmiRTar (http://cosbi.ee.ncku.edu.tw/CSmiRTar/) (43) allowed identifying testis-expressed genes of hsa-miR-942-5p and hsa-miR-1208 that have been validated as miRNA targets by functional studies. Despite 157 and 42 target genes were identified for of hsa-miR-942-5p and hsa-miR-1208 respectively, Gene Ontology analyses of functional annotation clustering showed no significant biological processes associated to those genes (data not shown). Nevertheless, an overexpression of hsa-miR-942-5p have been associated with low sperm motility in a previously published study, suggesting a possible association between this miRNA and spermatogenesis (19).

Regarding UMI biomarkers, although the initial analysis showed a full classification of infertile patients by the hsa-miR-942-5p/hsa-miR-1208 pair, its presence was only detected in a 25% of the UMI samples (**Figure 2A**). Therefore, the expression of these miRNAs is predicted to be very limited in sperm cells of UMI patients, which implies poor suitability as a biomarker of this pathology. This limitation implies a necessity of including an additional UMI-specific miRNA pair to the final biomarker panel. For this purpose, other two pairs (hsa-miR-296-5p/hsa-miR-328-3p and hsa-miR-34b-3p/hsa-miR-93-3p) showed good potential for detecting infertile individuals with no seminal alterations concerning ubiquity, differential relative expression, and TP percentages in the initial analyses **(Figure 2B and 2D)**. In particular, the hsa-miR-34b-3p/hsa-miR-93-3p pair appears to be an optimal biomarker candidate since it was originally selected from the specific transcriptome of UMI individuals, being the only one to display a 100% of presence in UMI patients and a statistically significant decrease in the relative expression when compared to controls (**Figure 2D).** Regarding the possible biological relevance of these miRNAs in fertility-related pathways, although several validated target genes expressed in testis were identified for both miRNAs in miRTarBase and CSmiRTar (95 for hsa-miR-34b-3p and 175 for hsa-miR-93-3p), no statistically significant gene ontology clustering was found (data not shown). Nevertheless, other

research studies have revealed that hsa-miR-34b-5p has been found to be especially enriched in germ cells of the adult testis (44)**,** and participates in the regulation of genes involved in spermatogenic processes (45). In addition, miR-34b/c has also been detected in zona pellucida–bound sperm cells, supporting their possible role in fertilization (46). Moreover, a possible relationship between miR-34b expression and ICSI success has also been described (47).

Unfortunately, although a great biomarker potential of the hsa-miR-34b-3p/hsa-miR-93-3p pair was observed in the initial analyses, no patients with UMI could be included in the infertile cohort of the qRT-PCR validation study. This represents the main limitation of the study as the biomarker suitability of this miRNA pair still needs to be verified before considering its use for clinical diagnosis. Nevertheless, the preliminary results of the hsa-miR-34b-3p/hsa-miR-93-3p pair are promising regarding UMI diagnostic.

CONCLUSIONS

The results of the present study suggest that the expression analysis of the miRNA pairs hsa-miR-942-5p/hsa-miR-1208 and hsa-miR-34b-3p/hsa-miR-93-3p constitute an efficient tool that could help to assess male infertility in patients with seminal alterations and UMI, respectively. This approach opens the possibility of including a molecular categorization of patients with different infertility phenotypes in clinical diagnosis, becoming especially relevant in cases of infertile individuals with seminal parameters that are within or close to the threshold values. Nevertheless, the transition of the results from basic research to clinics requires some further steps that will primarily include a validation in a much larger and heterogeneous infertile population.

AUTHORS' ROLES

C.C-V.: data collection, data analysis and interpretation, manuscript preparation, and critical discussion. A.S-H.: data collection, data analysis and interpretation, and critical discussion. J.B.: conception and design, data analysis and interpretation, manuscript preparation, and critical discussion. F.V.: data analysis and interpretation and critical discussion. Z.S.: data analysis and interpretation and critical discussion. E.A.: conception and design, data analysis and interpretation, manuscript preparation, and critical discussion. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the 2017/SGR-503 project (Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya, Spain) and UAB CF-180034 grant (Universitat Autònoma de Barcelona). Celia Corral-Vazquez is a recipient of a Personal Investigador en Formació grant UAB/PIF2015 (Universitat Autònoma de Barcelona).

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

1. World Health Organization. WHO laboratory manual for the examination and processing of human semen. 5th ed. Geneva: World Health Organization; 2010.
2. Kumar N, Singh AK. Trends of male factor infertility, an important cause of infertility: A review of literature. J Hum Reprod Sci 2015;8:191–6.
3. Glazener C, Coulson C, Lambert P, Watt E, Hinton R, Kelly N, et al. The value of artificial insemination with husband's semen in infertility due to failure of postcoital sperm-mucus penetration-controlled trial of treatment. Br J Obstet Gynaecol 1987;94:774–8.
4. Tomlinson MJ. Uncertainty of measurement and clinical value of semen analysis: has standardisation through professional guidelines helped or hindered progress? Andrology 2016;4:763–70.
5. World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction. 4th ed. Cambridge: Cambridge University Press; 1999.
6. World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction. 3rd ed. Cambridge: Cambridge University Press; 1992.
7. Lewis SEM. Is sperm evaluation useful in predicting human fertility? Reproduction 2007;134:31–40.
8. Ray A, Shah A, Gudi A, Homburg R. Unexplained infertility: An update and review of practice. Reprod Biomed Online 2012;24:591–602.
9. Anton E, Krawetz S. Spermatozoa as biomarkers for the assessment of human male infertility and genotoxicity. Syst Biol Reprod Med 2012;58:41–50.
10. Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, et al. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum Mol Genet 2007;16:763–73.
11. Krawetz S. Paternal contribution: new insights and future challenges. Nat Rev Genet 2005;6:633–42.
12. Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA. The presence, role and clinical use of spermatozoal RNAs. Hum Reprod Update 2013;19:604–24.
13. Mott JL, Mohr AM. Overview of MicroRNA Biology Justin. Semin Liver Dis 2015;35:3–11.

14. Aslani F, Modarresi MH, Soltanghoraee H, Akhondi MM, Shabani A, Lakpour N, et al. Seminal molecular markers as a non-invasive diagnostic tool for the evaluation of spermatogenesis in non-obstructive azoospermia. Syst Biol Reprod Med 2011;57:190–6.
15. 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;29:3–10.
16. Garrido N, Martínez-Conejero JA, Jauregui J, Horcajadas JA, Simón C, Remohí J, et al. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. Fertil Steril 2009;91:1307–10.
17. Abu-Halima M, Hammadeh M, Schmitt J, Leidinger P, Keller A, Meese E, et al. Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. Fertil Steril 2013;99:1249–55.
18. Abu-Halima M, Hammadeh M, Backes C, Fischer U, Leidinger P, Lubbad AM, et al. Panel of five microRNAs as potential biomarkers for the diagnosis and assessment of male infertility. Fertil Steril 2014;102:989–97.
19. Salas-Huetos A, Blanco J, Vidal F, Godo A, Grossmann M, Pons MC, et al. Spermatozoa from patients with seminal alterations exhibit a differential micro-ribonucleic acid profile. Fertil Steril 2015;104:591–601.
20. Ji Z, Lu R, Mou L, Duan YG, Zhang Q, Wang Y, et al. Expressions of miR-15a and its target gene HSPA1B in the spermatozoa of patients with varicocele. Reproduction 2014;147:693–701.
21. Zhou J-H, Zhou QZ, Lyu X-M, Zhu T, Chen Z-J, Chen M-K, et al. The expression of cysteine-rich secretory protein 2 (CRISP2) and its specific regulator miR-27b in the spermatozoa of patients with asthenozoospermia. Biol Reprod 2015;92:1–9.
22. Steger K, Wilhelm J, Konrad L, Stalf T, Greb R, Diemer T, et al. Both protamine-1 to protamine-2 mRNA ratio and Bcl2 mRNA content in testicular spermatids and ejaculated spermatozoa discriminate between fertile and infertile men. Hum Reprod 2008;23:11–6.
23. Aoki V, Liu L, Carrell D. A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. Mol Hum Reprod 2006;12:41–50.
24. Kempisty B, Depa-Martynow M, Lianeri M, Jedrzejczak P, Darul-Wasowicz A, Jagodzinski P. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. Folia Histochem Cytobiol 2007;45:109–13.

519 25. Lambard S, Galeraud-Denis I, Martin G, Levy R, Chocat A, Carreau S. Analysis and
520 significance of mRNA in human ejaculated sperm from normozoospermic donors:
521 Relationship to sperm motility and capacitation. Mol Hum Reprod 2004;10:535–41.
522 26. Jedrzejczak P, Kempisty B, Bryja A, Mostowska M, Depa-Martynow M, Pawelczyk L, et al.
523 Quantitative assessment of transition proteins 1, 2 spermatid-specific linker histone H1-like
524 protein transcripts in spermatozoa from normozoospermic and asthenozoospermic men.
525 Arch Androl 2007;53:199–205.
526 27. Ni K, Steger K, Yang H, Wang H, Hu K, Chen B. Sperm Protamine mRNA Ratio and DNA
527 Fragmentation Index Represent Reliable Clinical Biomarkers for Men with Varicocele after
528 Microsurgical Varicocele Ligation. J Urol 2014;192:170–6.
529 28. Lima-Souza A, Anton E, Mao S, Ho WJ, Krawetz SA. A platform for evaluating sperm RNA
530 biomarkers: Dysplasia of the fibrous sheath - Testing the concept. Fertil Steril
531 2012;97:1061-1066.e3.
532 29. Salas-Huetos A, Blanco J, Vidal F, Mercader JM, Garrido N, Anton E. New insights into the
533 expression profile and function of micro-ribonucleic acid in human spermatozoa. Fertil
534 Steril 2014;102:213–22.
535 30. Salas-Huetos A, Blanco J, Vidal F, Grossmann M, Pons MC, Garrido N, et al. Sperm from
536 normozoospermic fertile and infertile individuals convey a distinct miRNA cargo. Andrology
537 2016;4:1028–36.
538 31. Kruger TF, Acosta AA, Simmons KF, Swanson RJ, Matta JF, Oehninger S. Predictive value of
539 abnormal sperm morphology in in vitro fertilization. Fertil Steril 1988;49:112–7.
540 32. Björndahl L, Barratt CLR, Mortimer D, Jouannet P. 'How to count sperm properly': Checklist
541 for acceptability of studies based on human semen analysis. Hum Reprod 2016;31:227–32.
542 33. Goodrich RJ, Johnson GD, Krawetz S. The preparation of human spermatozoal RNA for
543 clinical analysis. Arch Androl 2007;53:161–7.
544 34. Wylie D, Shelton J, Choudhary A, Adai AT. A novel mean-centering method for normalizing
545 microRNA expression from high-throughput RT-qPCR data. BMC Res Notes 2011;4:555.
546 35. R Core Team. R: A language and environment for statistical computing. R Found. Stat.
547 Comput. Vienna, Austria. 2014;
548 36. Dvinge H, Bertone P. HTqPCR: high-throughput analysis and visualization of quantitative
549 real-time PCR data in R. Bioinformatics 2009;25:3325–6.
550 37. Corral-Vazquez C, Blanco J, Salas-Huetos A, Vidal F, Anton E. Normalization matters:

551 tracking the best strategy for sperm miRNA quantification. Mol Hum Reprod 2017;23:45–
552 53.
553 38. Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, Krawetz SA. Identification of
554 human sperm transcripts as candidate markers of male fertility. J Mol Med 2009;87:735–
555 48.
556 39. Dohle G, Colpi G, Hargreave T, Papp G, Jungwirth A, Weidner W. EAU guidelines on male
557 infertility. Eur Urol 2005;48:703–11.
558 40. World Health Organization. WHO manual for the standardized investigation and diagnosis
559 of the infertile couple. 4th ed. Cambridge UK: Cambridge University Press; 2000.
560 41. Abhari A, Zarghami N, Shahnazi V, Barzegar A, Farzadi L, Karami H, et al. Significance of
561 microRNA targeted estrogen receptor in male fertility. Iran J Basic Med Sci 2014;17:81–6.
562 42. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. MiRTarBase update 2018:
563 A resource for experimentally validated microRNA-target interactions. Nucleic Acids Res
564 2018;46:D296–302.
565 43. Wu WS, Tu BW, Chen T Te, Hou SW, Tseng JT. CSmiRTar: Condition-specific microRNA
566 targets database. PLoS One 2017;12:1–16.
567 44. Mclver SC, Roman SD, Nixon B, Mclaughlin EA. miRNA and mammalian male germ cells.
568 Hum Reprod Update 2012;18:44–59.
569 45. Wu J, Bao J, Kim M, Yuan S, Tang C, Zheng H, et al. Two miRNA clusters, miR-34b/c and miR-
570 449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis.
571 Proc Natl Acad Sci U S A 2014;111:E2851-7.
572 46. Liu W, Pang R, Chiu P, Wong B, Lao K, Lee K, et al. Sperm-borne microRNA-34c is required
573 for the first cleavage division in mouse. Proc Natl Acad Sci USA 2012;109:490–494.
574 47. Cui L, Fang L, Shi B, Qiu S, Ye Y. Spermatozoa micro ribonucleic acid-34c level is correlated
575 with intracytoplasmic sperm injection outcomes. Fertil Steril 2015;104:312-317.e1.
576
577

TABLES

Table 1. Pairs of ubiquitous microRNAs (miRNAs) with a correlated expression in the control (fertile) population but uncorrelated in the infertile populations. Area under the curve (AUC) values of each non-correlated pair are indicated. AUC values are classified as *excellent* (0.90 ≤ AUC ≤ 1.00), *good* (0.80 ≤ AUC < 0.90), *fair* (0.70 ≤ AUC < 0.80), *poor* (0.60 ≤ AUC < 0.70), and *failed* (AUC < 0.60).

Asthenozoospermia			Teratozoospermia			Oligozoospermia			Unexplained Male Infertility		
Pairs of miRNAs		AUC	Pairs of miRNAs		AUC	Pairs of miRNAs		AUC	Pairs of miRNAs		AUC
miRNA 1	miRNA 2		miRNA 1	miRNA 2		miRNA 1	miRNA 2		miRNA 1	miRNA 2	
miR-942-5p	miR-1208	0.91	miR-296-5p	miR-328-3p	0.87	miR-139-5p	miR-1260a	1.00	miR-34b-3p	miR-93-3p	1.00
miR-92a-3p	miR-636	0.91	miR-200b-3p	miR-491-5p	0.84	miR-17-5p	miR-19b-3p	0.94	miR-99b-5p	miR-125b-5p	0.97
let-7d-5p	miR-92a-3p	0.90	miR-92a-3p	miR-636	0.80	miR-146b-5p	miR-517c-3p	0.88	miR-99b-5p	miR-193a-5p	0.96
let-7c	miR-92a-3p	0.88	miR-130a-3p	miR-184	0.79	miR-99b-5p	miR-193a-5p	0.88	miR-190b	miR-151a-5p	0.91
miR-200b-3p	miR-491-5p	0.85	miR-20a-5p	miR-20b-5p	0.78	miR-92a-3p	miR-636	0.87	miR-92a-3p	miR-636	0.90
miR-139-5p	miR-1260a	0.83	miR-25-3p	miR-296-5p	0.75	miR-100-5p	miR-125b-5p	0.83	miR-17-5p	miR-19b-3p	0.84
miR-190b	miR-151a-5p	0.82	miR-17-5p	miR-19b-3p	0.74	miR-25-3p	miR-186-5p	0.80	miR-200b-3p	miR-491-5p	0.80
miR-192-5p	miR-628-3p	0.82	miR-148a-3p	miR-150-5p	0.71	miR-200b-3p	miR-491-5p	0.78	let-7c	miR-92a-3p	0.79
miR-30a-5p	miR-622	0.82	miR-92a-3p	miR-125b-5p	0.71	let-7c	miR-92a-3p	0.72	miR-148a-3p	miR-150-5p	0.77
miR-99b-5p	miR-193a-5p	0.79	miR-99b-5p	miR-193a-5p	0.67	miR-99a-5p	miR-100-5p	0.71	miR-21-5p	miR-148a-3p	0.67
miR-146b-5p	miR-517c-3p	0.74	miR-20a-5p	miR-106a-5p	0.66	miR-92a-3p	miR-125b-5p	0.70	miR-92a-3p	miR-125b-5p	0.66
miR-149-5p	miR-190b	0.73	miR-152	miR-218-5p	0.64	miR-20a-5p	miR-20b-5p	0.69	miR-25-3p	miR-186-5p	0.66
miR-92a-3p	miR-125b-5p	0.72	let-7d-5p	miR-92a-3p	0.62	miR-20a-5p	miR-106a-5p	0.66	miR-152-3p	miR-218-5p	0.62
miR-125a-3p	miR-371-3p	0.71	miR-99a-5p	miR-100-5p	0.62	miR-149-5p	miR-190b	0.66	miR-197-3p	miR-1291	0.61
miR-21-5p	miR-148a-3p	0.71	miR-25-3p	miR-186-5p	0.61	miR-517a-3p	miR-517c-3p	0.64	miR-20a-5p	miR-106a-5p	0.57
miR-100-5p	miR-125b-5p	0.70	miR-197-3p	miR-1291	0.55	miR-519d	miR-190b	0.62	miR-20a-5p	miR-20b-5p	0.57
miR-512-3p	miR-517a-3p	0.69	let-7c	miR-92a-3p	0.55	let-7d-5p	miR-92a-3p	0.60	miR-29a-3p	miR-625-5p	0.56
miR-10a-5p	miR-628-3p	0.69	miR-21-5p	miR-148a-3p	0.55	miR-99b-5p	miR-125b-5p	0.58	miR-100-5p	miR-125b-5p	0.52
miR-148a-3p	miR-150-5p	0.62	miR-30a-5p	miR-30d-5p	0.54	miR-148a-3p	miR-150-5p	0.56			
miR-20a-5p	miR-20b-5p	0.60	miR-100-5p	miR-125b-5p	0.53						
miR-34b-3p	miR-93-3p	0.59	miR-30a-3p	miR-30e-3p	0.53						
miR-30a-5p	miR-30d-5p	0.58	miR-146b-5p	miR-517c-3p	0.52						
miR-17-5p	miR-19b-3p	0.57	miR-99b-5p	miR-125b-5p	0.52						
miR-25-3p	miR-186-5p	0.57									
miR-99b-5p	miR-125b-5p	0.56									
miR-517a-3p	miR-517c-3p	0.56									
miR-152	miR-218-5p	0.55									
miR-20a-5p	miR-106a-5p	0.53									
miR-324-3p	miR-622	0.40									
miR-30a-3p	miR-30e-3p	0.34									

Supplemental Table 1. Demographic and seminal parameter data from all the individuals included in the several study populations and in the validation cohort. Spz = Spermatozoa; PR = Progressive; NP = Non-progressive; NF = Normal Forms; N = Normozoospermia; A = Asthenozoospermic; T = Teratozoospermic; O = Oligozoospermic; AT = Asthenoteratozoospermia; OA = Oligoasthenozoospermia; OT = Oligoteratozoospermia.

Supplemental Table 2. Number of ubiquitous and uncorrelated sperm miRNA pairs in each infertile population from the 48 stable pairs detected in the control population. Among the pairs with a ubiquitous expression (first row), it is indicated how many pairs show a disrupted correlation and therefore are selected for subsequent analyses (second row).

Supplemental Table 3. Genome annotation of the miRNA pairs selected from the studied infertile populations. Chromosome locations are described as well as genes situated in these regions.

Supplemental Table 4. Results of the qRT-PCR validation and description of the control and infertile cohorts. Normalized Ct (normCt) and ΔnormCt values of the analyzed miRNA pairs are displayed.

FIGURE LEGENDS

Figure 1. Schematic representation of the experimental design of the study. A = Asthenozoospermic; T = Teratozoospermic; O = Oligozoospermic; UMI = Unexplained Male Infertility.

Figure 2. ΔnormCt distribution per population of the four miRNA pairs with the highest Area Under the Curve**.** Red rectangles indicate data related to the specific population from which the miRNA pair was selected. Asterisks indicate a significant ΔnormCt difference ($p<0.05$) between the control and infertile populations (p-values are specified in the first row of the tables; p values are specified as N/A when $n<3$). ΔnormCt values of infertile individuals were considered as True Positives (TP) when they were not comprised within the range of minimum and maximum ΔnormCt values defined by the control population. C = Control; A = Asthenozoospermic; T = Teratozoospermic; O = Oligozoospermic; UMI = Unexplained Male Infertility.

Figure 3. ΔnormCt distribution of the four selected miRNA pairs in the validation cohorts. Asterisks indicate a significant ΔnormCt difference ($p<0.05$) between the control and infertile populations. ΔnormCt values of infertile individuals were considered as True Positives (TP) when they were not comprised within the range of minimum and maximum ΔnormCt values defined by the control population. C = Control; I = Infertile.