RBMO

ELSEVIER

REVIEW

Systematic review and evidence-based classification of differentially expressed mRNA in human spermatozoa: insights to improve the diagnosis and prognosis of male infertility

BIOGRAPHY

Dr. Ester Anton is a researcher in the Genetics of Male Fertility Group at Universitat Autònoma de Barcelona (Spain), where she also holds an Associate Professor position in the Department of Cell Biology, Physiology and Immunology. Her research explores the genetic causes of male infertility, and the impact of external factors on reproductive health.

Ester Anton, Coral Zurera-Egea, Roser Farriol, Zaida Sarrate, Joan Blanco*

KEY MESSAGES

This systematic review analyses 451 differentially expressed genes from 67 studies, identifying nine sperm mRNA linked to seminal abnormalities and reduced success of assisted reproductive technology. These genes represent a key panel of possible biomarkers associated with broader transcriptomic changes, offering insights into potential disruptions in spermatogenesis.

ABSTRACT

Since its emergence in the 2000s, the study of sperm RNA has revealed their influence in sperm functions and early embryo development, positioning them as promising markers for the diagnosis of male infertility. This systematic review aimed to evaluate the potential of differentially expressed sperm mRNA associated with infertility to improve diagnostic accuracy, and to identify predictive biomarkers for successful assisted reproductive technology (ART) outcomes. A comprehensive search of the MEDLINE-PubMed database was conducted from inception to 28 April 2023, and a total of 67 eligible articles were selected. Data on 451 genes were extracted and analysed using a custom scoring system, revealing a strong association between altered seminal parameters and the expression of nine mRNA: *AKAP4, DDX4, PGK2, PIWIL1, PRM1, PRM2, TNP1, TNP2* and *PLCZ1*. Notably, aberrant expression of *PLCZ1, PRM1, PRM2* and *PIWIL1* was closely linked with reduced ART success rates. These findings suggest that these genes constitute a panel of the most prominent biomarkers of broader transcriptomic changes, potentially reflecting disruptions in spermatogenesis. Given the complexity of diagnosing male fertility at a molecular level, integrating sperm transcriptome analysis with conventional semen assessments could enhance fertility predictions, although standardization of methodologies is essential to improve the reliability of the results.

INTRODUCTION

Infertility has emerged as one of the most widespread health issues of the 21st century, with well-documented evidence that male and female factors contribute equally (*Vander Borght and Wyns, 2018*). Focusing on male infertility, assessment of its prevalence in the general population remains challenging (*Barratt et al., 2017*), yet its incidence has been increasing steadily in recent years. This increase is largely attributed to factors such as delayed parenthood, obesity, substance abuse, and exposure to

Genetics of Male Fertility Group, Unitat de Biologia Cellular (Facultat de Biociències), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

*Corresponding author. E-mail address: joan.blanco@uab.cat (J. Blanco). https://doi.org/10.1016/j.rbmo.2025.104993 1472-6483/

Declaration: The authors report no financial or commercial conflicts of interest.

KEY WORDS

Assisted reproductive technology
Biomarker
Gene prioritization
Male infertility
Sperm mRNA

environmental pollutants and toxins (*Vander Borght and Wyns, 2018*).

Male infertility, which arises from a variety of factors including reproductive tract abnormalities, hormonal imbalances, inflammation of the reproductive system, and genetic factors, is a complex condition which exhibits high phenotypic variability. In some patients, assisted reproductive technology (ART) is employed to address infertility, yet success rates remain relatively low, with only approximately 30% of IVF/intracytoplasmic sperm injection (ICSI) cycles resulting in a live birth (*Wyns et al., 2020*). This low success rate, largely due to the poor quality of gametes and embryos, and the limitations in selecting high-quality cells, results in substantial emotional and financial costs for affected couples.

Current diagnostic methods for identifying male infertility primarily involve assessing seminal parameters such as sperm count, morphology and motility. However, these parameters often fail to predict ART outcomes, including pregnancy rate, leading to questions about their reliability (*Del Giudice et al., 2022*). Additionally, semen analyses offer little diagnostic insight for patients with normal seminal values, who represent up to 30% of those seeking infertility treatment (*Ray et al., 2012*). This indicates that routine semen analyses may not always detect the underlying causes of infertility. In this context, while the molecular causes of abnormal seminal parameters are relatively well understood in severe cases of male infertility, they often remain unidentified in milder instances, leaving many cases undiagnosed (*Punab et al., 2017*). Consequently, the search for molecular biomarkers for the diagnosis and prognosis of male infertility has been a prominent focus in reproductive biology since the advent of ART. The lack of molecular evaluations for diagnosis complicates the process of counselling couples about the potential success of ART and the reproductive health of their future children.

In the early 2000s, the study of sperm RNA emerged as a promising avenue for enhancing the diagnostic capabilities of semen analysis. Over the years, accumulated research has revealed that mature human spermatozoa contain a complex population of RNA, including both coding (mRNA) and non-coding RNA, which play crucial roles in sperm function and early embryo development (*Corral-Vazquez et al., 2021*). Given the important functions of sperm mRNA, research has explored their potential as molecular markers of seminal abnormalities and predictors of ART outcomes. Several studies have identified altered gene expression in sperm mRNA from infertile patients, suggesting that these alterations could be a potential cause of infertility. However, the clinical application of these findings remains limited due to a lack of sufficient conclusive evidence, which hinders their adoption in routine medical practice. The complexities of sperm mRNA expression and the variability of the results make it challenging to establish definitive correlations between these genetic alterations and male infertility. As a result, more robust studies are needed to validate these findings before they can be used reliably for diagnostic or prognostic purposes in clinical settings.

This systematic review aimed to summarize the human sperm mRNA molecules linked to male infertility, and identify clinically useful biomarkers to improve its diagnosis. Additionally, it explored the potential of sperm mRNA as predictive indicators for ART outcomes. To accomplish this, genes selected from eligible articles were submitted to a gene scoring system to objectively assess the link between a specific change in gene expression and male infertility. Using the framework proposed for the categorization of gene variants associated with male infertility (*Oud et al., 2019*), levels of evidence were categorized as weak, moderate or strong, and their potential for incorporation in the clinical setting to enhance the diagnosis and prognosis of male infertility was evaluated.

MATERIALS AND METHODS

Registration

The present study and the corresponding search protocol were registered in the PROSPERO registry (http://www.crd.york.ac.uk/PROSPERO) as PROSPERO CRD42024547710.

Search strategy

A systematic search of the literature was undertaken using the MEDLINE-PubMed database from its inception to 28 April 2023. The analysis was performed following the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (*Page et al., 2021*). This study employed the Population, Intervention, Context, and Outcome (PICO) framework to ensure systematic alignment of the study selection with the research question. The population (P) of interest comprised human male individuals diagnosed with infertility, encompassing all related conditions. The intervention (I) assessed across the studies was sperm RNA analysis, used to extract data of mRNA expression profiles. The comparison (C) group was fertile individuals who served as a refence control within each study. The outcomes (O) focused on identifying and characterizing differentially expressed transcripts within populations. Accordingly, the search strategy employed a combination of terms related to male infertility, transcriptomic analysis, seminal alterations, and ART outcomes. The search was restricted to case reports, classical articles, clinical studies, clinical trials, journal articles, letters, multicentre studies, observational studies and randomized controlled trials written in English and involving human males (Supplemental Table 1).

Study selection

The titles and abstracts of all articles were screened independently for eligibility by at least two specialists in male infertility. Articles that analysed mRNA molecules in semen and testicular tissue from human infertile patients were included. Conversely, non-human studies, studies lacking original data, and studies that did not meet the scope of the review (describing mRNA profiles in patients with known genetic or hormonal causes of male infertility, or studies focusing on RNA types other than mRNA) were excluded. Any discrepancies encountered during the primary screening were resolved through joint assessment of the article by the specialists.

Data extraction

Following the primary screening, the full text of the selected articles was obtained, and the same specialists extracted the data independently. For each article, the analysed genes were annotated, and their HUGO Gene Nomenclature Committee name, Online Mendelian Inheritance in Man (OMIM) number, chromosome location, GeneCards link, and OMIM male-infertility-associated phenotype were recorded. Furthermore, patient details such as the biological sample analysed, type of infertility, and the number and age of infertile individuals, along with their corresponding controls, were also

extracted for each gene. Data extraction also encompassed methodological aspects including sperm isolation technique, RNA isolation method, use of DNase treatment, RNA quality control procedures, mRNA analysis technique, data normalization, expression quantification approaches, statistical tests applied, and the result of the expression analysis (up- or down-regulation). Any disagreements in the extracted data were resolved through discussion and consensus between the analysers.

After data extraction, each article was categorized into either the semen group or the testicular group. Within the semen group, further distinction was made between studies conducted on spermatozoa and studies conducted on seminal plasma. This review conducted a detailed analysis on articles within the spermatozoa category.

Gene scoring

An internal scoring system aimed at assessing the strength of evidence regarding the relationship between each gene and male infertility was developed (TABLE 1). This system incorporates various criteria related to the genes themselves, the number of individuals and studies analysing the genes, and specific methodological parameters used in the analysis. Accordingly, one point was assigned to those genes that have been proven to display ubiquitous expression in human spermatozoa based on previously published research (*Corral-Vazquez et al., 2021*), thus ensuring reliable detection in prospective analyses. Genes linked to functions related directly to male fertility (according to the GeneCards database) or associated with a male infertility phenotype (according to the OMIM database) were given an additional point. The total number of patients and controls included in the analysis was considered by calculating the 95% CI for the number of individuals evaluated for every specific gene. Accordingly, a specific gene received two points when the total number of individuals was higher than the upper limit of the CI, the gene received one point when the total number was within the CI, and the gene received no points when the total number was lower than the lower limit of the CI. These calculations were performed separately for patients and controls. Regarding the number of articles, two points were assigned to genes reported by more than two independent studies, one point was assigned for genes reported by two studies, and no points were assigned for genes reported by a single study. The scoring system also considered methodological aspects of the experimental design, such as sperm isolation techniques, RNA quality controls and DNase treatments. For each gene, one point per treatment was assigned if all published studies applied the procedure. If not, the score for each procedure was adjusted based on the percentage of articles that included the treatment. The uniformity of the reported alterations was also considered in the scoring. For those genes with significantly altered expression detected by multiple evaluations, a maximum score of one point was given when all studies agreed on the dysregulation direction (either up- or down-regulation). In the case of discrepancy, the score was adjusted to reflect the highest percentage of consensus (whether up or down) across all studies conducted. If only one comparison between patients and controls existed, uniformity in altered expression received 0.5 points. Finally, the presence of

TABLE 1 SCORING CRITERIA FOR EVALUATING EVIDENCE OF CHANGES IN GENE EXPRESSION ASSOCIATED WITH MALE INFERTILITY

Criterion	Scoring
Ubiquitous sperm expression	1 point
Gene function directly related to male infertility (Genecard)	1 point
Male-infertility-associated phenotype (OMIM)	1 point
Number of patients	>Upper limit CI: 2 points Within CI: 1 point <Lower limit CI: 0 points
Number of controls	>Upper limit CI: 2 points Within CI: 1 point <Lower limit CI: 0 points
Number of articles	More than two studies: 2 points Two studies: 1 point One study: 0 points
Sperm isolation	Values range between 0 and 1, depending on the percentage of studies that apply the method
DNase I	Values range between 0 and 1, depending on the percentage of studies that apply the method
RNA quality control	Values range between 0 and 1, depending on the percentage of studies that apply the method
Altered expression uniformity	If there is only one altered expression: 0.5 points For more than one comparison: 1 point. Values range between 0 and 1, adjusted according to the percentage of coincident changes (up- or down-regulated) relative to the total (up, down, no change)
Supportive functional studies	1 point
Maximum score	14 points

supportive functional results confirming the association between differential expression and male infertility was awarded an additional point. Overall, genes in sperm studies could attain a maximum score of 14 points.

Gene scores were ranked from highest to lowest, and the values were subsequently categorized into 10 distinct binned intervals (bins). For this categorization, the use of equal-width intervals was excluded, ensuring that the bins were not evenly distributed but were based on the distribution of the data. Genes in Bins 10, 9 and 8 were classified as having strong, moderate and weak association with male infertility, respectively. These categorized gene sets were analysed for functional significance using Gene Ontology analysis in DAVID v.2021. The entire *Homo sapiens* gene set provided by the database served as the background. Gene Ontology terms were considered to be significantly enriched if they exhibited Bonferroni-corrected *P*-values <0.05.

RESULTS

Search strategy and study selection

The overall search strategy and selection process are detailed in FIGURE 1. The search yielded a total of 559 publications spanning from 1982 to 2023. Based on title and abstract analysis, 358 studies were excluded as they did not align with the scope of the review, they were conducted on non-human subjects, or they lacked original data.

Subsequently, the full text of the remaining 201 publications was sought, but two articles were excluded due to inaccessible full text. The remaining 199 articles underwent thorough analysis for data extraction. Of these, 98 articles were deemed ineligible for inclusion because their results were not directly relevant to the review, they presented no original data, or the experimental design was unsuitable.

Thus, a final list comprising 101 records was compiled for data extraction. Among these, 69 were attributed to the semen group, and 32 were attributed to the testicular group. The semen group was further divided into studies conducted on spermatozoa ($n = 67$) and studies conducted on seminal plasma ($n = 2$). For the purposes of this article, only the findings pertaining to the sperm group were considered.

Data extraction and analysis

Supplemental Table 2 summarizes the results of data extraction from the 67 eligible articles. The extracted data were analysed to identify the primary objectives of the published studies, assess the application of key transcriptomic analysis procedures, and identify the genes associated with male infertility, along with their levels of evidence.

Primary objective analysis

Most of the studies performed in the sperm fraction aimed to analyse the differential mRNA cargo between patients with altered seminal parameters and their corresponding controls (48/67 studies, ~72%) (Supplemental Table 2, Column J). Accordingly, these studies focused primarily on identifying differentially expressed mRNA that could explain, at least in part, the seminal alteration observed in the infertile patients. Besides, approximately 15% of the comparisons (10 studies) aimed to associate male infertility phenotypes with the outcomes of ART treatment, mainly fertilization and pregnancy rates. Such studies were principally designed to identify biomarkers able to predict the results of IVF/ICSI treatments. Some additional studies (6/67 studies, ~9%) %) tried to discern the impact of substance use on male infertility by investigating the differential mRNA expression in spermatozoa from smokers ($n = 5$) and heroin users ($n = 1$). The other four studies (~6%) delved into the analysis of spermatogenic impairments attributed to varicocele, aiming to elucidate the molecular mechanisms of the spermatogenic defects resulting from the enlargement of scrotal veins. Furthermore, two more studies (3%) were dedicated to elucidating the consequences of an altered DNA fragmentation index, aiming to understand its implications on sperm quality and its potential role in male infertility. Lastly, five studies (~7%) focused on individuals with less common or heterogeneous infertility phenotypes, encompassing a variety of clinical situations not covered by the other categories.

Methodological evaluation

Sperm transcriptomic analyses are challenging due to several factors, such as the presence of non-sperm cells in the ejaculate, the inherently fragmented and low-concentration nature of sperm RNA, and a reduced number of gametes in certain male infertility cases. Furthermore, most procedures associated with RNA isolation and data analysis have been optimized for other cell types, making their application in spermatozoa more challenging. These issues underscore the importance of the experimental design in any work addressed to analyse these molecules in sperm samples.

Sperm purification should be mandatory when analysing ejaculated spermatozoa, as semen samples may also contain somatic cells and germ cells. Even low levels of non-sperm-cell contamination are problematic because these cells have 100- to 200-fold more RNA than sperm, significantly skewing the proportion of sperm-specific RNA (*Cappallo-Obermann and Spiess, 2016*). Therefore, efficient elimination of non-sperm cells is crucial in sperm studies before the application of RNA isolation procedures. Surprisingly, 23 of the 67 eligible articles (34.3%) did not provide any information about sperm isolation methods, raising serious doubts about the reliability of their results (Supplemental Table 2, Column P). The remaining 44 studies employed three different methods: density gradient centrifugation ($n = 32$), swim-up ($n = 7$), and somatic cell lysis (SCL) ($n = 5$). SCL is particularly advantageous for samples with low motility and low sperm count, where other methods may fail to recover sufficient spermatozoa for downstream procedures. However, gradient-based methods were the preferred choice in the selected studies, likely due to their ability to isolate spermatozoa with the highest fertilizing potential — a key feature in ART — despite the fact that these spermatozoa do not fully represent the entire sperm sample. Additionally, gradient-based methods offer commercially standardized procedures, ensuring consistency and minimizing subjectivity, which further contributes to their widespread use over SCL.

Concerning RNA purification, two main strategies have been applied in the articles selected in this review: organic-based methods (44.8%) and column-based methods (53.7%) (Supplemental Table 2, Column Q). In two studies (3%), a combination of the two methods was used. Certainly, the choice of method depends on the specific requirements of the experiment, such as the desired purity and yield of RNA, the starting material, the downstream applications, and the specific expertise of the research team. Organic-based methods are known to yield high RNA purity and quantity, and to remove protein contaminants efficiently. However, they involve handling toxic chemicals, are time-consuming, and require multiple

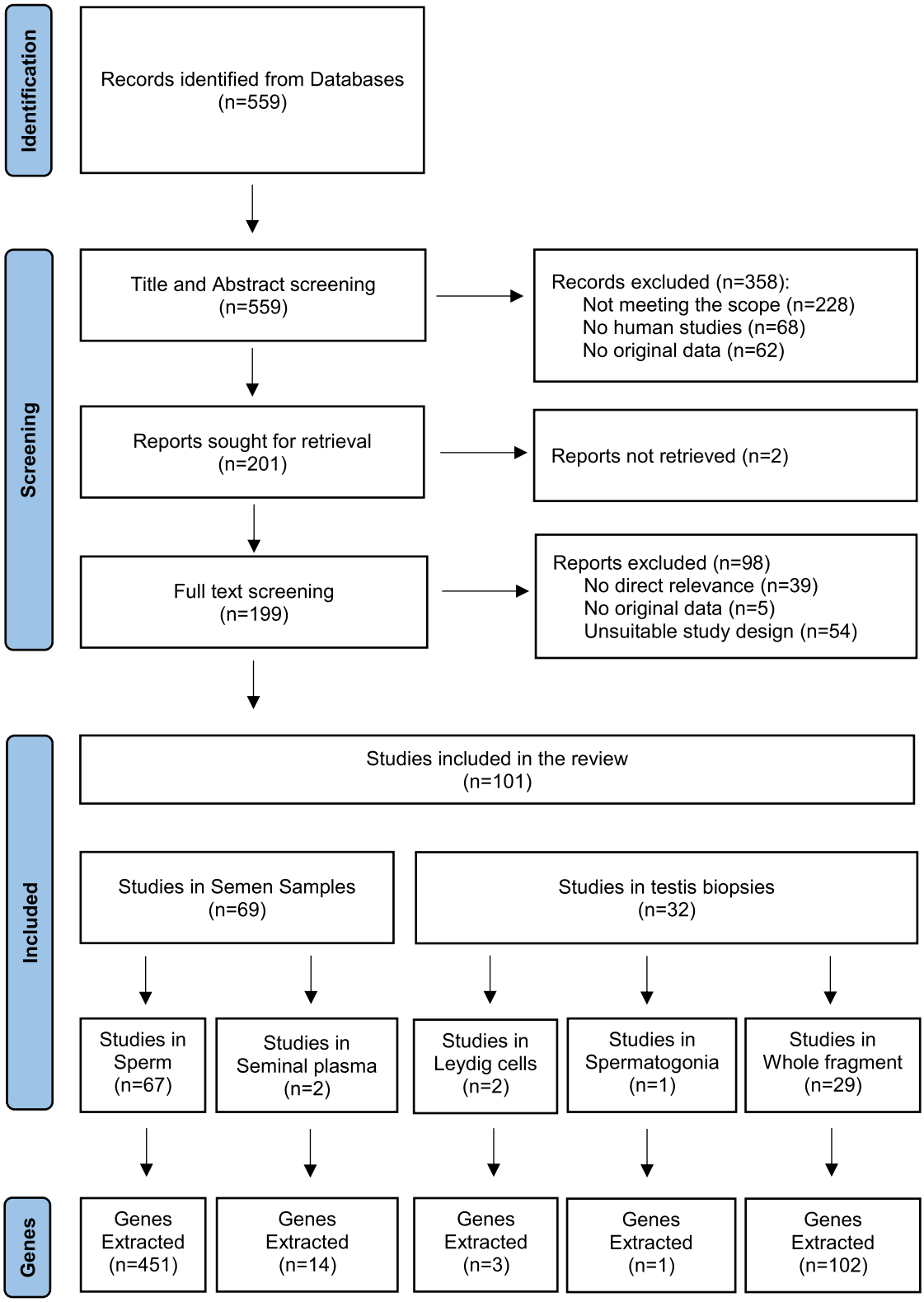

FIGURE 1 Flow chart showing the search and screening strategy to identify eligible publications and sperm mRNA transcripts associated with male infertility.

steps, which may increase the handling-related risk of RNA degradation without proper precautions. Column-based methods, on the other hand, are quick and easy to use with consistent and reproducible results, making them suitable for small samples. While these methods can sometimes yield less RNA than organic methods, this can vary depending on the sample type and protocol. Considering the low RNA content and small input of sperm samples, exploring alternative methods is

advisable. Based on the authors' experience, magnetic-bead-based methods can offer high sensitivity and efficiency in sperm RNA isolation (*Corral-Vazquez et al., 2021*). These methods facilitate automation and scalability when specialized equipment is available, although they may be more expensive due to equipment costs.

Additionally, it has been well documented that the method of RNA isolation can significantly affect the identification of differentially expressed transcripts, with different techniques potentially yielding varying results (*Barragán et al., 2015*; *Scholes and Lewis, 2020*). Therefore, it is advisable to avoid comparing results obtained using different RNA isolation methodologies. In this review, given the considerable variability in both the methods and specific kits used for extraction, the decision was made not to segment the results based on this variable. However, this remains an open question that should be addressed in future research. Any laboratory applying the findings from this review should be mindful of the critical impact that RNA isolation methods can have on the outcomes.

Assessing RNA quality before conducting RNA studies is crucial to guarantee the robustness of the results. In this sense, ensuring the absence of DNA in the final sample is a primary concern. Accordingly, the application of a DNase treatment to remove any trace of contaminating genomic DNA is highly recommended, as DNA contamination can interfere with the detection and accurate quantification of some transcripts, leading to misleading results. This step is particularly critical for targeted sperm RNA studies, which have very low amounts of RNA compared with other cell types, so even minimal DNA contamination can compromise the purity of the RNA sample. In this context, the analysis of the eligible articles has yielded surprising results (Supplemental Table 2, Column R). More than half of the articles (41/67 studies, 61.2%) did not apply or describe the application of a DNase treatment. This generalized omission undermines the consistency and robustness of the particular and global findings from these studies.

It is important to note a limitation of RNA studies in spermatozoa related to the quality assessment of these molecules. RNA quality is usually determined by the RNA integrity number (RIN), which estimates integrity based on the ratio of 28S to 18S ribosomal RNA. The RIN is presented as a numerical value from 1 to 10, where 10 represents intact RNA and 1 indicates highly degraded material (*Schroeder et al., 2006*). However, this value is useless in sperm transcriptomic studies since human sperm lacks intact molecules of ribosomal RNA. Accordingly, alternative methods to assess RNA quality should be considered for sperm RNA studies (Supplemental Table 2, Column S). One option is microfluidic-based RNA analysis using capillary electrophoresis. This method enables the identification of degradation patterns, a crucial aspect of evaluating sperm RNA sample quality. However, only a small proportion of eligible articles applied this method (5/67 studies, 7.5%). Another possibility is the visualization of bulk RNA using electrophoresis; this method was employed in a limited number of studies (4/67 studies, 6%). Alternatively, some studies focused on identifying specific sperm RNA transcripts (3/67 studies, 4.5%), while others targeted transcripts from non-sperm cells (7/67 studies, 10.4%). Although the primary aim in all cases is to assess RNA integrity, the analysis of non-sperm-cell transcripts also offers insights into the efficacy of the sperm isolation method utilized. In any case, electrophoresis-based methodologies are not direct or quantitative methods of analysis. Therefore, it is highly recommended to use them in conjunction with other techniques for a more comprehensive assessment of RNA quality. Overall, it is concerning that a substantial majority of studies (48/67 studies, 71.6%) lack any procedure for assessing RNA quality. This highlights a critical gap in current research practices that needs to be addressed to ensure the reliability and reproducibility of findings in sperm RNA studies.

Regarding the method of analysis (Supplemental Table 2, Column T), targeted polymerase chain reaction (PCR) amplification is the technique of choice for most of the eligible articles (64/67 studies, 95.5%). In these types of studies, the expression evaluation of a gene of interest relies on the use of normalizers, which are reference molecules with a supposed invariant expression across samples. Normalizers are used to adjust for technical variability, and ensure accurate quantification of target gene expression (*Peltier and Latham, 2008*). Data normalization guarantees that the measured gene expression levels reflect the true biological differences accurately. Accordingly, unstable reference genes can alter the expression pattern significantly, and introduce errors into the results (*Ferguson et al., 2010*). Therefore, identifying appropriate reference normalizers is crucial in quantitative reverse transcription PCR assays. In the sperm RNA studies selected in this review, various normalizers have been employed (Supplemental Table 2, Column V). *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) was used most frequently, with 30 articles describing its use, accounting for 44.8% of the total. *ACTB* (beta-actin) was used in 22 studies, representing 32.8%. *HBB* (hamoglobin subunit beta) was used in three studies, comprising 4.5%. *PRM2* (protamine 2) was used in two studies, equating to 3%. *B2M* (beta-2-microglobulin) was used in one study as a sole normalizer (1.5%), although it served as a secondary normalizer in most studies that combined multiple molecules ($n = 4$, 6%).

In any case, it is essential to determine whether these molecules exhibit invariant expressions across samples. To investigate this, the list of sperm normalizers from this review was compared with the list of ubiquitously expressed mRNA transcripts identified by the authors' team through the analysis of sperm RNA from a population of fertile control donors (*Corral-Vazquez et al., 2021*). The results were very clear: only *PRM2* and *SAP130* were expressed consistently across all samples. The most used transcripts, *GAPDH* and *ACTB*, were not present in all samples, raising serious doubts about their reliability as normalizers. This confirms previous concerns raised by other researchers (*Barragán et al., 2015*), highlighting the need for careful consideration when selecting reference genes.

Genes and male infertility

Following analysis of the 67 eligible articles, data from 451 genes were extracted (**FIGURE 1**; Supplemental Table 2) and evaluated using the scoring system (Supplemental Table 3). As mentioned, the goal was to assess the level of evidence between changes in gene expression and male infertility. Nine genes with a strong association (Bin 10), 31 genes with a moderate association (Bin 9), and 73 genes with a weak association (Bin 8) with male infertility were identified. The remaining genes ($n = 338$) were excluded from further analyses because of their limited evidence

of association with male infertility based on the scoring system (Supplemental Table 3).

Overview of the biological functions associated with genes in the top bins (Bins 8, 9 and 10)

The list of 113 genes, categorized as strong, moderate and weak, was subjected to Gene Ontology analysis to assess functional significance. A total of 18 significantly enriched Gene Ontology terms were identified (Supplemental Table 4A). Of them, 14 terms appeared to be related to key biological processes in reproductive biology: spermatogenesis ($n = 1$); sperm motility ($n = 8$); sperm capacitation ($n = 1$); spermatid development ($n = 3$); and fertilization ($n = 1$). This association should not be entirely surprising given that the list of genes was obtained from articles employing targeted strategies, which were selected for their known association with male infertility. However, it is important to emphasize the robustness of these associations, as indicated by the low *P*-values of the enriched processes.

To further investigate these findings, the remaining 338 genes from Bins 1–7 were submitted to Gene Ontology analysis to see if they also enriched the same processes with similar significance (Supplemental Table 4B). In this case, 12 Gene Ontology terms were identified as enriched. However, they were notably more diverse compared with the Gene Ontology terms associated with Bins 8–10, with only four genes linked directly to processes relevant to sperm biology.

These findings not only validate the criteria used for gene selection in this work, but also emphasize the functional significance of the genes included in Bins 8–10. The relationship between the enriched Gene Ontology terms and male-fertility-related processes highlights the potential of these genes to serve as biomarkers for diagnosing and understanding male infertility.

Association between altered gene expression and seminal parameters

As mentioned previously, most studies conducted on the sperm fraction aimed to identify differentially expressed mRNA that could explain the seminal alterations observed in patients. Among the differentially expressed genes identified, 179 were associated with oligozoospermia (patients exhibiting at least a reduced sperm count among other seminal alterations), 240 were associated with asthenozoospermia (patients exhibiting at least reduced sperm motility), and 10 were associated with teratozoospermia (patients exhibiting at least increased sperm morphological alterations). Additionally, although not classified as patients with seminal alterations, nine genes were associated with infertile normozoospermic patients (unexplained male infertility) (Supplemental Table 5).

To identify which of these genes have the potential to serve as biomarkers for alterations in sperm count, motility and/or morphology, the lists were filtered further to include only those genes whose alterations were strongly associated with male infertility (i.e. they were also present among the nine genes included in Bin 10) (TABLE 2).

Seven genes showed a strong association with both oligozoospermia and asthenozoospermia, likely due to the simultaneous occurrence of alterations in sperm count and motility in most patients (Supplemental Table 5). Among these genes, two (*AKAP4* and *PGK2*) were identified as crucial for sperm motility, while one gene product was essential for maintaining germline integrity (*DDX4*). Additionally, four genes (*TNP1, TNP2, PRM1* and *PRM2*) were involved in organizing the structure of sperm chromatin (TABLE 2).

Focusing on genes that impact a single sperm parameter, only one gene (*PIWIL1*), was associated specifically with asthenozoospermia (Supplemental Table 5). PIWIL1 plays a crucial role in repressing transposable elements, which is essential for maintaining genome integrity in the germline (TABLE 2). However, its link to reduced sperm motility remains unclear, as its primary function is more related to genomic stability than directly influencing motility. In fact, this gene has also been linked to cases of unexplained male infertility, adding further uncertainty to its association with specific male infertility conditions.

Finally, *PLCZ1* was associated specifically with teratozoospermia (Supplemental Table 5). PLCZ1 plays a crucial role in fertilization and early embryonic development by inducing oocyte activation, which triggers calcium oscillations essential for the initial stages of embryo development (TABLE 2).

The analysis of gene functions discussed above highlights potential connections between specific genes and particular sperm abnormalities. However, these associations were not observed consistently across all examined genes, raising important questions about the link between changes in gene expression and the development of specific sperm defects. This suggests that sperm abnormalities may not result from straightforward, cause-and-effect relationships with single genes. Instead, they likely arise from a complex interplay of multiple factors. This intricate issue will be explored below.

Association between altered gene expression and ART outcome

This study also aimed to investigate whether specific changes in gene expression were linked to lower ART success rates. As mentioned, some of the comparisons in the eligible articles revealed changes in gene expression associated with lower ART success rates (Supplemental Table 2, Column J). These studies primarily aimed to identify biomarkers that could predict IVF/ICSI outcomes. From the review of these articles, 15 genes associated with a low fertilization rate, six genes associated with abnormal embryo development, 82 genes associated with a reduced pregnancy rate, and one gene associated with recurrent miscarriage were identified. To determine which of these genes could serve as biomarkers for changes in ART outcomes, the list was further refined to include only those genes with strong associations with male infertility (i.e. they were also present among the nine genes included in Bin 10) (Supplemental Table 5).

In IVF/ICSI-derived zygotes, fertilization is typically assessed 16–20 h after sperm entry by observing the presence of two pronuclei and two polar bodies. Among the top bin (Bin 10), altered expression of four genes – *PLCZ1, PRM1, PRM2* and *PIWIL1* – was specifically associated with a lower fertilization rate (Supplemental Table 5). As noted previously, PLCZ1 is a crucial sperm factor for oocyte activation and early embryo development. Additionally, proper expression of *PRM1* and *PRM2* is essential for sperm chromatin formation; their dysregulation can hinder the replacement of protamines with histones after fertilization, which is critical for formation of male pronuclei and can therefore impact the success of fertilization. Furthermore, PIWIL1 plays a key role in maintaining genome integrity in germ cells and regulating gene expression

TABLE 2 GENES WITH A STRONG (BIN 10) ASSOCIATION WITH INFERTILITY ACROSS DIFFERENT POPULATIONS THAT COULD SERVE AS POTENTIAL MOLECULAR BIOMARKERS BASED ON THEIR SCORING

Gene symbol	Gene name	Molecular function	Score	Semen analysis				ART success			Var	Subst users	Reference
				Oligo	Astheno	Terato	UMI	Fert	Embr dev	Preg			
AKAP4	A-kinase anchoring protein 4	Plays a role in sperm motility	11.66	Yes	Yes	-	-	-	-	-	-	-	*Abu-Halima et al., 2023*; *Montjean et al., 2012*; *Sadakierska-Chudy et al., 2020*
DDX4	DEAD-box helicase 4	Required to repress transposable elements, maintaining germline integrity	11.50	Yes	Yes	-	-	-	-	-	-	-	*Abu-Halima et al., 2023*; *Guo et al., 2007*; *Montjean et al., 2012*; *Sadakierska-Chudy et al., 2020*
PIWIL1	PIWI-like RNA-mediated gene silencing 1	Plays a role in repressing transposable elements, crucial for germline integrity	11.35	-	Yes	-	Yes	Yes	-	-	-	-	*Cheung et al., 2019*; *Giebler et al., 2018*; *Giebler et al., 2021*
TNP1	Transition protein 1	Plays a key role in the replacement of histones to protamines in the elongating spermatids of mammals	11.32	Yes	Yes	-	-	-	-	-	-	Yes	*Abu-Halima et al., 2023*; *Amor et al., 2021*; *Jedrzejczak et al., 2007*;
TNP2	Transition protein 2	Plays a key role in the replacement of histones to protamine in the elongating spermatids of mammals	11.32	Yes	Yes	-	-	-	-	-	-	Yes	*Abu-Halima et al., 2023*; *Amor et al., 2021*; *Jedrzejczak et al., 2007*
PLCZ1	Phospholipase C zeta 1	Involved in inducing oocyte activation and initiating embryonic development up to the blastocyst stage	11.10	-	-	Yes	-	Yes	-	-	Yes	-	*Aghajanpour et al., 2011*; *Amor et al., 2021*; *Janghorban-Laricheh et al., 2016*; *Tavalaee et al., 2018*; *Tavalaee and Nasr-Esfahani, 2016*
PRM2	Protamine 2	Compacts sperm DNA into a stable, inactive complex	11.06	Yes	Yes	-	-	Yes	Yes	-	-	Yes	*Depa-Martynów et al., 2007*; *Hamad, 2019*a; *Hamad et al., 2019*b; *Jodar et al., 2012*; *Kempisty et al., 2007*; *Montjean et al., 2012*; *Sadakierska-Chudy et al., 2020*; *Shabani Nashtaei et al., 2018*
PGK2	Phosphoglycerate kinase 2	Essential for sperm motility and male fertility	10.99	Yes	Yes	-	-			-	-		*Abu-Halima et al., 2023*; *Liu et al., 2018*; *Shen et al., 2013*
PRM1	Protamine 1	Compacts sperm DNA into a stable, inactive complex	10.89	Yes	Yes	-	-	Yes	Yes	-	-	Yes	*Amor et al., 2021*; *Depa-Martynów et al., 2007*; *Hamad et al., 2019*b; *Jodar et al., 2012*; *Kempisty et al., 2007*; *Shabani Nashtaei et al., 2018*; *Sadakierska-Chudy et al., 2020*

ART, assisted reproductive technology; Oligo, oligozoospermia; Astheno, asthenozoospermia; Terato, teratozoospermia; UMI, unexplained male infertility; Fert, fertilization; Embr dev, embryo development; Preg, pregnancy; Var, varicocele; Subst users, substance users.

during spermatogenesis (TABLE 2). As gamete quality is fundamental to the quality and development of the zygote, alterations in the expression of these genes could impair sperm quality and potentially reduce the fertilization rate.

Altered expression of *PRM1* and *PRM2* was also associated with altered embryo development (Supplemental Table 5). Given that a reduction in fertilization logically leads to decreased embryo development and/or a lower pregnancy rate, the observed negative effects on these outcomes may simply result from a diminished fertilization rate, although a more direct influence on these variables cannot be ruled out.

Association between altered gene expression and varicocele

This analysis found differentially expressed genes in spermatozoa from patients with varicocele, a condition known to have an adverse effect on semen quality and sperm function, potentially influencing ART outcomes (*Jensen et al., 2017*). A total of nine genes exhibited differential expression in these patients (Supplemental Table 5). Among them, *PLCZ1* was ranked in the top bin (Bin 10), indicating a strong association with varicocele. Given that varicocele is associated with altered semen parameters, the observed changes in gene expression may indicate spermatogenic defects resulting from impaired testicular blood flow. This impairment can have an adverse effect on spermatogenesis and disrupt crucial genes essential for the functionality of spermatozoa.

Association between altered gene expression and substance users

A limited number of studies have investigated the impact of substance use on male infertility by analysing differentially expressed transcripts in spermatozoa. Among the 11 genes identified in these studies (Supplemental Table 5), four were strongly associated (Bin 10) with sperm chromatin organization: *TNP1, TNP2, PRM1* and *PRM2* (TABLE 2). These findings suggest that these genes could serve as potential molecular biomarkers for assessment of the detrimental effects of substance consumption on male fertility.

DISCUSSION

This review provides a comprehensive analysis of high-quality research examining the relationship between altered sperm mRNA expression and male infertility. Following application of the gene scoring system, a strong association was identified between changes in seminal parameters and the expression of nine specific sperm mRNA: *AKAP4* (*Abu-Halima et al., 2023*; *Montjean et al., 2012*; *Sadakierska-Chudy et al., 2020*); *DDX4* (*Abu-Halima et al., 2023*; *Guo et al., 2007*; *Montjean et al., 2012*; *Sadakierska-Chudy et al., 2020*); *PIWIL1* (*Cheung et al., 2019*; *Giebler et al., 2018*); *TNP1* (*Abu-Halima et al., 2023*; *Jedrzejczak et al., 2007*); *TNP2* (*Abu-Halima et al., 2023*; *Jedrzejczak et al., 2007*); *PLCZ1* (*Aghajanpour et al., 2011*; *Janghorban-Laricheh et al., 2016*; *Tavalaee et al., 2018*; *Tavalaee and Nasr-Esfahani, 2016*); *PRM2* (*Hamad, 2019*; *Jodar et al., 2012*a; *Kempisty et al., 2007*; *Montjean et al., 2012*c; *Sadakierska-Chudy et al., 2020*c; *Shabani Nashtaei et al., 2018*); *PGK2* (*Abu-Halima et al., 2023*; *Liu et al., 2018*; *Shen et al., 2013*); and *PRM1* (*Jodar et al., 2012*b; *Kempisty et al., 2007*b; *Sadakierska-Chudy et al., 2020*d; *Shabani Nashtaei et al., 2018*). Furthermore, the aberrant expression of four of these transcripts was strongly correlated with reduced ART success rates: *PIWIL1* (*Giebler et al., 2021*), *PLCZ1* (*Aghajanpour et al., 2011*), *PRM1* and *PRM2* (*Depa-Martynów et al., 2007*). Additionally, *PLCZ1* has been implicated in varicocele (*Janghorban-Laricheh et al., 2016*b), while transition proteins and protamines have been associated with substance users (*Amor et al., 2021*; *Hamad et al., 2019*).

All these genes encode proteins critical to male fertility, highlighting their potential as biomarkers for male reproductive health. Specifically, AKAP4 plays a crucial role in sperm motility by regulating flagellar function. DDX4 and PIWIL1 are involved in repressing transposable elements, maintaining genomic stability within the germline, and suggesting that alterations in these genes may impair sperm quality and fertilization success. TNP1 and TNP2 are key to chromatin remodelling during spermatogenesis, facilitating the replacement of histones with protamines in elongating spermatids, which is critical for sperm DNA packaging, condensation, and subsequent fertilization and embryonic development. PLCZ1 triggers oocyte activation and initiates embryonic development, making it a crucial factor for successful fertilization in ART. PRM1 and PRM2 compact sperm DNA into a stable, inactive form, maintaining DNA integrity, with dysfunction linked to poor ART outcomes. Finally, PGK2 is essential for sperm motility and male fertility through its role in ATP production, supporting sperm function, especially motility, which is critical for fertilization and ART success. These nine genes, as demonstrated by their functions, are key contributors to sperm functionality, and their altered expression is closely linked to changes in seminal parameters and ART outcomes, offering potential for future fertility diagnostics and therapeutic strategies.

Methodological considerations

In this review, data were collected from the experimental designs of various studies, which proved useful for gene classification. These data clearly revealed a high degree of heterogeneity between studies. While heterogeneity itself is not inherently negative — different techniques or methods can yield similar results — the issue arises when the divergences compromise critical aspects of the protocols that are objectively essential. In this sense, a substantial number of studies suffer from experimental shortcomings, such as the absence of proper sperm selection methods, DNase treatment, and quality controls of isolated sperm RNA fractions (see section on methodological evaluation). These deficiencies highlight the need for standardization in this type of research. In the future, only studies that meet these minimal methodological requirements should be considered for publication to ensure the reliability of findings.

This need for standardization becomes even more evident when considering the evolution of sperm RNA research in terms of protocols and technological advancements. Older studies likely employed less accurate and refined methodologies compared with more recent studies, although there are exceptions. While recent research benefits from continuous improvements in techniques, instrumentation and analytical tools leading to generally more precise and reliable results, it remains essential to claim rigorous quality standards. The scientific community, including authors, reviewers and editors, must ensure that robust quality control measures are implemented to address these limitations and uphold the integrity of the field.

Scoring system reliability

After the initial screening of eligible articles, 451 candidate genes were identified whose altered expression was described as potentially linked to male infertility. The large number of genes, combined with the notable heterogeneity

in experimental design — both in terms of study populations and experimental methodologies — highlighted the need to implement a system to curate and prioritize genes with the strongest associations. In this system, data from three distinct levels were incorporated: the functionality of the genes; the number of individuals and studies involved; and the quality of the experimental approaches applied in the different articles. The use of multiple levels of evidence is essential for establishing consistent associations between gene defects and male infertility.

Different pieces of evidence highlight the reliability of the implemented strategy. First, Gene Ontology analysis for Bins 8, 9 and 10 (genes with weak, moderate and strong associations, respectively) showed significant associations with processes related to sperm functionality. Conversely, analysis of Bins 1–7, which include genes with limited evidence of association, yielded different results as most of the enriched processes were not related to sperm functionality. These findings underscore the reliability of the scoring system implemented in identifying the most relevant genes.

Second, upon examining the functions of the genes classified within Bin 10, it becomes evident that all these genes were intricately connected to the processes of spermatogenesis and the functions of spermatozoa. These genes play critical roles in various stages of sperm development, as well as influencing the functional capabilities of sperm cells, such as motility and fertilization (TABLE 2).

Third, the curated list of genes was cross-referenced with those identified by *Oud et al. (2019)* as having a pathogenic gene variant associated with a male infertility phenotype. For this comparison, the list of genes with at least a moderate association in both studies was used. Five genes — *CFAP43, CFAP44, CFAP69, DNAH1* and *PLCZ1* —appeared on both lists, which represents only 12.5% of the genes selected in the present study (5/40 genes) (Supplemental Table 6). While this percentage may seem modest, it is important to note that this systematic review specifically targeted studies involving patients with isolated infertility, without additional phenotypic manifestations. In contrast, the study by *Oud et al. (2019)* encompassed a broader spectrum of infertility types, including syndromic infertility and endocrine disorders. To refine the comparison, the focus was narrowed to the 15 genes reported by *Oud et al. (2019)* as having gene variants in patients with isolated infertility. In this more targeted analysis, five of the 15 genes (33.33%) overlapped with the list from the present study (Supplemental Table 6). This comparative approach holds great potential for deepening understanding of the pathophysiology of male infertility. The identification of common genes through different approaches suggests the existence of converging intersections — shared dysregulated processes that may arise from distinct underlying mechanisms. This paves the way for the development of molecular markers, offering promising opportunities for improving the diagnosis and prognosis of male infertility.

Finally, the authors wish to highlight the flexibility of the model proposed, as well as its results. As the scoring system is built on evidence that can change over time as additional studies are conducted, genes can be reclassified from their original categorization. This adaptability ensures that the system stays current with new results, and enhances its accuracy continuously. Such flexibility is especially valuable for potential clinical applications, where updated data can improve future implementations of the results.

Cause or consequence?

A key question remains: is the altered expression of certain genes associated with male infertility the cause of the condition, or is it merely a consequence of the infertility phenotype? As discussed earlier, several differentially expressed genes are involved in biological functions linked directly to abnormal seminal parameters. For example, AKAP4 protein is localized in the sperm flagellum and regulates sperm motility, so its association with sperm motility is easy to establish. However, alterations in the expression of this gene have also been associated with oligozoospermia. Similarly, PLZC1 has been implicated with a lower fertilization rate, which aligns well with its known function. However, its association with teratozoospermia and varicocele is less intuitive, given the current understanding of its role. The most illustrative examples are the protamines, which were linked in this study to a range of infertility-related conditions, including oligozoospermia, teratozoospermia, impaired fertilization, embryo development issues, and substance use (TABLE 2). Supporting this interpretation, it is also remarkable that the top five genes ranked in Bin 9 are involved in alterations that extend beyond seminal parameters (TABLE 3). This suggests a more generalized effect rather than a specific effect.

Based on this analysis, it is proposed that alterations in the expression of specific sperm transcripts are the consequence of undefined disruptions in spermatogenesis. In other words, abnormal spermatogenesis influences the mRNA cargo of spermatozoa. In this scenario, the nine genes identified in Bin 10 would likely represent the 'tip of the iceberg' — the most prominent markers of more extensive transcriptomic changes driven by unknown underlying mechanisms.

Therefore, within the scope of this study, which sought to identify novel biomarkers for the diagnosis and prognosis of male infertility, it is suggested that the abnormal expression of these nine genes could serve as an indicator of a more generalized underlying condition. The same principle can be extended to other cases of altered gene expression, such as in patients with substance use or varicocele. While one might assume that the list of dysregulated genes is shorter in these cases, it is important to note that the number of studies — and thus, the genes analysed under these conditions — is fewer compared with those examining seminal alterations and ART outcomes. Therefore, sufficient data to confirm whether the list of affected genes is truly smaller are currently lacking. In these instances, detecting disruptions in the expression of this same set of nine genes may be indicative of broader reproductive issues, highlighting the potential of these markers as indicators of overall reproductive health.

What is next?

In the dairy breeding industry, predicting male fertility remains an important challenge. In recent years, alongside traditional seminal parameter analysis, factors such as sperm DNA integrity and transcriptome analysis have become increasingly common (*Turri et al., 2021*). Studies in different farm animals have shown that the expression of some sperm miRNA correlates with the fertility rate, making it a valuable tool for predicting male fertility potential (*Indriastuti et al., 2022*).

Similarly, the present authors believe that, in the future, the fertility potential of infertile patients should be evaluated using a combination of techniques. When integrated with semen quality assessments, these techniques could lead to a better

TABLE 3 GENES WITH A MODERATE (BIN 9) ASSOCIATION WITH INFERTILITY ACROSS DIFFERENT POPULATIONS THAT COULD SERVE AS POTENTIAL MOLECULAR BIOMARKERS BASED ON THEIR SCORING

Gene symbol	Gene name	Molecular function	Score	Semen analysis				ART success				
				Oligo	Astheno	Terato	UMI	Fer	Embr dev	Pre	Var	Subst users
WBP2NL	WBP2 N-terminal like	May contribute to meiotic resumption and pronuclear formation during fertilization	10.32	-	-	Yes	-	Yes	-	Yes	-	-
PIWIL2	PIWI-like RNA-mediated gene silencing 2	Plays a role in repressing transposable elements, crucial for germline integrity	10.16	Yes	-	-	-	Yes	-	-	-	-
GPX4	Glutathione peroxidase 4	Required for normal sperm development and male fertility	10.00	-	Yes	-	-	-	Yes	-	-	Yes
ROPN1	Rhophilin associated tail protein 1	Involved in fibrous sheath integrity and sperm motility. Required for spermatozoa capacitation	10.00	Yes	Yes	-	-	-	-	-	-	-
HSPA2	Heat shock protein family A (Hsp70) member 2	Plays a role in spermatogenesis. May participate in the maintenance of spindle integrity during meiosis in male germ cells	10.00	Yes	-	-	-	-	-	-	Yes	-
ANXA5	Annexin A5	Potential role in cellular signal transduction, inflammation, growth and differentiation	9.50	Yes	Yes	-	-	-	-	-	-	-
ARMC2	Armadillo repeat containing 2	Required for sperm flagellum axoneme organization and function	9.50	Yes	Yes	-	-	-	-	-	-	-
CATSPER1	Cation channel sperm associated 1	Central to sperm hyperactivation, acrosome reaction, and chemotaxis towards the oocyte.	9.50	Yes	Yes	-	-	-	-	-	-	-
CATSPER2	Cation channel sperm associated 2	Central to sperm hyperactivation, acrosome reaction, and chemotaxis towards the oocyte	9.50	Yes	Yes	-	-	-	-	-	-	-
CFAP251	Cilia and flagella associated protein 251	Involved in spermatozoa motility	9.50	Yes	Yes	-	-	-	-	-	-	-
CFAP43	Cilia and flagella associated protein 43	Involved in sperm flagellum axoneme organization and function	9.50	Yes	Yes	-	-	-	-	-	-	-
CFAP44	Cilia and flagella associated protein 44	Involved in sperm flagellum axoneme organization and function	9.50	Yes	Yes	-	-	-	-	-	-	-
CFAP65	Cilia and flagella associated protein 65	Plays a role in flagellar formation and sperm motility	9.50	Yes	Yes	-	-	-	-	-	-	-
CFAP69	Cilia and flagella associated protein 69	Required for sperm flagellum assembly and stability	9.50	-	-	-	-	-	-	-	-	-
CLU	Clusterin	Involved in spermatogenesis	9.50	Yes	Yes	Yes	-	-	-	-	-	-
DNAH1	Dynein axonemal heavy chain 1	Required for sperm flagellum motility	9.50	Yes	Yes	-	-	-	-	-	-	-
FSIP2	Fibrous sheath interacting protein 2	Plays a role in spermatogenesis	9.50	Yes	Yes	-	-	-	-	-	-	-

(*continued on next page*)

TABLE 3 (*Continued*)

Gene symbol	Gene name	Molecular function	Score	Semen analysis				ART success				
				Oligo	Astheno	Terato	UMI	Fer	Embr dev	Pre	Var	Subst users
GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	Plays an important role in energy production during spermiogenesis and sperm motility	9.50	Yes	Yes	-	-	-	-	-	-	-
HILS1	H1.9 linker histone, pseudogene	Implicated in chromatin remodelling during spermiogenesis	9.50	-	Yes	-	-	-	-	-	-	-
IQCG	IQ motif containing G	Required for normal axoneme assembly in sperm flagella, normal sperm tail formation, and male fertility	9.50	Yes	Yes	-	-	-	-	-	-	-
ODF2	Outer dense fibre of sperm tails 2	Seems to be a component of sperm tail outer dense fibres. May have a modulating influence on sperm motility	9.50	Yes	Yes	-	-	-	-	-	-	-
QRICH2	Glutamine rich 2	Essential role in the formation of sperm flagella and flagellar structure maintainance	9.50	Yes	Yes	-	-	-	-	-	-	-
SEPTIN12	Septin 12	Involved in the morphogenesis of sperm heads and the elongation of sperm tails	9.50	Yes	Yes	-	-	-	-	-	-	-
SLC26A8	Solute carrier family 26 member 8	Required for sperm motility and sperm capacitation. May play a role in sperm tail differentiation	9.50	Yes	Yes	-	-	-	-	-	-	-
SMCP	Sperm mitochondria associated cysteine rich protein	Involved in sperm motility	9.50	Yes	Yes	-	-	-	-	-	-	-
SPEF2	Sperm flagellar 2	Required for correct axoneme development and sperm head morphology	9.50	Yes	Yes	-	-	-	-	-	-	-
TEKT2	Tektin 2	Plays a key role in the assembly or attachment of the inner dynein arm to microtubules in sperm flagella	9.50	Yes	Yes	-	-	-	-	-	-	-
TEKT3	Tektin 3	Forms filamentous polymers in the walls of flagellar microtubules. Required for normal sperm mobility	9.50	Yes	Yes	-	-	-	-	-	-	-
TPPP2	Tubulin polymerization promoting protein family member 3	Probable regulator of microtubule dynamics required for sperm motility	9.50	Yes	Yes	-	-	-	-	-	-	-
TTC21A	Tetratricopeptide repeat domain 21A	Required for sperm flagellar formation and intraflagellar transport	9.50	Yes	Yes	-	-	-	-	-	-	-
TXNDC2	Thioredoxin domain containing 2	Probably plays a regulatory role in sperm development	9.50	Yes	Yes	-	-	-	-	-	-	-

ART, assisted reproductive technology; Oligo, oligozoospermia; Astheno, asthenozoospermia; Terato, teratozoospermia; UMI, unexplained male infertility; Fert, fertilization; Embr dev, embryo development; Preg, pregnancy; Var, varicocele; Subst users, substance users.

understanding of sperm function and more accurate predictions of male fertility. The results presented in this review show a short list of nine sperm mRNA with the potential to take part in this group of biomarkers. However, further exploration of other transcripts, including non-coding RNA, remains a promising avenue that could broaden this field of research and uncover additional biomarkers of important value (*Corral-Vazquez et al., 2019*). Analysing these transcripts in prospective studies will help to determine whether their dysregulation, individually or alongside other factors, could enhance diagnosis and prognosis, particularly in the context of ART outcomes.

CONCLUSIONS

Many studies have focused on identifying differentially expressed mRNA or predicting ART outcomes, yet methodological inconsistencies raise concerns about the reliability of results. Standardizing procedures and ensuring rigorous methodologies are crucial for enhancing research validity and reproducibility.

The gene prioritization system advances the understanding of male infertility by integrating diverse evidence levels. It identifies genes crucial for sperm function and fertility, validated through Gene Ontology analysis and cross-referenced studies. The adaptability of the system allows it to incorporate new research.

Abnormal gene expression may indicate disrupted spermatogenesis in male infertility. Although a direct cause–effect relationship has not been established to date, the nine genes identified in Bin 10 show promise as markers of molecular level changes. Overall, dysregulated expression may enhance diagnosis and prognosis more effectively than focusing on individual gene alterations.

Predicting male fertility is complex; accordingly, combining sperm transcriptome analysis with semen quality assessments could refine fertility predictions. This review highlights nine sperm mRNA biomarkers that could potentially improve diagnosis and prognosis, particularly in ART outcomes, by offering a deeper understanding of sperm function.

ACKNOWLEDGEMENTS

This study was supported by Instituto de Salud Carlos III, Ministry of Science and Innovation, Government of Spain (PI21/00564) and Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya (2021SGR00122). Coral Zurera is the recipient of a grant from Universitat Autònoma de Barcelona (UAB) (UAB/PIF2022). The authors wish to thank Professor Mercedes Campillo Grau from the Department of Paediatrics, Obstetrics and Gynaecology, and Preventive Medicine and Public Health, Area of Preventive Medicine and Public Health at UAB for her assistance with the bin analysis of the gene list.

AUTHOR CONTRIBUTIONS

Ester Anton: conceptualization, data curation, formal analysis, funding acquisition, validation, writing – review and editing. Coral Zurera-Egea: data curation, formal analysis, validation, writing – review and editing. Roser Farriol: data curation, writing – review and editing. Zaida Sarrate: formal analysis, funding acquisition, writing – review and editing. Joan Blanco: conceptualization, data curation, funding acquisition, validation, writing – original draft, writing – review and editing.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used OpenAI ChatGPT to improve the readability and language of some parts of the manuscript. After using this tool, the authors reviewed and edited the content as needed, and take full responsibility for the content of the publication.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.rbmo.2025.104993.

REFERENCES

Abu-Halima, M., Becker, L.S., Al Smadi, M.A., Abdul-Khaliq, H., Raeschle, M., Meese, E., 2023. Sperm Motility Annotated Genes: Are They Associated with Impaired Fecundity? Cells 12, 1239. https://doi.org/10.3390/cells12091239.

Aghajanpour, S., Ghaedi, K., Salamian, A., Deemeh, M.R., Tavalaee, M., Moshtaghian, J., Parrington, J., Nasr-Esfahani, M.H., 2011. Quantitative expression of phospholipase C zeta, as an index to assess fertilization potential of a semen sample. Hum Reprod 26, 2950–2956. https://doi.org/10.1093/humrep/der285.

Amor, H., Zeyad, A., Hammadeh, M.E., 2021. Tobacco smoking and its impact on the expression level of sperm nuclear protein genes: H2BFWT, TNP1, TNP2, PRM1 and PRM2. Andrologia 53, e13964. https://doi.org/10.1111/and.13964.

Barragán, M., Martínez, A., Llonch, S., Pujol, A., Vernaeve, V., Vassena, R., 2015. Effect of ribonucleic acid (RNA) isolation methods on putative reference genes messenger RNA abundance in human spermatozoa. Andrology 3, 797–804. https://doi.org/10.1111/ANDR.12053.

Barratt, C.L.R., Björndahl, L., De Jonge, C.J., Lamb, D.J., Martini, F.O., McLachlan, R., Oates, R.D., van der Poel, S., John, B.S., Sigman, M., Sokol, R., Tournaye, H., 2017. The diagnosis of male infertility: an analysis of the evidence to support the development of global WHO guidance-challenges and future research opportunities. Hum Reprod Update 23, 660–680 . https://doi.org/10.1093/HUMUPD/DMX021.

Cappallo-Obermann, H., Spiess, A.N., 2016. Comment on "Absence of sperm RNA elements correlates with idiopathic male infertility. Sci Transl Med 8, 353tc1. https://doi.org/10.1126/SCITRANSLMED.AAF2396.

Cheung, S., Parrella, A., Rosenwaks, Z., Palermo, G.D., 2019. Genetic and epigenetic profiling of the infertile male. PLoS One 14, e0214275. https://doi.org/10.1371/journal.pone.0214275.

Corral-Vazquez, C., Blanco, J., Aiese Cigliano, R., Sarrate, Z., Rivera-Egea, R., Vidal, F., Garrido, N., Daub, C., Anton, E., 2021. The RNA content of human sperm reflects prior events in spermatogenesis and potential post-fertilization effects. Mol Hum Reprod 27, gaab035. https://doi.org/10.1093/molehr/gaab035.

Corral-Vazquez, C., Salas-Huetos, A., Blanco, J., Vidal, F., Sarrate, Z., Anton, E., 2019. Sperm microRNA pairs: new perspectives in the search for male fertility biomarkers. Fertil Steril 112, 831–841. https://doi.org/10.1016/j.fertnstert.2019.07.006.

Del Giudice, F., Belladelli, F., Chen, T., Glover, F., Mulloy, E.A., Kasman, A.M., Sciarra, A., Salciccia, S., Canale, V., Maggi, M., Ferro, M., Busetto, G.M., De Berardinis, E., Salonia, A., Eisenberg, M.L., 2022. The association of impaired semen quality and pregnancy rates in assisted reproduction technology cycles: Systematic review and meta-analysis. Andrologia 54, e14409. https://doi.org/10.1111/AND.14409.

Depa-Martynów, M., Kempisty, B., Lianeri, M., Jagodziński, P.P., Jędrzejczak, P., 2007. Association between fertilin β, protamines 1 and 2 and spermatid-specific linker histone H1-like protein mRNA levels, fertilization ability of human spermatozoa, and quality of preimplantation embryos. Folia Histochem Cytobiol 45, 79–85.

Ferguson, B.S., Nam, H., Hopkins, R.G., Morrison, R.F., 2010. Impact of reference gene

selection for target gene normalization on experimental outcome using real-time qRT-PCR in adipocytes. PLoS One 5, e15208. https://doi.org/10.1371/JOURNAL.PONE.0015208.
Giebler, M., Greither, T., Handke, D., Seliger, G., Behre, H.M., 2021. Lower Spermatozoal PIWI-LIKE 1 and 2 Transcript Levels Are Significantly Associated with Higher Fertilization Rates in IVF. Int J Mol Sci 22, 11320. https://doi.org/10.3390/IJMS222111320.
Giebler, M., Greither, T., Müller, L., Mösinger, C., Behre, H., 2018. Altered PIWI-LIKE 1 and PIWI-LIKE 2 mRNA expression in ejaculated spermatozoa of men with impaired sperm characteristics. Asian J Androl 20, 260–264. https://doi.org/10.4103/aja.aja_58_17.
Guo, X., Gui, Y.T., Tang, A.F., Lu, L.H., Gao, X., Cai, Z.M., 2007. Differential expression of VASA gene in ejaculated spermatozoa from normozoospermic men and patients with oligozoospermia. Asian J Androl 9, 339–344. https://doi.org/10.1111/j.1745-7262.2007.00253.x.
Hamad, M., Shelko, N., Montenarh, M., Hammadeh, M.E., 2019. The impact of cigarette smoking on protamines 1 and 2 transcripts in human spermatozoa. Hum Fertil 22, 104–110. https://doi.org/10.1080/14647273.2017.1382733.
Hamad, M.F., 2019. Quantification of histones and protamines mRNA transcripts in sperms of infertile couples and their impact on sperm's quality and chromatin integrity. Reprod Biol 19, 6–13. https://doi.org/10.1016/j.repbio.2019.03.001.
Indriastuti, R., Pardede, B.P., Gunawan, A., Ulum, M.F., Arifiantini, R.I., Purwantara, B., 2022. Sperm Transcriptome Analysis Accurately Reveals Male Fertility Potential in Livestock. Animals (Basel) 12, 2955. https://doi.org/10.3390/ANI12212955.
Janghorban-Laricheh, E., Ghazavi-Khorasgani, N., Tavalaee, M., Zohrabi, D., Abbasi, H., Nasr- Esfahani, M.H., 2016a. An association between sperm PLCζ levels and varicocele? J Assist Reprod Genet 33, 1649–1655. https://doi.org/10.1007/s10815-016-0802-5.
Jedrzejczak, P., Kempisty, B., Bryja, A., Mostowska, M., Depa-Martynow, M., Pawelczyk, L., Jagodzinski, P.P., 2007. Quantitative assessment of transition proteins 1, 2 spermatid-specific linker histone H1-like protein transcripts in spermatozoa from normozoospermic and asthenozoospermic men. Arch Androl 53, 199–205. https://doi.org/10.1080/01485010701426430.
Jensen, C.F.S., Østergren, P., Dupree, J.M., Ohl, D.A., Sønksen, J., Fode, M., 2017. Varicocele and male infertility. Nat Rev Urol 14, 523–533. https://doi.org/10.1038/NRUROL.2017.98.
Jodar, M., Kalko, S., Castillo, J., Ballescà, J.L., Oliva, R., 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Human Reproduction 27, 1431–1438. https://doi.org/10.1093/humrep/des021.
Kempisty, B., Depa-Martynow, M., Lianeri, M., Jedrzejczak, P., Darul-Wasowicz, A., Jagodzinski, P.P., 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. Folia Histochem Cytobiol 45, 109–113.
Liu, X.X., Cai, L., Liu, F.J., 2018. An in silico analysis of human sperm genes associated with asthenozoospermia and its implication in male infertility. Medicine (United States) 97, e13338. https://doi.org/10.1097/MD.0000000000013338.
Montjean, D., De La Grange, P., Gentien, D., Rapinat, A., Belloc, S., Cohen-Bacrie, P., Menezo, Y., Benkhalifa, M., 2012. Sperm transcriptome profiling in oligozoospermia. J Assist Reprod Genet 29, 3–10. https://doi.org/10.1007/S10815-011-9644-3.
Oud, M.S., Volozonoka, L., Smits, R.M., Vissers, L.E.L.M., Ramos, L., Veltman, J.A., 2019. A systematic review and standardized clinical validity assessment of male infertility genes. Hum Reprod 34, 932. https://doi.org/10.1093/HUMREP/DEZ022.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/BMJ.N71.
Peltier, H.J., Latham, G.J., 2008. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. RNA 14, 844–852. https://doi.org/10.1261/RNA.939908.
Punab, M., Poolamets, O., Paju, P., Vihljajev, V., Pomm, K., Ladva, R., Korrovits, P., Laan, M., 2017. Causes of male infertility: a 9-year prospective monocentre study on 1737 patients with reduced total sperm counts. Hum Reprod 32, 18–31. https://doi.org/10.1093/HUMREP/DEW284.
Ray, A., Shah, A., Gudi, A., Homburg, R., 2012. Unexplained infertility: an update and review of practice. Reprod Biomed Online 24, 591–602. https://doi.org/10.1016/J.RBMO.2012.02.021.
Sadakierska-Chudy, A., Patrylak, J., Janeczko, J., Chudy, J., 2020. Downregulation of gene expression and the outcome of ICSI in severe oligozoospermic patients: A preliminary study. Mol Reprod Dev 87, 1219–1230. https://doi.org/10.1002/MRD.23442.
Scholes, A.N., Lewis, J.A., 2020. Comparison of RNA isolation methods on RNA-Seq: implications for differential expression and meta-analyses. BMC Genomics 21, 249. https://doi.org/10.1186/S12864-020-6673-2.
Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., Ragg, T., 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7, 3. https://doi.org/10.1186/1471-2199-7-3.
Shabani Nashtaei, M., Nekoonam, S., Naji, M., Bakhshalizadeh, S., Amidi, F., 2018. Cryoprotective effect of resveratrol on DNA damage and crucial human sperm messenger RNAs, possibly through 5′ AMP-activated protein kinase activation. Cell Tissue Bank 19, 87–95. https://doi.org/10.1007/S10561-017-9642-5.
Shen, S., Wang, J., Liang, J., He, D., 2013. Comparative proteomic study between human normal motility sperm and idiopathic asthenozoospermia. World J Urol 31, 1395–1401. https://doi.org/10.1007/s00345-013-1023-5.
Tavalaee, M., Nasr-Esfahani, M.H., 2016. Expression profile of PLCζ, PAWP, and TR-KIT in association with fertilization potential, embryo development, and pregnancy outcomes in globozoospermic candidates for intra-cytoplasmic sperm injection and artificial oocyte activation. Andrology 4, 850–856. https://doi.org/10.1111/andr.12179.
Tavalaee, M., Nomikos, M., Lai, F.A., Nasr-Esfahani, M.H., 2018. Expression of sperm PLCζ and clinical outcomes of ICSI-AOA in men affected by globozoospermia due to DPY19L2 deletion. Reprod Biomed Online 36, 348–355. https://doi.org/10.1016/j.rbmo.2017.12.013.
Turri, F., Capra, E., Lazzari, B., Cremonesi, P., Stella, A., Pizzi, F., 2021. A Combined Flow Cytometric Semen Analysis and miRNA Profiling as a Tool to Discriminate Between High- and Low-Fertility Bulls. Front Vet Sci 8, 703101. https://doi.org/10.3389/FVETS.2021.703101.
Vander Borght, M., Wyns, C., 2018. Fertility and infertility: Definition and epidemiology. Clin Biochem 62, 2–10. https://doi.org/10.1016/J.CLINBIOCHEM.2018.03.012.
Wyns, C., Bergh, C., Calhaz-Jorge, C., de Geyter, C., Kupka, M.S., Motrenko, T., Rugescu, I., Smeenk, J., Tandler-Schneider, A., Vidakovic, S., Goossens, V., 2020. ART in Europe, 2016: results generated from European registries by ESHRE. Hum Reprod Open 2020, hoaa032. https://doi.org/10.1093/HROPEN/HOAA032.

Received 29 December 2024; received in revised form 29 March 2025; accepted 1 April 2025.