Universitat Autònoma de Barcelona

Human sperm transcriptome: characterization, biological relevance, and biomarker functionality

Celia Corral Vázquez

Human sperm transcriptome: characterization, biological relevance, and biomarker funcionality

Celia Corral Vázquez

UAB
2019

Universitat Autònoma de Barcelona

Human sperm transcriptome: characterization, biological relevance, and biomarker functionality

Celia Corral Vázquez

Tesis doctoral presentada para optar al Grado de Doctora con Mención Internacional por la Universidad Autónoma de Barcelona

Directores:

Dra. Ester Anton Martorell y Dr. Joan Blanco Rodríguez

Programa de Doctorado en Biología celular

Departamento de Biología celular, Fisiología e Inmunología

Bellaterra, 2019

La Dra. **Ester Anton Martorell**, profesora del Departamento de Biología Celular, Fisiología e Inmunología de la Universitat Autònoma de Barcelona,

y el Dr. **Joan Blanco Rodríguez**, profesor del Departamento de Biología Celular, Fisiología e Inmunología de la Universitat Autònoma de Barcelona,

CERTIFICAN

Que **Celia Corral Vázquez** ha realizado bajo su dirección el trabajo de investigación que se expone en la memoria titulada "Human sperm transcriptome: characterization, biological relevance, and biomarker functionality" para optar al grado de Doctora por la Universitat Autònoma de Barcelona.

Que este trabajo se ha llevado a cabo en la Unidad de Biología Celular del Departamento de Biología Celular, Fisiología e Inmunología de la Universitat Autònoma de Barcelona.

Y para que así conste, firman el presente certificado.

Dra. Ester Anton Martorell

Dr. Joan Blanco Rodríguez

Celia Corral Vázquez

Bellaterra, 19 de septiembre de 2019

La presente Tesis Doctoral se ha llevado a cabo con la financiación de los siguientes proyectos:

- 2014/SGR00524 (Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya).
- 2017/SGR-503 (Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya).
- UAB CF-180034 (Universitat Autònoma de Barcelona).

La autora, Celia Corral Vázquez, se ha beneficiado de una beca predoctoral para Personal Investigador en Formación (PIF, 2015) concedida por la Universitat Autònoma de Barcelona.

Las imágenes usadas en el presente trabajo, extraídas o adaptadas de otras publicaciones, han recibido los permisos necesarios para ser utilizadas.

Diseño de cubierta y separadores de sección: SHOOK Studio

Summary

The biological relevance of sperm contribution to the embryo has been shown to go beyond a mere transmission of the paternal genome. Several findings revealed that human spermatozoa carry a complex population of coding and non-coding RNAs with potential implications in multiple fertility-related pathways. Accordingly, the consideration of these molecules as simple residual pools of earlier processes has been left behind. This new paradigm also opens the possibility for potential applications in the field of male fertility biomarkers. However, sperm transcriptomic analysis has several limitations due to the heterogeneity and delicate nature of these molecules, besides the small amount of RNA contained in spermatozoa.

In this context, the objective of this Doctoral Thesis is to characterize the human sperm transcriptome to set up the basis for developing new biomarkers of male fertility. Within this goal, the following aims were undertaken: 1) to optimize specific methodologies of sperm RNA analysis using qRT-PCR and RNA-seq strategies; 2) to provide an integrative profiling and functional characterization of sperm mRNAs and lncRNAs by RNA-seq technologies; and 3) to establish new fertility biomarkers among the transcriptomic cargo of the human spermatozoa.

For this purpose, the experimental protocols and data analysis were adapted to the inherent limitations of sperm RNA and to the used transcriptomic technology. Therefore, methods for the elimination of non-sperm cells from semen samples were implemented, together with strict quality controls for ensuring the absence of DNA and non-sperm RNA. Besides, an organic solvent-based method was used for qRT-PCR studies, and non-organic solvent kits were employed for RNA-seq.

The obtained data were normalized by specific methods depending on the used technique. In particular, the normalization of sperm miRNA qRT-PCR singleplex studies required the determination of a suitable set of normalizing miRNAs molecules. This was achieved by comparing the results derived from a sperm miRNA expression dataset normalized by: i) the reference Mean Centering Restricted (MCR) method; and ii) the expression level of different miRNAs. The miRNAs hsa-miR-100-5p and

hsa-miR-30a-5p showed ubiquitous and stable expressions, and data normalized by their mean expression led to results with an appropriate quality when compared to MCR. Therefore, this miRNA combination was suggested as the most suitable choice for data normalization in further sperm singleplex studies.

RNA-seq analysis was used to characterize the sperm transcriptome cargo of fertile individuals. Results revealed a complex network of mRNAs and lncRNAs with a high fragmentation status, but containing a host of ubiquitous transcripts. Gene ontology analyses of the whole set of expressed mRNAs showed an enrichment of spermatogenesis and reproduction processes, which was more significant in the sets of highly expressed, ubiquitous, and highly stable mRNAs. Additionally, the functional profiling of potential *cis*-target genes of the observed lncRNAs showed a significant involvement in embryo development and cell adhesion. This implication became more evident in those *cis*-target genes that were not present among the sperm mRNA cargo.

Finally, the detection of ubiquitous transcripts and pairs of RNAs with correlated expressions suggested a potential use of these molecules as fertility biomarkers. Accordingly, the presence of sperm miRNA pairs with a correlated expression in fertile individuals that was disrupted in infertile patients of different ethiologies (asthenozoospermia, teratozoospermia, oligozoospermia, and Unexplained Male Infertility or UMI) was evaluated and validated by qRT-PCR. The hsa-miR-942-5p/hsa-miR-1208 pair allowed correctly classifying the 85.71% of infertile individuals, thus achieving the highest potential for discerning infertility cases with seminal alterations. Additionally, the pair hsa-miR-34b-3p/hsa-miR-93-3p was highlighted due to its high potential for discerning UMI patients. Besides, several pairs of ubiquitous lncRNAs and mRNAs were also observed to display a correlated expression in fertile individuals, becoming potential candidates for further biomarker studies.

Index

1 Introduction

1.1. Human male infertility

Infertility is defined by the World Health Organization (WHO) (World Health Organization, 2010) and the International Committee for Monitoring Assisted Reproductive Technology (ICMART) (Zegers-Hochschild et al., 2009) as «the inability of a sexually active, non-contracepting couple to achieve pregnancy in one year». This worldwide disorder affects around 25% of couples in reproductive age (Dohle et al., 2005), although this percentage is variable depending on geographical zones (Agarwal et al., 2015). Among the infertile couples, the origin of around 20-30% of cases lies solely in a male factor, 50% of them are due to female factor, and a 20-30% is due to a combination of male and female factor (Sharlip et al., 2002).

1.1.1. Clinical assessment of male fertility

The evaluation and diagnosis of male infertility usually start with the physical examination of the urogenital tract, followed by a laboratory procedure in order to assess semen quality, endocrine profile, and genetic anomalies associated with infertility (World Health Organization, 2000). These analyses allow classifying male infertility in a wide variety of causes including physiological alterations, diseases, physical injuries or exposure to external factors. Nevertheless, it is important to mention that, in a significant amount of patients, the molecular cause of infertility remains unknown after the application of these analyses (see section 1.1.2. Infertility of Unknown Origin) (Table 1).

1.1.1.1. Physical examination

The first steps of male infertility evaluation aim to detect possible pathologies or urogenital abnormalities. This assessment is achieved by a detailed medical history assessment and a urogenital exploration (e.g. genital examination, evaluation of epididymis and *vasa deferentia*, and possible diagnosis of varicocele, among others) (Esteves et al., 2011; World Health Organization, 2000).

Additional ultrasonography, imaging, and thermography techniques can be necessary to complete the physical observations (World Health Organization, 2000).

Table 1. Distribution of diagnostic categories of male infertility (Sigman et al., 2009).

Category	Frequency (%)
Unknown Origin	43.3
Varicocele	26.6
Obstruction	15.3
Cryptorchidism	2.7
Immunological	2.6
Ejaculatory failure	2.0
Endocrinologic	1.5
Drug/radiation	1.4
Genetic	1.2
Testicular failure	1.1
Sexual dysfunction	0.7
Pyospermia	0.5
Cancer	0.4
Systemic disease	0.3
Infection	0.2
Torsion	0.1
Ultrastructural	0.1

1.1.1.2. Semen analysis

Semen analysis, commonly known as seminogram or spermiogram, is a laboratory technique based on the evaluation of semen quality in order to provide a fertility prognosis of the patient. This analysis is based on the measure of seminal alterations as a possible indication of an altered germ cell epithelium, epididymides, or accessory sexual glands (Esteves et al., 2011).

These analyses consist of a microscopic assessment and several biochemical assays, with the purpose of measuring physical characteristics of the ejaculate (volume, pH, viscosity, liquefaction, and appearance), and sperm microscopic parameters (sperm count, motility, and morphology). Additionally, evaluation of leukocyte quantification, presence of non-spermatic cellular elements, vitality, sperm agglutination, fructose detection (when no spermatozoa are found in the semen sample), sperm DNA integrity, and anti-sperm antibodies can be also

included in the analysis (World Health Organization, 2010). To facilitate a diagnostic assessment of the male fertility status by semen evaluation, the WHO provides reference limits for the aforementioned parameters (Table 2) (Cooper et al., 2009).

Table 2. Lower reference thresholds for semen parameters established by the World Health Organization (WHO) from 1992 to 2010.

Semen parameters	**WHO, 1992**	**WHO, 1999**	**WHO, 2010**
Volume (mL)	2	2	1.5
Sperm concentration (spz/mL)	$20 \cdot 10^6$	$20 \cdot 10^6$	$15 \cdot 10^6$
Total sperm number	$40 \cdot 10^6$	$40 \cdot 10^6$	$39 \cdot 10^6$
Total motility (% PR+NP)	50	50	40
PR (%)	25	25	32
Vitality (% alive)	75	75	58
Morphology (% NF)	30	14	4
Leukocyte count	$1.0 \cdot 10^6$	$1.0 \cdot 10^6$	$1.0 \cdot 10^6$

PR = progressive motility; NP = non-progressive motility; NF = normal forms; spz = spermatozoa.

The measure of semen parameters with a value *equal to* or *above* the reference limits is defined as normozoospermia. Regarding the described thresholds, the assigned nomenclature depends on which parameter is altered. Therefore, the main parameter alterations are named asthenozoospermia (percentage of spermatozoa with progressive motility below the lower reference limit), teratozoospermia (low percentage of morphologically normal spermatozoa), and oligozoospermia (decreased number of spermatozoa) (Table 3) (World Health Organization, 2010). These anomalies take place together in many cases, described as the oligo-astheno-teratozoospermia syndrome (Dohle et al., 2005).

The alteration of semen parameters is associated with fertility problems, and the coexistence of several imbalances supposes an increment of the risk for infertility (Guzick et al., 2001). Nevertheless, it has been stated that reference limits of semen parameters should not be interpreted as conclusive data for distinguishing infertile patients. Several studies have revealed the limitations of this analysis as a predictive measure of male infertility, since there is no clear correspondence between seminal parameters and pregnancy achievement (Jensen et al., 2002; Van Der

Steeg et al., 2011). In fact, the official established thresholds by WHO have suffered substantial modifications throughout the last decades (Table 2).

Table 3. Nomenclature related to semen quality (World Health Organization, 2010).

Aspermia	No semen (or retrograde ejaculation)
Asthenozoospermia	Percentage of progressively motile (PR) spermatozoa below the lower reference limit
Asthenoteratozoospermia	Percentages of both progressively motile (PR) and morphologically normal spermatozoa below the lower reference limits
Azoospermia	No spermatozoa in the ejaculate (given as the limit of quantification for the assessment method employed)
Cryptozoospermia	Spermatozoa absent from fresh preparations but observed in a centrifuged pellet
Haemospermia (haematospermia)	Presence of erythrocytes in the ejaculate
Leukospermia (leukocytospermia, pyospermia)	Presence of leukocytes in the ejaculate above the threshold value
Necrozoospermia	Low percentage of live, and high percentage of immotile, spermatozoa in the ejaculate
Normozoospermia	Total number (or concentration, depending on the outcome reported) of spermatozoa, and percentages of progressively motile (PR) and morphologically normal spermatozoa, equal to or above the lower reference limits
Oligoasthenozoospermia	Total number (or concentration, depending on the outcome reported) of spermatozoa, and percentage of progressively motile (PR) spermatozoa, below the lower reference limits
Oligoasthenoteratozoospermia	Total number (or concentration, depending on the outcome reported) of spermatozoa, and percentages of both progressively motile (PR) and morphologically normal spermatozoa, below the lower reference limits
Oligoteratozoospermia	Total number (or concentration, depending on the outcome reported) of spermatozoa, and percentage of morphologically normal spermatozoa, below the lower reference limits
Oligozoospermia	Total number (or concentration, depending on the outcome reported) of spermatozoa below the lower reference limit
Teratozoospermia	Percentage of morphologically normal spermatozoa below the lower reference limit

PR = progressive motility.

1.1.1.3. Endocrine evaluation

Endocrine assessment is advised in cases of low sperm concentration or absence of spermatozoa in the ejaculate, as well as erectile dysfunction, hypospermia or symptoms of hypogonadism. Although basic evaluation includes the assessment of serum Follicle-Stimulating Hormone (FSH) and testosterone levels, the measuring of other hormones may also be included, as the Luteinizing Hormone (LH), prolactin, gonadotropin or estradiol (Table 4) (Esteves et al., 2011).

Table 4. Normal serum levels of hormones in fertile men (American College of Physicians, 2007).

Follicle-Stimulating Hormone (FSH)	5.0-15.0 mIU/mL
Luteinizing Hormone (LH)	3.0-15.0 mIU/mL
Testosterone	300-1,200 ng/dL
Prolactin	<15 ng/mL
Estradiol	10-30 pg/mL

1.1.1.4. Genetic evaluation

Physical abnormalities, endocrine irregularities, and seminogram alterations can be originated by genetic factors. Accordingly, genetic evaluation is performed in certain cases of infertility involving low sperm concentration, the absence of at least one *vas deferens*, Infertility of Unknown Origin (IUO) (see section 1.1.2. Infertility of Unknown Origin), or patients with a personal or familiar medical history of genetic syndromes (Esteves et al., 2011).

The most frequent genetic factors that affect fertility are the presence of numerical or structural chromosome abnormalities in the karyotype, Y chromosome microdeletions, and pathological variants of the Cystic Fibrosis Transmembrane Conductance Regulator gene (*CFTR*) (Centers for Disease Control and Prevention, 2002). In laboratory evaluation, chromosomal abnormalities are mainly detected by G-band karyotyping, while *CFTR* gene mutations and Y chromosome microdeletions are detected by DNA amplification using the Polymerase Chain Reaction (PCR) (Esteves et al., 2011).

1.1.2. Infertility of Unknown Origin

IUO refers to the occurrence of fertility impairments due to a spontaneous, obscure, or unknown cause. Patients with IUO endure infertility diagnosis, and display no abnormalities in physical examinations, genetic factor analyses, or endocrine clinical testing. Attending to male factor, this condition affects around 30-40% of the male infertile population (Table 1) (Sigman et al., 2009). Within this percentage, IUO is classified in two categories depending on the normality or abnormality of semen parameters: Idiopathic Male Infertility (IMI) and Unexplained Male Infertility (UMI) (Hardin and Kim, 2015).

IMI patients are defined as infertile men with no history of physical, genetic or endocrine alterations, but still showing an unexplained reduction in semen quality regarding microscopic semen parameters (Dohle et al., 2005). Approximately 67% of males with IUO origin are categorized as IMI patients: around 30% of them present abnormalities in sperm motility, morphology, density or seminal volume, while 37% of the patients show alterations in two or more of these parameters (Table 5) (Sigman et al., 2009). Among the possible causing factors of IMI, chronic stress and unknown genetic and epigenetic abnormalities have been suggested (Dohle et al., 2005).

Table 5. Distribution of abnormalities of semen parameters in infertile patients (Sigman et al., 2009).

Abnormality in semen parameters	Frequency (%)
Azoospermia	4
Motility	18
Morphology	7
Density	2
Volume	2
Defects in two or more parameters	37
All parameters normal (unexplained male infertility)	30

The UMI category comprises infertile patients with no history of physical, genetic or endocrine alterations, and normal semen parameters (Hamada et al., 2011), from infertile couples in which any female contribution to infertility has been discarded. The prevalence of this condition was

determined to range from 6 to 30% of infertile males (Moghissi and Wallach, 1983; Sigman et al., 2009). UMI has been attributed to failures in oocyte fertilization associated with autoimmune causes, DNA integrity or the presence of Reactive Oxygen Species (ROS). It has also been associated with a deficiency of some sperm function such as capacitation, zona pellucida binding and acrosome reaction (Hamada et al., 2011). Moreover, UMI may also be the consequence of post-fertilization failures, by the inability of sperm to generate embryos of appropriate quality to carry on their development program. Although some studies have linked these failures with alterations of sperm DNA integrity (Host et al., 2000; Saleh et al., 2003), it has recently been suggested that the incorporation of an altered sperm transcriptome in the oocyte can also result in dysfunctional embryo development (Yuan et al., 2016).

1.1.3. Molecular insight of male infertility through *omics* technologies

The field of *omics* is based on the large-scale analysis of the genome, methylome, transcriptome, and proteome of certain cells and tissues. This state-of-the-art research allows studying the interconnections of macromolecules, providing a complete modular viewpoint of the whole process of gene expression by integrating datasets of thousands of molecules produced by high-throughput approaches (Yadav, 2017).

In the last years, biomedical research has been enquiring into the biochemical and molecular background related to IUO. The process of sperm generation and maturation involves a complex network of modifications including DNA packaging, and different types of epigenetic marks that regulate gene expression. By the use of *omics* technologies, alterations produced throughout those processes can be assessed with the purpose of discovering new possible causes of IUO (Yadav, 2017).

Genomic studies in infertile men have found different Single Nucleotide Polymorphisms (SNPs) associated with oligozoospermia, idiopathic azoospermia (Aston et al., 2010; Aston and Carrell, 2009), Non-obstructive Azoospermia (NOA) (Zhao et al., 2012; Zou et al., 2014), and mixed alterations of semen parameters (Kosova et al., 2012). Besides SNPs, the occurrence of rare Copy Number Variations (CNVs) has also been

identified as a cause of spermatogenic failure in testicular tissues. Certain CNVs have been found associated with patients with Sertoli-cell-only Syndrome (SCOS) (Tüttelmann et al., 2011), meiotic arrest, and azoospermia (Eggers et al., 2015). Additionally, some genomic studies have established a relation between low sperm count and X-linked genetic factors (Krausz et al., 2012; Lopes et al., 2013). Regarding pathogenic genetic variations, several studies have found gene mutations associated with infertility profiles as NOA (*TEX11, ZMYND15, TAF4B, SYCE1, MCM8,* and *TEX15*), teratozoospermia (*AURKC, DPY19L2, SPATA16,* and *DNAH1*), asthenoteratozoospermia (*CATSPER1, CATSPER2, SEPT12* and *SLC26A8*), and UMI (*PLCZ1*) (Bracke et al., 2018).

Beyond genetic code, the involvement of methylome variations as a cause of male infertility has been also analyzed using *omics* approaches. Many researchers have identified altered methylation at CpG sites in sperm cells as a possible cause of altered semen parameters (Camprubí et al., 2016; Du et al., 2016; Montjean et al., 2015; Schütte et al., 2013), UMI (Urdinguio et al., 2015), and pregnancy failure (Benchaib et al., 2005).

Concerning RNA profiling, several researchers have detected differential quantification levels of various genes in sperm cells of infertile men (Bansal et al., 2015; Montjean et al., 2012; Pacheco et al., 2011; Sato et al., 2015). Besides messenger RNAs (mRNAs), the study of non-coding RNA (ncRNA) is important due to their regulatory involvement: they are related to the control of gene expression by interacting with other elements like DNA sequences, RNAs, or proteins (Hamatani, 2012). Therefore, they constitute an important mechanism during spermatogenesis, displaying differential fingerprints associated to male infertility (Abhari et al., 2014; Abu-Halima et al., 2014b; Salas-Huetos et al., 2016, 2015).

The main objective of Proteomics is to assess the translational status of cells, offering a global view of functionality, structure, abundance, post-translational modifications and alterations caused by diseases (Horgan et al., 2009). In the field of infertility, different studies have aimed to find an association between this disease and proteomic alterations (Parte et al., 2012; Shen et al., 2013), as well as post-translational modifications in sperm proteins (Rahman et al., 2013), or different profiles of sperm tail proteins (Hashemitabar et al., 2015) and protein targets of anti-sperm antibodies (Zangbar et al., 2016).

In summary, the unraveling of the causes of male infertility through *omics* technologies is a novel and complex research field with many open possibilities. Among them, the study of sperm RNA appears to be a promising area since it constitutes a nexus between the above mentioned *omics* fields. This approach provides an in-depth vision of the sperm gene expression cargo, which can also be related to regulatory epigenetic factors, including the action of ncRNAs. In turn, these transcripts eventually will regulate the formation of proteins encoded by mRNAs. Altogether, this technology offers an integrative view of gene expression and regulation from several standpoints of the RNA biological pathway.

1.2. Methods of RNA profiling

Technologies of RNA profiling have undergone a rapid development involving improved sensitivity and decreasing allocated costs. These techniques are focused on the identification and/or quantification of both large and small RNA molecules (Lowe et al., 2017). Although several classification criteria can be followed, in general terms, RNA profiling methods can be categorized into single transcript profiling techniques and whole transcriptomic technologies.

1.2.1. Single-transcript profiling techniques

Individual transcript profiling methods were developed for individual expression studies or for the construction of basic libraries. Specifically, the Sanger method started to be used in the 80s for sequencing short random transcripts (known as expressed sequence tags or EST) from previously constructed complementary DNA (cDNA) libraries (Sim et al., 1979). This technology was predominant during the 1990s since it provided information of the transcript cargo of an organism without the necessity of sequencing the whole genome or having prior knowledge of its genetic content (Marra et al., 1998). Additionally, other individual transcript quantification methods like Northern Blotting, *In Situ* Hybridization (ISH) (Parker and Barnes, 1999), Nylon Membrane Arrays

(Alwine et al., 1977), and Reverse Transcription PCR (RT-PCR) (Becker-André and Hahlbrock, 1989) became popular.

Among single-transcript analysis methods, Quantitative Reverse Transcription PCR (qRT-PCR) has become the golden standard technique for the detection and quantification of RNA molecules. This technology is based on conventional RT-PCR, which requires previous knowledge about the sequence of the genes of interest. qRT-PCR has a wide dynamic range, and allows the quantification of the amplified transcripts (Bustin and Mueller, 2005); after the enzymatic retro-transcription of the RNA molecule of interest into cDNA, its amplification is monitored in real time by detecting the fluorescence emitted by the reporter molecules. This signal intensity is relative to the starting copy number of RNA transcripts, so the higher amount of cDNA, the sooner detection of a significant increase in fluorescence (a point known as Cycle Threshold or Ct).

Fluorescence detection methods can be either probe or non-probe based. Non-probe based methods measure the binding of non-specific reporters (like SYBR Green) to double-stranded DNA when it is synthetized during the PCR amplification, producing the emission of an increased detectable fluorescence. Alternatively, probe-based methods make use of amplicon-specific fluorescence probes (like TaqMan assays). In this case, the fluorescence emission is only generated when the probe hybridizes with its complementary sequence in the cDNA molecule (Bustin and Mueller, 2005).

Besides the analysis of mRNAs, qRT-PCR supposes an opportunity of performing individual transcript profiling of small molecules of RNA (see section 1.3. Sperm transcriptome), since it allows detecting such low-abundant targets with high sensitivity and specificity. Their quantification can be performed by probe-based methods, adding an initial reverse-transcription step based on stem-loop primers. These primers are composed by three main elements: a short single-stranded sequence (complementary to the 3' end of a given small RNA), a double-stranded stem, and a loop with a primer-binding sequence (Salone and Rederstorff, 2015).

With the purpose of simultaneously studying the expression of several transcripts, multiple qRT-PCR can be performed in a single reaction tube, which is known as *multiplex* qRT-PCR (Bustin and Mueller, 2005).

Additionally, qPCR arrays are also commonly used to evaluate the expression profiling of a predesigned set of genes.

1.2.2. Transcriptomics technologies.

Earliest transcriptomic studies were performed during the 1990s, with the development of the Serial Analysis of Gene Expression (SAGE) (Bartlett et al., 2003), based on EST technology. This method consists on a cDNA synthesis followed by digestion with restriction enzymes into short fragments (11 base pairs or bp). These short sequences are lately concatenated into longer strands (>500 bp), which can be sequenced by Sanger sequencing or aligned to a reference genome for gene identification and relative quantification. Additionally, the Cap Analysis of Gene Expression (CAGE) technique is a variant of SAGE that is focused on the sequencing of gene start sites (cDNA fragments from the 5' end of mRNA), mostly for promoter analyses (Lowe et al., 2017).

Microarray technology development supposed an overtaking of the limitations of the previous transcriptomic methods. This technique allows the relative simultaneous quantification of thousands of predefined transcripts, although prior knowledge of genome sequences of the analyzed organism is required, like libraries or annotated genome sequences. This technique is based on the hybridization of fluorescently labeled transcripts to probes (short nucleotide oligomers) that are attached to a solid substrate (e.g., a glass platform). Depending on probe density, microarrays can be classified into low-density and high-density platforms. Low-density spotted arrays are composed of longer probes, and based on a test-control comparison of samples marked with different fluorophores, allowing the relative quantification of transcripts. High-density short probe arrays have higher transcript resolution, and allow the hybridization and transcript quantification of individual samples (Lowe et al., 2017). The microarray was the predominant high-throughput transcriptome technique during 1990s-2000s (Figure 1) (Nelson, 2001), but

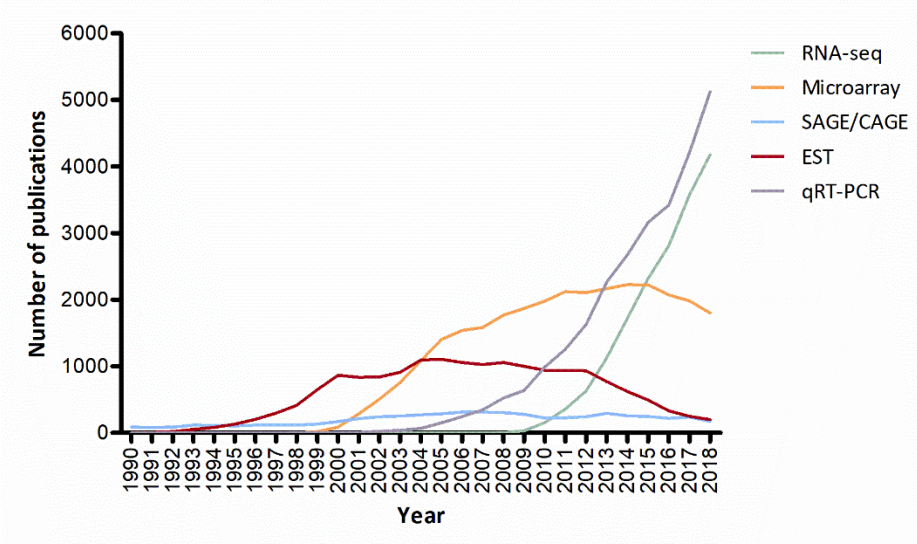


Figure 1. Number of published studies based on different transcriptome profiling techniques (data obtained from Corlan, 2004).

it was surpassed by the rising of RNA sequencing (RNA-seq) methods due to their technical advantages (Table 6) (Zhong Wang and Snyder, 2001).

RNA-seq combines high-throughput sequencing with deep computational reconstruction analyses in order to capture and quantify transcripts of a certain RNA sample. It constitutes the actual state-of-the-art transcriptomic technology, since it achieves better sequence resolution and sensitivity, besides having a higher dynamic range (5 orders of magnitude) than microarrays. Additionally, RNA-seq requires a lower input RNA amount (nanogram scale), and no prior knowledge of genome sequences. RNA-seq can be performed for sequencing large RNAs, but also small RNAs (Lowe et al., 2017). Nowadays, advances in RNA-seq sensitivity have allowed implementing single-cell transcriptome studies, both *in vitro* (Hashimshony et al., 2012) and *in situ* (Lee et al., 2014).

RNA-seq methodology starts with an initial retrotranscription to cDNA, a preamplification step, and library construction, followed by sequencing. The optimal experimental design is variable depending on factors such as the main objective of the study, the RNA quantity input, or the size of the transcripts of interest (Conesa et al., 2016).

Table 6. Comparison of RNA-seq and Microarray methods (Lowe et al., 2017).

	RNA-seq	**Microarray**
Input RNA amount	Low (~ 1 ng total RNA)	High (~ 1 µg mRNA)
Labor intensity	High (sample preparation and data analysis)	Low
Prior knowledge of reference genome	None required, though useful	Required
Quantitation accuracy	~ 90% (limited by sequence coverage)	>90% (limited by fluorescence detection accuracy)
Sensitivity	10^{-6} (limited by sequence coverage)	10^{-3} (limited by fluorescence detection)
Dynamic range	>10^{5} (limited by sequence coverage)	10^{3} - 10^{4} (limited by fluorescence saturation)
Technical reproducibility	>99%	>99%

In mRNA sequencing, the desired molecules are usually pulled apart by using oligonucleotide probes with poly-T sequences that specifically bind the poly(A) tails of mRNAs. Alternatively, depletion of ribosomal RNA (rRNA, which usually constitutes over 90% of total RNA content in non-sperm cells, and therefore can mask the sequencing results) can be also performed. After selection, a reverse transcription is performed with the resulting molecules; alternatively, the selected RNAs are often fragmented before this step. The resulting cDNA molecules are ligated to adapters and usually amplified by PCR to enrich the samples (transcripts that contain the expected 5′ and 3′ adapter sequences), especially in cases of very low RNA input. Library sequencing design may differ depending on several variables so the fragments can be sequenced producing smaller or larger reads (usually <500 bp). Additionally, reads can be sequenced in a single direction (single-end reads, which is cheaper and suitable for well annotated-organisms) or in both directions (paired-end reads, which facilitates *de novo* transcript discovery or isoform analyses). Also, variable sequencing depth can be achieved, which refers to the number of produced reads for a given sample and determines the sensitivity and accuracy of the library. Sequencing depth can be very variable, so libraries with 1-100 million reads can be produced (Conesa et al., 2016). The obtainment of 5 million mapped reads has been considered by some authors as the minimum depth for an expression profiling study (Sims et al., 2014). After library sequencing, computational data processing usually comprises the alignment of the sequenced reads to a reference genome

for transcript identification and quantification. Alternatively, the reads can also be assembled and reconstructed without a reference genome (*de novo* assembly) (Conesa et al., 2016).

Regarding small RNA sequencing (sRNA-seq), the experimental procedure begins with a small RNA purification based on the molecule size, usually by gel electrophoresis of by a specialized RNA extraction method (Lowe et al., 2017). Subsequently, adapter ligation, RT-PCR amplification, and library sequencing are performed (Zheng et al., 2010). Small RNA library depth is usually lower than mRNA-seq, ranging from 2-10 million reads. Like mRNAs, small RNAs can also be aligned to a reference genome or identified *de novo*. Quality controls are used to ensure that the quantified transcripts are not mRNA degradation products (e.g., verifying unexpected read coverage over highly expressed genes as Glyceraldehyde 3-phosphate Dehydrogenase; *GAPDH*) (Conesa et al., 2016).

1.2.3. Data normalization

The whole RNA extraction and profiling process from different assays implies the introduction of variations regarding the total quantity and quality of the obtained RNA. Therefore, the elimination of such variability is mandatory if the results of these assays have to be compared (Chapman and Waldenström, 2015). The process of reducing technical errors or variation between different molecules, samples or experiments is known as normalization. It constitutes an essential step in expression studies since null or incorrect data normalization can dramatically affect the obtained results and the final interpretation of the experimental outcome (Ferguson et al., 2010).

There are numerous normalization methods that are often specific for each employed experimental design. Regarding qRT-PCR, several normalizing methods have been established depending on the number of studied molecules. When normalizing the expression of a large number of different transcripts, a common strategy is to correct raw data by the mean expression value of the whole dataset. Several algorithms derived from this method have been developed, like the global-mean or Mean-Centering (MC) normalization (Mestdagh et al., 2009), and the Mean-Centering Restricted (MCR) strategy, which is specific for miRNA qRT-PCR

studies (Wylie et al., 2011). Nevertheless, mean-based methods are not suitable for singleplex or low numbers of qRT-PCR assays, so alternative strategies are needed in these cases. A recurrent option is the correction by the total amount of obtained RNA, although this strategy is strongly dependent on the accuracy of the used RNA quantification method. The introduction of external RNA molecules (either synthetically generated or cloned and transcribed *in vitro* from another species) as internal normalizing references are also frequently used. Nevertheless, these molecules do not undergo the exact same biosynthesis and endogenous cellular processes than the analyzed transcripts, so they can include additional variability (Huggett et al., 2005). Generally, the use of reference RNA molecules is considered the most effective normalization method for single qRT-PCR assays. This strategy overcomes the previously described limitations, since they correspond to molecules of the same nature than the transcripts of interest, and undergo the same experimental procedure (Huggett et al., 2005).

Importantly, an appropriate reference gene also should display other specific requirements: being consistently expressed, not being co-regulated with the genes of interest, and have minimal expression variations between different individuals, tissues, and physiological states (Chervoneva et al., 2010; Kozera and Rapacz, 2013). Accordingly, many studies rely on the use of housekeeping genes as reference controls, since they are involved in basic essential cellular processes, and are considered to be ubiquitous and stably expressed (Thellin et al., 1999). Some of the commonly used housekeeping genes in mRNA expression studies are *GAPDH*, β-actin (*ACTB*), and different rRNAs (Bustin, 2000). For miRNA qRT-PCR assays, RNU6B (U6), and 5S rRNA are the commonly used reference transcripts (Peltier and Latham, 2008). Nevertheless, several studies have revealed important variations in the expression level of these genes in different tissues or altered physiological status that can lead to biased or incorrect results (Barber et al., 2005; Barbu and Dautry, 1989; Benz et al., 2013; Dheda et al., 2005; Ferguson et al., 2010; Tricarico et al., 2002). To overcome this obstacle, it has been recommended to perform a previous validation of the candidate reference genes for each cell type, tissue, and experimental design over the use of pre-established reference genes (Bustin et al., 2009). Many computational strategies have been suggested for the selection of suitable reference genes in specific experimental

conditions, like geNorm (Vandesompele et al., 2002), and NormFinder (Andersen et al., 2004) methods. For miRNA qRT-PCR studies, an specific approach named Concordance Correlation Restricted (CCR) algorithm was developed to adapt the MCR strategy to a normalizer-selection method for singleplex qRT-PCR assays (Wylie et al., 2011).

Concerning the normalization of transcriptomic technology outcomes, the correction of microarray data commonly relies on specific intensity-dependent strategies. These methods often make use of the intensities emitted by a selected group of genes to normalize the obtained data (Park et al., 2003). For this purpose, the global median of the log intensity ratios can be used, by a nonlinear method based either on the Locally Weighted Scatterplot Smoothing (LOWESS) or Locally Estimated Scatterplot Smoothing (LOESS) algorithms of Cleveland (Cleveland and Devlin, 1988; Cleveland and Grosse, 1991), or by a linear normalization (Yang et al., 2002).

Regarding RNA-seq data normalization, the most recurrent strategies are based on correcting read counts of each gene by the total number of reads accumulated in each sample (Evans et al., 2018). These methods are known as Reads Per Kilobase per Million mapped reads (RPKM, used for single-end reads), Fragments per Kilobase per Million mapped reads (FPKM, used for paired-end reads), and Transcripts Per Million (TPM) (Figure 2). These measures remove possible variations associated with gene-length and library-size differences, and make possible the expression comparison between different genes and assays (Conesa et al., 2016).

$$RPKM/FPKM_i = \frac{X_i}{l_i N} \cdot 10^9$$

$$TPM_i = \frac{X_i}{l_i} \cdot \frac{1}{\sum_j \frac{X_j}{l_j}} \cdot 10^6$$

Figure 2. Formulas for the calculation of Reads Per Kilobase per Million mapped reads (RPKM) / Fragments per Kilobase per Million mapped reads (RPKM), and Transcripts Per Million (TPM) measures.

i = measured feature (e.g. gene or exon); j = all features in a sample; X = number of reads; l = length (kilobases); N = total number of sequenced reads.

1.3. Sperm transcriptome

Spermatozoa are highly-differentiated cells that play an essential role in reproduction by providing the haploid paternal genome to the embryo. For a long time, this contribution was considered the only function of spermatozoa, but several studies have revealed that these cells encompass a wider range of implications at both spermatic and post-fertilization levels. Therefore, the biological relevance of sperm cells is not merely based on DNA, but also proteins, and RNA. Although the scarce sperm RNA amount (about 200 times less than somatic cells) was either ignored or considered a mere residual pool from previous spermatogenic stages (Jodar et al., 2013), recent findings have revealed that sperm RNAs include a functional and complex population of molecules, with a wide range of subtypes and roles (Hamatani, 2012; Hosken and Hodgson, 2014). For a detailed revision of this topic see Corral-Vazquez et al., 2017 in Publication 3.

Sperm RNAs comprise a coding fraction named mRNAs, but also a non-coding set of RNAs. ncRNAs can be divided, depending on their length, into long non-coding RNA (lncRNA), and small non-coding RNA (sncRNA) (Figure 3) (Corral-Vazquez et al., 2017). It has been estimated that mRNAs and lncRNAs constitute about 50 femtograms (fg) of the total sperm RNA fraction, while sncRNAs contribute with ~0.3 fg (Goodrich et al., 2013). Recently, the presence of circular RNAs (circRNAs) has been also characterized in testicular tissues and seminal plasma (Dong et al., 2016), although its identification in sperm cells is still in process.

Although ncRNAs were initially considered as junk RNA, many biological functions have been recently described for these molecules, highlighting their involvement in the regulation of gene expression (Luk et al., 2014).

1.3.1. Coding RNA

mRNAs are the transcript fraction containing coding sequences that can be translated into proteins. Sperm mRNAs are mainly located within the entire head and midpiece of the cell, concentrated close to the nuclear envelope, and constituting a structural part of the nuclear matrix

(Dadoune et al., 2005; Modi et al., 2005; Pessot et al., 1989; Wykes et al., 1997). Although intact mRNA molecules can be found in the sperm cell, a biased enrichment towards 3' ending sequences has been found, which has been attributed to a fragmentation tendency of these transcripts (Jodar et al., 2013).

The first discovered sperm transcript was the mRNA of the proto-oncogene *C-MYC*, involved in sperm capacitation and acrosomal reaction processes (Naz et al., 1991). From that finding, many studies have tried to characterize the sperm mRNA content, revealing the presence of transcripts with marked biological roles. An RNA profiling study performed by SAGE led to the classification of sperm mRNAs into 25 functional groups. The main categories were composed of 96 nuclear protein genes involved in the transcription process, and 84 ribosomal subunit genes related to protein synthesis (Zhao et al., 2006). A clear example was mRNAs encoding transcription factors like Nuclear Factor Kappa B (*NFjB*), Homeobox Protein Hox-2a (*HOX2A*), or Interferon Consensus Sequence-Binding Protein (*ICSBP*), which were found accumulated in human sperm nucleus (Dadoune et al., 2005). Additionally,

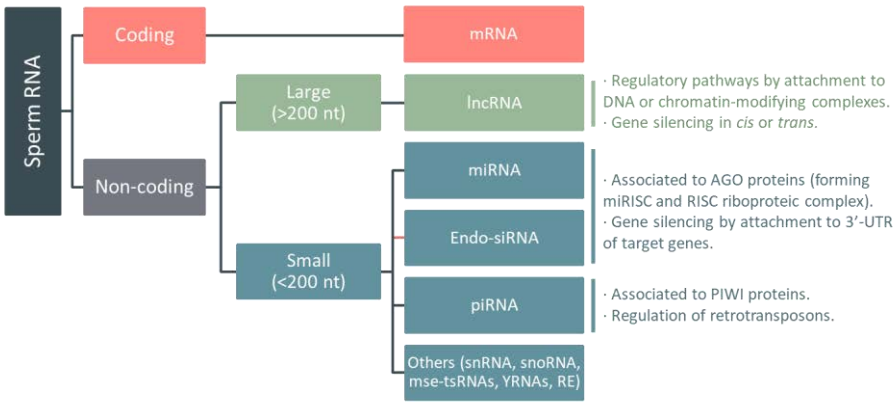


Figure 3. Classification of sperm RNA families and their main regulatory mechanisms.

mRNA = messenger RNA; lncRNA = long non-coding RNA; miRNA = microRNA; endo-siRNA = endogenous small interfering RNA; piRNA = PIWI-interacting RNA; snRNA = small nuclear RNA; snoRNA = small nucleolar RNA; mse-tsRNA = mature- sperm-enriched tRNA-derived small RNA; RE = Repetitive Rlements; AGO = Argonaute; miRISC = microRNA-induced Silencing Complex; RISC = RNA-induced Silencing Complex; UTR = Untranslated Region.

the analysis of non-fragmented mRNAs with higher expression levels indicated the enrichment of roles related to spermatogenesis, sperm functionality, fertilization, male fertility, and embryo development (Ostermeier et al., 2002; Zhao et al., 2006). Other sperm transcriptomic studies have described common abundant transcripts related to RNA binding, protein binding, zinc finger, transferase activity, adenosine triphosphate binding, and DNA binding among other functions (Lalancette et al., 2009).

1.3.2. Non-coding RNA

The non-coding fraction of the transcriptome (non-coding RNA or ncRNA) is categorized by molecule length. Therefore, monocatenary ncRNAs with a length of >200 nucleotides (nt) are named lncRNAs, while the sncRNA family is composed of transcripts of 20–300 nt. For a detailed revision of classification, biosynthesis, functions and implications of these non-coding transcripts see Corral-Vazquez and Anton, 2019 in Publication 4.

The biological functions of lncRNAs mainly comprise several epigenetic regulation pathways, affecting either the transcription of single mRNAs or even whole chromosomes (Bao et al., 2013). The main mechanism of this regulation consists on the attachment of lncRNAs to certain DNA regions or to diverse chromatin-modifying complexes (e.g., transcription factors or epigenetic modifiers) (Luk et al., 2014). By this recruitment, gene regulation in *cis* (flanking genes) or *trans* (genes from distal locations) is performed (Kanduri, 2016). Regarding sperm lncRNAs, they include specific transcripts that are especially abundant in mature sperm cells. Some of these transcripts have been found to be associated with chromatin, known as Chromatin-Associated RNAs (CARs). It has been suggested that CARs can influence genome architecture and regulate gene expression (Jodar et al., 2013).

Regarding sncRNAs, although all of them are defined as small molecules with a protein-attachment dependent regulatory functionality, they can be classified according to their size, biosynthesis pathway, type of associated proteins, and mechanism of action (Jodar et al., 2013). Mainly, they divide into microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and endogenous small interfering RNAs (endo-siRNAs) (Röther and Meister,

2011). In addition, other less abundant sncRNAs are also found in spermatozoa: repetitive elements, Transcription Start Sites (TSS)/promoter associated RNAs, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), mature-sperm-enriched tRNA-derived small RNAs (mse-tsRNAs), and YRNAs, which are components of the Ro60 ribonucleoprotein particle (Krawetz et al., 2011).

miRNAs are monocatenary RNA molecules of approximately 22 nt in size. Their regulatory mechanism is generally based on their complementary attachment to the 3' Untranslated Region (UTR) of target mRNAs, consequently inhibiting their expression (Luo et al., 2015). Alternatively, miRNAs can bind to other genomic regions, such as promoters, and act as transcriptional regulators (Place et al., 2008). A single miRNA can target multiple genes, and a single gene can be targeted by several miRNA. It is estimated that the expression of more than 60% of human genes is regulated by miRNAs (Luo et al., 2015). The characterization of human sperm miRNAs has revealed the existence of a specific expression profile in fertile males. A study performed in 10 fertile individuals showed a panel of 221 sperm miRNAs that were present in all samples (Salas-Huetos et al., 2014). The regulatory pathways of these ncRNAs were found to be involved in biological processes related to spermatogenesis and embryogenesis. Another characterization study of sperm ncRNA from fertile men revealed the presence of 35 enriched miRNAs involved in embryo development and cell growth processes (Krawetz et al., 2011). Besides, it has been found that several miRNAs, such as hsa-miR-34b-3p, hsa-miR-375, and hsa-miR-191-5p, are overexpressed in human sperm cells of infertile men (G. C. Ostermeier et al., 2005; Salas-Huetos et al., 2014).

piRNAs are 24-30 nt monocatenary RNAs, which are bound to the germ-line specific proteins PIWI (Chuma and Nakano, 2013). These molecules have been found to be the most abundant ncRNA in both human and mice sperm transcriptomes (Pantano et al., 2015; Röther and Meister, 2011). Biological functions of piRNAs mainly rely on retrotransposon regulation, serving as a protection mechanism against genome modifications produced by these elements (Chuma and Nakano, 2013). In sperm cells, Long Interspersed Nuclear Elements 1 (LINE1) constitute the most frequently regulated retrotransposons by piRNAs (Pantano et al.,

2015). The regulation of transposable elements can be performed either post-transcriptionally (degradation of complementary retrotransposon sequences by direct attachment) or by epigenetic modifications (e.g., participation in the regulation of *de novo* methylation in embryos and prospermatogonias) (Chuma and Nakano, 2013).

Endo-siRNAs are 22 nt RNA molecules whose regulatory mechanisms are similar to the protein-dependent gene silencing pathways of miRNAs, although both RNAs differ in their biosynthetic process. Besides gene silencing, it has been discovered that endo-siRNAs also guide epigenetic elements, such as histone methyltransferases, and promote the modification of chromatin conformation (Song et al., 2011). Some studies performed in mice have revealed higher expression levels of endo-siRNAs in male germ cells in comparison with other tissues. Their target sequences mainly corresponded to mRNAs (92%), but also ncRNAs (4%), pseudogenes (3%), and retrotransposons (1%) (Song et al., 2011). Additionally, it has been suggested that endo-siRNAs are necessary for post-fertilization processes such as the correct development of preimplantational embryos (Suh et al., 2010).

1.3.3. Sperm RNA and spermatogenesis

Spermatogenesis is a complex process that encompasses the differentiation from spermatogonia to mature spermatozoa. This process is composed of three main stages: the mitotic phase (division of spermatogonia to generate the primary spermatocytes), the meiotic phase (primary spermatocytes undergo a first meiotic division into secondary spermatocytes, and then a second meiotic division to generate spermatids), and the post-meiotic phase, also named spermiogenesis (spermatid differentiation into mature sperm cells) (Figure 4) (Gilbert, 2000; Hamatani, 2012). Most changes related to cell morphology that occur during spermatogenesis take place post-meiotically. In this last stage, round spermatids undergo structural modifications that include cytoplasm elimination and formation of acrosome and flagellum. Besides, most nuclear histones are replaced by protamines by the action of transition proteins, and thus chromatin becomes condensed in a highly-

packaged structure (Figure 4) (Cappallo-Obermann et al., 2011; Dadoune, 2009).

Transcriptional activity is continuous from the beginning of spermatogenesis until the formation of round spermatids when transcription is arrested coinciding with acrosome formation and chromatin repackaging (Miller et al., 2005). During the transcriptionally active phases, RNA synthesis significantly increases in two waves: One of them takes place during the premeiotic cell stage (spermatogonia), while the other one occurs at the pachytene stage of meiosis (Figure 4). During these highly regulated waves, specific transcripts needed for the correct development of each stage are generated (De Mateo and Sassone-corsi, 2014).

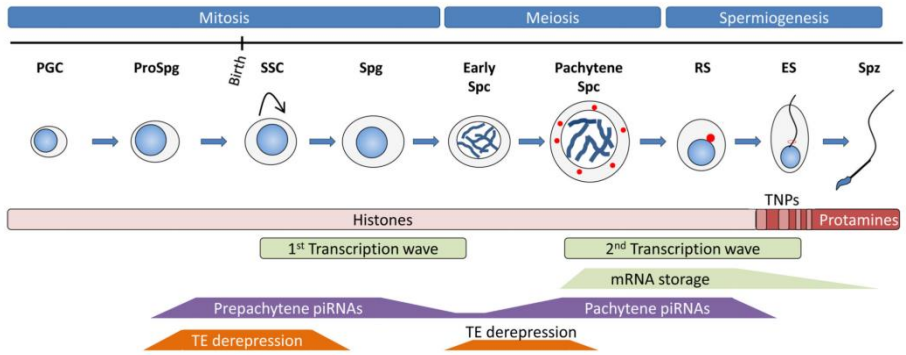


Figure 4. General view of the spermatogenesis process. The evolution of cell differentiation throughout the three stages of spermatogenesis is schematized chronologically. The replacement of histones by protamines, the two transcriptional waves, the action of piRNAs over transposable elements and the storage of meiotic mRNA are also indicated (adapted from De Mateo and Sassone-corsi, 2014).

PGC = Primordial Germ Cells; ProSpg = Prospermatogonia; SSC = Spermatogonial Stem Cell; Spg = Spermatogonia; Spc = Spermatocyte; RS = Round Spermatid; ES = Elongating Spermatid; Spz = Spermatozoa; TNP = Transition Protein; TE = Transposable Elements.

A search in Gene Ontology (GO) databases displayed 795 annotated human genes involved in spermatogenesis pathways (Carbon et al., 2009). During the mitotic stage, a study revealed the important expression of the transcription factors genes ju-nana (*JUN*) and the Nuclear Factor of Activated T-cells protein family (*NFAT*). *NFAT* is involved in immune response, and *JUN* expression has been associated with tumor genesis.

Therefore, a relation of both transcription factors with pluripotency preservation of undifferentiated spermatogonial preservation has been suggested (Zhu et al., 2016). At the meiotic stage, the expression of *HOX* genes has been associated with chromosomal organization. Besides, transcription factor Specificity Protein 1 (*SP1*) has a role in spermatocyte chromatin remodeling process, as well as in cell grown and differentiation, apoptosis, immune response, and DNA damage response (Zhu et al., 2016). Regarding spermiogenesis, the role of Transcription Factor 3 (*TCF3*) is involved in Wingless int-1 (Wnt) signaling pathway regulation, which is essential in the control process of spermatid elongation phase (Zhu et al., 2016). Transcripts expressed along spermiogenesis together with some previously synthetized RNAs are stored within ribonucleoproteic complex in the cytoplasm of mature spermatozoa (Jodar et al., 2013) as a preserving mechanism for later phases of transcriptional arrest (Dadoune, 2009; Miller et al., 2005).

Apart from stage-specific transcripts, it has been suggested that mRNAs have alternative mechanisms for regulating spermatogenesis**.** It has been discovered that sperm nucleosomes are non-randomly placed in certain loci that contain key genes with important functions in embryo development. This disposition promotes a more accessible transcription of such chromatin regions packaged by histones. It has been stated that sperm mRNAs could have a passive role in chromatin reorganization by influencing the protamine packaging process. Specifically, they could enhance nucleosome distribution in important sequences for embryo development, like key embryonic transcription factors and signaling pathway proteins (Hamatani, 2012).

Besides the importance of coding RNAs, several pathways involved in spermatogenesis regulation have been attributed to ncRNAs (Jodar et al., 2013). Among them, several lncRNAs have been characterized in male germ cells, with potential influence over the whole spermatogenetic process. Nevertheless, only a few lncRNAs have been actually detected in differentiated spermatozoa. Some examples are DMRT1 related gene (*DMR*), which targets the mRNAs of *DMRT1* (a transcription factor involved in germ cell development and related to spermatogonia differentiation regulation), and *Spga-lncRNA1/2*, which seems to play a role in the pluripotency preservation of spermatogonia. Additionally, Meiotic

Recombination Hot Spot Locus (*MRHL*) and Testis-specific X-linked (*TSX*) have been detected in pachytene spermatocytes and are involved in meiosis regulation (Luk et al., 2014).

The involvement of miRNAs in the regulation of human spermatogenesis has begun to be characterized. Although miRNA transcription is quite abundant throughout the whole process, the two described transcriptional waves also influence their biosynthesis. During the first wave, many miRNAs involved in the preservation of the spermatogonia niche are produced. Some examples discovered in mice are miR-20, miR-21, miR-34c, miR-106a, mir-135a, miR-146a, miR-182, miR-183 miR-17-92 cluster, miR-221, and miR-222 (De Mateo and Sassone-corsi, 2014; Kotaja, 2014). Other miRNAs take the opposite part in this regulation and promote spermatogonia differentiation, like let-7 family or miR-383 (De Mateo and Sassone-corsi, 2014; Tong et al., 2011). In the second transcriptional wave, miRNAs that regulate a wide range of cellular mechanisms are produced. These processes include massive apoptosis (regulated by miR-449 cluster and miR-34b/c), chromatin rearrangement, sperm maturation (both regulated by miR-18 and miR-214), or the transcription of Transition Proteins 1 and 2 (*TNP1* and *TNP2*) and Protamines 1 and 2 (*PRM1* and *PRM2*) genes (regulated by miR-469 and miR-122a) (De Mateo and Sassone-corsi, 2014; Kotaja, 2014).

piRNAs are considered as an important piece in the correct sperm differentiation process since retrotransposon control in germ cells is an essential issue for spermatogenesis development. In fact, the disruption of the biosynthetic process of piRNA or their associated proteins (PIWI) causes male infertility in human and mice (Carmell et al., 2007; Heyn et al., 2012; Kuramochi-Miyagawa, 2004). Generally, piRNAs are classified as pre-pachytene and pachytene piRNAs depending on the spermatogenesis stage in which they are synthesized (Figure 4) (Pillai and Chuma, 2012).

Regulatory role of endo-siRNA during sperm differentiation is not well characterized yet. Nevertheless, it has been suggested that they also take part in spermatogenesis regulation since hundreds of possible targets have been found to be expressed in male germ cells during spermatogenesis (Suh et al., 2010).

1.3.4. Sperm RNA and embryo development

Besides providing the paternal genome, human sperm contribution at fertilization also includes additional elements that are essential for early embryo activation and development. They comprise centrioles and proteins from the perinuclear theca of the sperm head, such as Phospholipase C-Z (PLC-Z), kinase signaling molecules, transcriptional factors, and structural proteins (Boerke et al., 2007). Additionally to these elements and paternal DNA, sperm RNAs are also released to the oocyte. However, these molecules were initially considered by several authors as remnants of late-spermatogenesis transcription with negligible influence on embryo development (Cummins, 2001). One of the main reasons for this stance was the limited amount of sperm RNA provided to the zygote (10-100 fg in humans) compared to the total oocyte RNA content (about 330,000 fg in humans) (Kocabas et al., 2006; Krawetz, 2005). Nevertheless, outcomes from multiple studies have proved that sperm RNAs are functional in the zygote.

A categorization of the total fraction of zygote-delivered sperm mRNA into three groups has been proposed, according to their origin and functionality. The first category is composed of spermatid-specific transcripts with no predicted embryo function (Boerke et al., 2007). *PRM2* or the putative fusogenic protein *GA17* would be included in this group since they have been described as potentially harmful for embryo development and are rapidly degraded after delivery (Ziyyat and Lefèvre, 2001). The second category comprises spermatid-specific mRNAs with a potential function for the zygote development, which remain stable until Zygote Genome Activation (ZGA) (Figure 5). Proteins encoded in this group are involved in stress-response pathways, as well as embryogenesis, morphogenesis, and implantation. Some of these transcripts would correspond to the A-kinase Anchoring Protein 4 (*AKAP4*, present in sperm flagellum and necessary for oocyte activation), *PLC-Z*, Forkhead Box G1 (*FOXG1B*, which is involved in early embryo development), Wnt Family Member 5A (*WNT5A*, related to cell differentiation), the Heat Shock Factor Binding Protein 1 (*HSBP-1*, involved in stress response after fertilization), the Signal Transducer and Activator of Transcription 4 (*STAT4*, potential modulator of transcription from the male pronucleus), and B1 cycline (*CCNB1*, important in cell cycle progression) (Boerke et al., 2007; Saunders

et al., 2007). Lastly, the third group includes the non-spermatic (extracellular) mRNAs that had been incorporated into sperm cells from the seminal plasma (Boerke et al., 2007; Valadi et al., 2007). These transcripts can be originated by multiple cell types from different tissues, and are mostly stored into <1µm extracellular microvesicles coming from the epididymis (epididymosomes) and the prostate (prostasomes) (H. Li et al., 2012; Raposo and Stoorvogel, 2013). Clusterin (*SGP2*) constitutes a well-known example of microvesicle-incorporated mRNA that has a role in sperm maturation, sperm-oocyte interaction, and embryo development (Boerke et al., 2007).

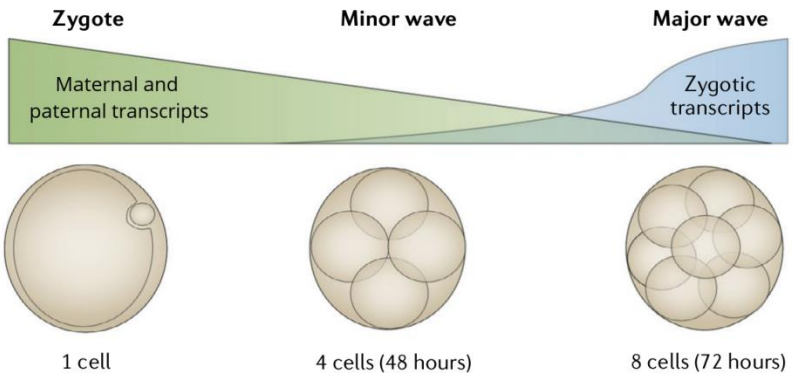


Figure 5. Zygote genome activation in human embryos. In the first 72 hours, the clearance of parental transcripts is coordinated with the activation of zygotic transcription. This process is initiated in an early minor wave of transcription at 4-cell stage, followed by a major wave in 8-cell stage embryos (Adapted from Schulz and Harrison, 2019).

Concerning ncRNAs, it has been described that these molecules are also involved in gene expression regulation during early embryo development. Specifically, sperm miRNAs and piRNAs remain intact after being released into the oocyte until ZGA (Boerke et al., 2007). This process is a period over which embryo transcription is gradually activated, comprising a minor transcription wave produced during the early cleavage divisions, and a major wave at the first division-cycle pause (Schulz and Harrison, 2019) (Figure 5). Additionally, some targets of sperm-specific miRNAs have been detected in metaphase II oocytes in mice (Amanai et al., 2006).

Moreover, it has also been observed that miRNAs can introduce epigenetic modifications [(1)] in early embryo stages (Pilu, 2011).

1.4. Sperm transcriptomic application to fertility biomarker research

It has been frequently suggested that expression alterations of some sperm RNA molecules in infertile patients could be used as fertility biomarkers (Anton and Krawetz, 2012). Accordingly, several transcriptomic studies have been developed to identify differentially expressed sperm transcripts associated with specific infertility status. These studies are of special interest for the diagnosis of IUO patients. Given that conventional clinical methods have failed to determine the molecular causes of this type of infertility, molecular biomarkers could be implemented as a complement for standard procedures.

1.4.1. Sperm mRNAs as infertility biomarkers

Through transcriptomic comparative studies, expression levels of sperm *PRM1* and *PRM2* have been strongly suggested as biomarkers of different seminal alterations, due to their essential role in sperm functionality (Table 7). Aberrant expression levels of both genes have been observed in infertile populations (Aoki et al., 2006; Rogenhofer et al., 2013; Steger et al., 2008). Altered protamine levels have also been associated to abnormal sperm features as motility (Kempisty et al., 2007; Lambard et al., 2004), and morphology (Savadi-Shiraz et al., 2015), as well as being suggested as biomarkers of cryptorchidism (Nguyen et al., 2009). A higher *PRM1* and *PRM2* expression have been also detected in testicular biopsies of azoospermic men (Haraguchi et al., 2009; Song et al., 2000).

Besides protamines, the differential expression of other sperm mRNAs in infertile populations has been identified (Table 7). It is the case of B Cell

[(1)] An example of these mechanisms would be paramutations. These are non-Mendelian mutations produced when the epigenetic alterations of one allele (paramutagenic allele) are transferred to the other allele (paramutable allele). This transmission results in a gene expression modification in the paramutable allele. These epigenetic changes remain even when the paramutagenic allele is gone (Pilu, 2011).

Lymphoma 2 (*BCL2*) (Steger et al., 2008), the Pregnancy-Specific B-1-glycoprotein 1 (*PSG1*), the Human Leukocyte Antigen-E (*HLA-E*), and Heat-Shock Protein A2 (*HSPA2*) (Avendaño et al., 2009; Motiei et al., 2013). Alterations of *HSPA2* level have also been detected in men with low sperm counts (Cedenho et al., 2006), along with Tyrosine Kinase Receptor A (*TRKA*) (Li et al., 2010), and *VASA* gene expression levels (Guo et al., 2007). Regarding morphology, some remarkable biomarker mRNAs that are highly related to spermatogenesis and sperm function have been suggested. Some examples are *PLC-Z1*, acrosomal vesicle protein 1 (*ACRV1*), and the genes of the outer dense fiber of sperm tails 1-4 (*ODF1-4*) (Platts et al., 2007; Savadi-Shiraz et al., 2015). Other spermatogenesis-related transcripts have been found as possible biomarkers of low sperm motility, like *TNP1* and *TNP2*, Annexin *A2* (*ANXA2*), Spermatid-specific Linker Histone H1-like Protein (*HILS1*), Lactate Dehydrogenase C (*LDHC*) or the Testis-specific Protein gene (*TPX-1*) (Table 7) (Jedrzejczak et al., 2007; Jodar et al., 2012; Wang et al., 2004).

Additionally to the search of possible biomarkers associated with fertility alterations and seminal abnormalities, RNA expression measures could be a novel approach that would supply the lack of effective diagnostic tools for detecting and predicting UMI. As commented above, no seminal alterations are detected in these individuals and therefore their diagnosis and prognosis represent an even bigger challenge. Although it is a relatively unexplored research area, Garrido et al (Garrido et al., 2009) found a set of mRNAs that displayed an altered expression in infertile individuals with normozoospermia: Trypsin 1 (*TRY1*), Gamma-Glutamyltransferase 1 Transcript Variant 3 (*GGT1*), and Calcium Binding Protein 39 Like gene (*CAB39L*) (Table 7).

Several studies have identified some sperm mRNAs with high potential to predict successful sperm retrieval from testicle biopsies of azoospermic men. Some examples are *ZMYND15*, *PRM1*, cyclin A1, B1 and B2 (*CCNA1*, *CCNB1* and *CNB2*), Chromodomain Y1 (*CDY1*), and the homolog of the *DAZ* gene, *BOULE* (Haraguchi et al., 2009; Hashemi et al., 2018; Kleiman et al., 2011). Additionally, some sperm transcripts have been associated to successful outcomes in assisted reproduction cycles, like *PRM1* and *PRM2*, and Alpha-2-macroglobulin (*A2M*) (Table 7) (Depa-Martynow et al., 2012; Depa-Martynów et al., 2007; Garcia-Herrero et al., 2011).

Table 7. List of possible infertility biomarker mRNAs described in human sperm cells.

Infertility						
BCL2 [1]	HLA-E [2]	HSPA2 [3]	PRM1 [4,5]	PRM2 [2, 4-6]	PSG1 [2]	CASP9 [6]
Oligozoospermia						
CREM [7]	HSF2 [8]	HSPA4 [8]	NIPBL [7]	PRM2 [7]	UBE2B [11]	SPZ-1 [7]
DDX3X [7]	HSP90 [8]	JMJD1A [7]	PARK7 [7]	SPATA4 [7]	VASA [12]	PIWI-LIKE-2 [13]
HSF1 [8]	HSPA2 [9]	MEA1 [7]	NRF2 [10]			
Teratozoospermia						
ACRV1 [14]	ODF1-4 [14]	PLCZ1 [14]	PRM1 [15]	PRM2 [15]	SPAM1 [14]	TNP2 [15]
Asthenozoospermia						
ANXA2 [16]	CRISP2 [19]	HILS1 [22]	nNOS [20]	OAZ3 [16]	PRM2 [24,25]	PIWI-LIKE-1 [13]
BRD2 [16]	eNOS [20]	LDHC [23]	NRF2 [10]	SPAG6 [27]	TNP1 [22]	SLC22A14 [27]
ROPN1 [17]	CABYR [17]	TNP2 [22]	HSPA4 [8]	TPX1 [23]	DKKL1 [21]	NDUFA13 [26]
ODF2 [18]	PRM1 [20,24,25]					
Oligoasthenozoospermia						
BDNF [28]	NRF2 [10]	TRKA [29]				
Non-obstructive azoospermia						
TGIFL [37]	CCNA1 [32]	PRM1 [32]	GGN [30]	TMEM225 [30]	SPATS1 [30]	UBQLN3 [30]
AKAP4 [24]	CCNB1 [32]	ESX1 [33]	GSG1 [30]	PRM2 [32,35]	RBMY1 [36]	WBSCR28 [30]
BOULE [31]	CCNB2 [32]	SPATA3 [30]	GTSF1L [30]	SPACA4 [30]	CAPN11 [30]	ZMYND15 [38]
HSFY [34]	CDY1 [31]	FSCN3 [30]	ADCY10 [30]	FAM71F1 [30]		
Unexplained male infertility						
TRY1 [39]	GGT1.3 [39]	CAB39L [39]				
Cryptorchidism						
AGP-1 [40]	EIF4G2 [40]	HNRPA1 [40]	PMS1 [40]	RPL9 [40]	SRP54 [40]	HSPCD35 [40]
ATF2 [40]	FKBP3 [40]	TR-α [40]	PPGB [40]	RPS10 [40]	SUI1 [40]	UBE2D2 [40]
CD47 [40]	GNAS [40]	KPNA2 [40]	TPX1 [40]	RPS3A [40]	TAF10 [40]	UBN1 [40]
CDC27 [40]	GSTM3 [40]	MCSP [40]	PRM1 [40]	SCAMP2 [40]	TCP11 [40]	USP25 [40]
CUL3 [40]	GUCA1A [40]	NUP155 [40]	VDAC3 [40]	TNPAIP3 [40]	SKP1 [40]	RANBP9 [40]
EEF1A1 [40]	GYG [40]	ODF1 [40]	RPL7 [40]	PPP2R2B [40]	SRP14 [40]	
Varicocele						
FasL [41]	HSF1 [8]	HSF2 [8]	HSPA1B [42]	HSPA4 [8]	PLC-Z [43]	
Altered outcomes in assisted reproduction						
A2M [44]	C1QB [44]	CFD [44]	DDIT4 [44]	LGALS1 [44]	MT1B [44]	S100A6 [44]
VSIG4 [44]	CAPG3 [44]	COX7B2 [44]	FTHL12 [44]	LGALS3 [44]	MT1F [44]	C19orf36 [44]
TGFB1 [44]	CCL18 [44]	CTSB [44]	ANKRD7 [44]	TMSL3 [44]	MT1G [44]	ANGPTL4 [44]
FTL [44]	CD163 [44]	CTSH [44]	GPX1 [44]	MARCO [44]	MT1X [44]	LOC645745 [44]
APOC1 [44]	CD63 [44]	CTSL [44]	GRN [44]	METRNL [44]	NPC2 [44]	TYROBP [44]
APOE [44]	CD68 [44]	CTSZ [44]	HMOX1 [44]	MPP1 [44]	PRM1 [45,46]	CDKN2D [44]
SPP1 [44]	VIM [44]	CXCR4 [44]	IFI30 [44]	MT1A [44]	PRM2 [45,46]	

(1) (Steger, 2008); (2) (Avendaño, 2009); (3) (Motiei, 2013); (4) (Aoki, 2006); (5) (Rogenhofer, 2013); (6) (Zalata, 2016); (7) (Montjean, 2012) (8) (Ferlin, 2010); (9) (Cedenho, 2006); (10) (Chen, 2012); (11) (Yatsenko, 2013); (12) (Guo, 2007); (13) (Giebler, 2018); (14) (Platts, 2007); (15) (Savadi-Shiraz, 2015); (16) (Jodar, 2012); (17) (Pelloni, 2018); (18) (Luo, 2017); (19) (Zhou, 2015); (20) (Lambard., 2004); (21) (Yan, 2018); (22) (Jedrzejczak, 2007); (23) (Wang, 2004); (24) (Aslani, 2011); (25) (Kempisty, 2007); (26) (Yang, 2017); (27) (Huo, 2017); (28) (Zheng, 2011); (29) (Li, 2010); (30) (Malcher, 2013); (31) (Kleiman, 2011); (32) (Haraguchi, 2009); (33) (Pansa, 2014); (34) (Stahl, 2011); (35) (Song, 2000); (36) (Kuo, 2004); (37) (Aarabi, 2008); (38) (Hashemi, 2018); (39) (Garrido, 2009); (40) (Nguyen, 2009); (41) (Del Giudice, 2010); (42) (Ji, 2014); (43) (Janghorban-Laricheh, 2016); (44) (Garcia-Herrero, 2011); (45) (Depa-Martynow, 2012); (46) (Depa-Martynów, 2007).

1.4.2. Sperm ncRNAs as infertility biomarkers

Although most research studies have been focused on coding RNAs, an increasing interest in sperm ncRNAs has recently emerged, based on the marked regulatory nature of these molecules.

Among small non-coding transcripts, the use of miRNAs as biomarker has been frequently considered. Comparative microarray analyses between fertile and infertile patients (asthenozoospermic and oligoasthenozoospermic) revealed a set of miRNAs overexpressed in these individuals (50 and 42 respectively), and a downregulation of others (27 and 44 respectively) (Abu-Halima et al., 2014a). Five of these miRNAs (miR-34b*, miR-34b, miR-34c-5p, miR-429, and miR-122) were validated by qRT-PCR and proposed by the authors as a panel of fertility biomarkers (Abu-Halima et al., 2014b). Another study compared the sperm miRNA expression levels between fertile and asthenozoospermic patients, and suggested miR-27b as a biomarker of low sperm motility (Zhou et al., 2015). This miRNA directly regulates the expression of Cysteine-rich Secretory Protein 2 gene (*CRISP2*), which is underexpressed in asthenozoospermic patients. Moreover, Salas-Huetos et al. compared the sperm miRNA profiles between fertile controls and individuals with asthenozoospermia, teratozoospermia, oligozoospermia, and UMI (Salas-Huetos et al., 2016, 2015, 2014). In these studies, specific miRNA expression profiles were detected in each group, and the differences were associated with the specific fertility problems present in each population. Regarding the biomarker research for other reproductive pathologies, some authors detected differential expression of sperm miR-15a in varicocele patients (Place et al., 2008).

Concerning the large fraction of ncRNAs, although the field of human sperm lncRNA research is less detailed, some comparative studies between fertile and infertile populations have revealed several potential fertility biomarkers. Zhan et al. found downregulation of the lncRNA HOTAIR in sperm cells of asthenozoospermic and oligozoospermic patients. This transcript is directly related to histone H4 acetylation of Nuclear Factor Erythroid-Derived 2-Like 2 (*Nrf2*), so the expression reduction of this lncRNA produces a downregulation of this gene, besides leading to ROS-related defects in sperm function (Zhang et al., 2015).

Besides, a recent study suggested three sperm lncRNAs (lnc32058, lnc09522, and lnc98487) with significant differential expression in asthenozoospermic patients when compared to fertile donors. Moreover, the expression of these three lncRNAs was negatively correlated with motility rate (X. Zhang et al., 2019).

1.4.3. Stable pairs of sperm transcripts as infertility biomarkers

Although most biomarker studies have been focused on the search of single-molecule approaches, the possibility of establishing a diagnostic panel of multiple related molecules has also been considered by some authors. Using the relative expression (or the expression ratio) of two molecules, any alteration can be detected from two different standpoints. Therefore, by increasing the measured inputs, a wider approach for detecting possible biological fluctuations is achieved, which allows rising the sensitivity of the biomarker system.

The presence of several pairs of sperm mRNAs with a relative stable expression in fertile men was previously been reported by Lalancette et al. (Lalancette et al., 2009). In patients with impaired spermatogenesis, a clear example of RNA pairs that show an alteration of their pairwise expression are *PRM1* and *PRM2* transcripts (Aoki et al., 2006; Steger et al., 2008). The alteration of the *PRM1* and *PRM2* ratio has also been found to be related with asthenozoospermia (Jedrzejczak et al., 2007; Kempisty et al., 2007; Lambard et al., 2004), and varicocele (Ni et al., 2014). Other examples of sperm biomarker transcripts pairs associated with infertility are Signal Recognition Particle 54 kDa Protein/Long-chain-fatty-acid—CoA Ligase (*SRP54*/*ACSBG2)*, *SRP54*/G Protein-coupled Receptor 137 (*GRP137)*, *SRP54*/Tetratricopeptide Repeat Domain 7A (*TTC7A*), *SRP54*/Ubiquitin-associated Domain-containing Protein 1 (*UBAC1.2*), and *UBAC1.2*/Ring Finger Protein 7 (*RNF7*), which exhibit a disrupted expression ratio in infertile patients with Dysplasia of the Fibrous Sheath (DFS) (Lima-Souza et al.,2012).

2 Objectives

The main objective of the present Doctoral Thesis is to characterize the human sperm transcriptome to set up the basis for developing new biomarkers of male fertility.

In order to fulfill the main goal of the project, the following specific objectives were undertaken:

1. To optimize specific methodologies of sperm RNA analysis using qRT-PCR and RNA-seq strategies.

2. To provide an integrative profiling and functional characterization of sperm mRNAs and lncRNAs by RNA-seq technologies.

3. To establish new fertility biomarkers among the transcriptomic cargo of the human spermatozoa.

3 Publications

Publication 1

Normalization matters: tracking the best strategy for sperm miRNA quantification

Corral-Vazquez, C., Blanco, J., Salas-Huetos, A.,
Vidal, F., Anton, E., (2017)

Mol Hum Reprod 23, 45–53

Quartile: Q1 (Reproductive Biology)

Impact Factor (2017): 3.449

Molecular Human Reproduction, Vol.23, No.1 pp. 45–53, 2017
Advanced Access publication on December 8, 2016 doi:10.1093/molehr/gaw072

molecular human reproduction

ORIGINAL ARTICLE

Normalization matters: tracking the best strategy for sperm miRNA quantification

Celia Corral-Vazquez, Joan Blanco, Albert Salas-Huetos, Francesca Vidal, and Ester Anton*

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

*Correspondence address. Genetics of Male Fertility Group, Unitat de Biologia Cel·lular (Facultat de Biociències), Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain; E-mail: ester.anton@uab.cat

Submitted on July 12, 2016; resubmitted on October 11, 2016; editorial decision on November 11, 2016; accepted on November 14, 2016

STUDY QUESTION: What is the most reliable normalization strategy for sperm microRNA (miRNA) quantitative Reverse Transcription Polymerase Chain Reactions (qRT-PCR) using singleplex assays?

SUMMARY ANSWER: The use of the average expression of hsa-miR-100-5p and hsa-miR-30a-5p as sperm miRNA qRT-PCR data normalizer is suggested as an optimal strategy.

WHAT IS KNOWN ALREADY: Mean-centering methods are the most reliable normalization strategies for miRNA high-throughput expression analyses. Nevertheless, specific trustworthy reference controls must be established in singleplex sperm miRNA qRT-PCRs.

STUDY DESIGN, SIZE DURATION: Cycle threshold (Ct) values from previously published sperm miRNA expression profiles were normalized using four approaches: (i) Mean-Centering Restricted (MCR) method (taken as the reference strategy); (ii) expression of the small nuclear RNA RNU6B; (iii) expression of four miRNAs selected by the Concordance Correlation Restricted (CCR) algorithm: hsa-miR-100-5p, hsa-miR-146b-5p, hsa-miR-92a-3p and hsa-miR-30a-5p; (iv) the combination of two of these miRNAs that achieved the highest proximity to MCR.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Expression profile data from 736 sperm miRNAs were taken from previously published studies performed in fertile donors ($n = 10$) and infertile patients ($n = 38$). For each tested normalizer molecule, expression ubiquity and uniformity across the different samples and populations were assessed as indispensable requirements for being considered as valid candidates. The reliability of the different normalizing strategies was compared to MCR based on the set of differentially expressed miRNAs (DE-miRNAs) detected between populations, the corresponding predicted targets and the associated enriched biological processes.

MAIN RESULTS AND THE ROLE OF CHANCE: All tested normalizers were found to be ubiquitous and non-differentially expressed between populations. RNU6B was the least uniformly expressed candidate across samples. Data normalization through RNU6B led to dramatically misguided results when compared to MCR outputs, with a null prediction of target genes and enriched biological processes. Hsa-miR-146b-5p and hsa-miR-92a-3p were more uniformly expressed than RNU6B, but their results still showed scant proximity to the reference method. The highest resemblance to MCR was achieved by hsa-miR-100-5p and hsa-miR-30a-5p. Normalization against the combination of both miRNAs reached the best proximity rank regarding the detected DE-miRNAs (Area Under the Curve = 0.8). This combination also exhibited the best performance in terms of the target genes predicted (72.3% of True Positives) and their corresponding enriched biological processes (70.4% of True Positives).

LARGE SCALE DATA: Not applicable.

LIMITATIONS, REASONS FOR CAUTION: This study is focused on sperm miRNA qRT-PCR analysis. The use of the selected normalizers in other cell types or tissues would still require confirmation.

WIDER IMPLICATIONS OF THE FINDINGS: The search for new fertility biomarkers based on sperm miRNA expression using high-throughput assays is one of the upcoming challenges in the field of reproductive genetics. In this context, validation of the results using singleplex assays would be mandatory. The normalizer strategy suggested in this study would provide a universal option in this area, allowing for

normalization of the validated data without causing meaningful variations of the results. Instead, qRT-PCR data normalization by RNU6B should be discarded in sperm-miRNA expression studies.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the 2014/SGR00524 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). The authors report no conflict of interest.

Key words: miRNA / sperm / normalization / infertility / qRT-PCR

Introduction

RNA expression profiling is a thorough approach to untangle complex pathways regarding gene transcription and underlying biological processes (Kanakachari *et al.*, 2015). It constitutes a key aspect in basic research, pharmacogenomics and molecular diagnostics (Bustin, 2002). Although RNA analysis can be performed by different techniques (Mestdagh *et al.*, 2009; Chu and Corey, 2012; Kanakachari *et al.*, 2015), quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR) constitutes one of the most widely used methods since it provides outstanding advantages of speed, accuracy and specificity (Mestdagh *et al.*, 2009). This procedure firstly encompasses a RNA template reverse-transcription to complementary DNA (cDNA) followed by a quantitative (q) PCR (Van Rooij, 2011). It is employed to identify differentially expressed genes in two situations, for the validation of high-throughput RNA screening (or multiplex analyses) and also in singleplex analyses (Pritchard *et al.*, 2015). Since the source of possible variations could be biological, but also artefactual, (e.g. differences in sample purification and collection, RNA isolation or target quantification) (Peltier and Latham, 2008), the accuracy of qRT-PCR is critically dependent on data normalization, meaning the process of reducing technical errors (Wylie *et al.*, 2011).

Depending on the number of inputs of each particular assay, proper normalization strategies should be selected. In high-throughput analysis, it is possible to take advantage of the huge amount of data generated in order to obtain a normalization reference value. Therefore, global-mean normalization (a.k.a. Mean-Centering/MC method) is commonly selected for this purpose (Mestdagh *et al.*, 2009). Instead, in singleplex qPCR assays (qRT-PCRs with a low number of inputs, generally tens of RNAs) (Wylie *et al.*, 2011), normalization against a single reference gene or a combination of few genes is generally accepted. The ideal standard of a normalizer molecule would consist of a single nucleic acid with an invariant expression across samples, displaying similar extraction properties, storage stability and quantification efficiency than the interrogated transcripts (Peltier and Latham, 2008). Unstable reference genes can drastically change the expression pattern and introduce flaws in results (Ferguson *et al.*, 2010). Thereby, the identification of appropriate reference normalizers is crucial in singleplex qRT-PCRs assays.

Sperm transcriptome analyses have become a main pillar of male fertility research (Li and Zhou, 2012). In this field, microRNA (miRNA) expression profiling is considered as a key to integrate gene expression regulation processes. MiRNAs are small monocatenary RNA molecules (22–24 nucleotides) that are classified into the family of small non-coding RNAs (sncRNAs) (Luo *et al.*, 2015). Their main function is post-transcriptionally regulating gene expression by altering their target mRNAs translation (Ambros, 2001). Generally, this process is triggered by the interaction of the miRNAs with the 3′ untranslated region of the target mRNA (Gangaraju and Lin, 2009). Their implication in many biological functions as development, cellular proliferation, differentiation and apoptosis has been extensively reviewed (Shenoy and Blelloch, 2014). Alterations in miRNA expression profiles have been related to human diseases, including male infertility (Lian *et al.*, 2009; Abu-Halima *et al.*, 2013; Salas-Huetos *et al.*, 2015).

MiRNA qRT-PCR data normalization supposes an even increased challenge as these molecules represent a tiny fraction within the total RNA mass amount. Moreover, this fraction can vary significantly across samples (Peltier and Latham, 2008), thus increasing the difficulty of detecting differential expression patterns. Mean-Centering Restricted (MCR) strategy, designed by Wylie *et al.*, (2011) as a variation of the normalizer MC method, was developed to calculate the mean expression value of fully-expressed miRNAs across samples. Nevertheless, this strategy loses robustness when less than a thousand inputs are analysed (Wylie *et al.*, 2011). In these cases, sncRNAs have been selected as reference controls, being RNU6B the most recurrent option for normalizing sperm miRNA qRT-PCR data (for a review of the literature see Table I). Nevertheless, different reasons put into question the accuracy of this strategy. First, the nature of these small nuclear RNAs (snRNAs) is different to the quantified molecules; second, it has been proved in different cell types that their expression profile is highly

Table I Normalization methods and reference genes used in previously published sperm microRNA (miRNA) quantification studies.

Normalization strategy	References
Mean of the 23 assays that were detected in all samples	Muñoz *et al.* (2015)
RNU6B	Cui *et al.* (2015)
RNU6B	Zhou *et al.* (2015)
MCR	Salas-Huetos *et al.* (2015)
miR-548q	Metzler-Guillemain *et al.* (2015)
RNU6B	Zhou *et al.* (2015)
RNU6B	Abu-Halima *et al.* (2014)
MCR	Salas-Huetos *et al.* (2014)
Calibrator sample from control group	Ji *et al.* (2014)
RNU6B	Abu-Halima *et al.* (2013)
Hsa-let-7b-5p	Belleannée *et al.* (2013)
RNU6B	Li *et al.* (2012)
RNU6B	Marczylo *et al.* (2012)

variable among individuals (Benz *et al.*, 2013; Lamba *et al.*, 2014); moreover, their invariant and uniform expression in human spermatozoa has not yet been documented.

Other strategies have been considered for the identification of normalizer miRNAs with a uniform and ubiquitous expression. Some algorithms have been designed to select panels of miRNA candidates from high-throughput data. For example, geNorm and NormFinder have been previously implemented (Peltier and Latham, 2008) and increasingly employed. These approaches were designed to select as normalizer miRNAs whose expression values better correlate to the average global expression. Another example is the Concordance Correlation Restricted (CCR) algorithm. This approach was implemented for identifying miRNAs whose expression better resembles MCR values based on the Pearson correlation coefficient yielded to correction by the Concordance Correlation Coefficient (Lin, 1989). In CCR, only fully-expressed miRNAs are considered. This restriction guarantees that the normalizers proposed are selected under an appropriate quality criteria (Wylie *et al.*, 2011).

In this study, we have compared four strategies to normalize sperm miRNA qRT-PCR data derived from studies performed in fertile and infertile individuals (Fig. 1); (i) MCR strategy (widely accepted due to its high reliability and here considered as the reference method); (ii) the expression value of the most commonly used normalizer in the literature: RNU6B; (iii) the individual expression of four miRNAs selected by the CCR algorithm for achieving the highest proximity to MCR method (hsa-miR-100-5p, hsa-miR-146b-5p, hsa-miR-92a-3p and hsa-miR-30a-5p); (iv) a combination of two of these miRNAs that displayed the better resemblance to MCR method.

The accuracy of these strategies when compared to MCR method was based on the set of differentially expressed miRNAs (DE-miRNAs) detected between populations, their predicted targets and the associated enriched biological processes. The primary endpoint of the study is to provide the best normalizer candidates for human sperm miRNA qRT-PCR singleplex assays.

Materials and Methods

Samples used and raw data analysis

Expression profile data were taken from previously published studies carried on 48 individuals classified into three groups: fertile donors (n = 10; Salas-Huetos *et al.*, 2014), infertile patients with different seminal alterations (n = 30; Salas-Huetos *et al.*, 2015) and normozoospermic infertile patients (n = 8; Salas-Huetos *et al.*, 2016). Samples evaluated in these studies were collected under the same conditions, processed through the same protocols and subjected to the same quality controls. Briefly, semen samples were obtained by masturbation after 3 days of sexual abstinence and processed by the somatic cell lysis method (Goodrich *et al.*, 2007). Isolation of total sperm RNA was performed by the TRIzol® protocol (Life Technologies, Carlsbad, CA, USA) and the resulting RNA fractions were treated with rDNase I (Life Technologies). The absence of DNA was confirmed by a RT-PCR followed by a PCR for Protamine one gene (*PRM1*) and for Glyceraldehyde 3-phosphate dehydrogenase gene (*GAPDH*). The absence of somatic RNA was checked by verifying the lack of ribosomal RNA in the samples using the RNA 6000 Nano chip (Agilent Technologies, Wilmington, DE, USA). Specifically, the absence of leukocyte RNA was verified by a RT-PCR followed by a PCR for the *CD45* gene (Supplementary Table I).

About 50 ng of sperm RNA were reverse transcribed (TaqMan MicroRNA Reverse Transcription kit, Life Technologies). cDNA was pre-amplified (TaqMan PreAmp Master Mix, and Megaplex PreAmp Primers, Life Technologies). Subsequently, a total of 736 miRNA was quantified by a qPCR (TaqMan Array Human MicroRNA A and B Cards Set v.3.0 kit, Life Technologies). Cycle threshold (Ct) values were processed by SDS and RQ Manager v.1.2 software (Life Technologies) and classified as Determined ($15 \leq Ct \leq 35$), Undetermined ($Ct \geq 35$) or Unreliable (no Ct value or $Ct < 15$).

Normalization methods

Expression data were normalized by subtracting the normalizing value from the raw Ct values. This process was performed by four different methods. Firstly, Ct values were normalized using the MCR method. R statistical computing environment v.3.2.3 (www.r-project.org) and the HTqPCR v.3.2 package for high-throughput analyses of qPCR data (www.bioconductor.org) were employed for calculations. Additionally, Ct values were normalized against the expression level of RNU6B (mean value of three replicates). Moreover, Ct values were also normalized against the expression of four miRNAs selected by CCR algorithm: hsa-miR-100-5p, hsa-miR-146b-5p, hsa-miR-92a-3p and hsa-miR-30a-5p. These four miRNAs were tested as normalizers separately by using their expression level as normalization value. Finally, Ct values were additionally normalized using the mean expression value of the two miRNAs that achieved the best results (i.e. hsa-miR-100-5p combined with hsa-miR-30a-5p).

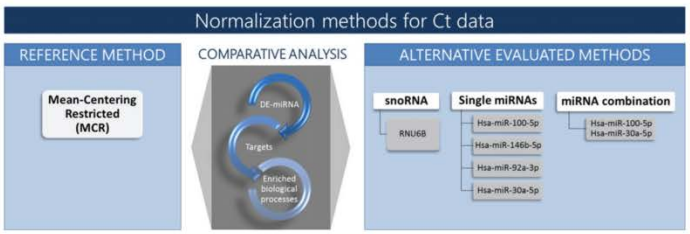


Figure 1 Normalization strategies for sperm microRNA (miRNA) expression data evaluated in this study. Comparative analyses between the reference method and the alternative strategies were based on the number of differentially expressed miRNAs (DE-miRNAs), the corresponding predicted target genes, and the associated enriched biological processes. Ct, cycle threshold; snoRNA, small nucleolar RNA.

Evaluation of ubiquity and uniformity across samples and populations

Expression ubiquity for RNU6B was verified across samples by checking the inclusion of its Ct values among the Determined group ($15 \leq Ct \leq 35$). The ubiquity of the hsa-miR-100-5p, hsa-miR-146b-5p, hsa-miR-92a-3p and hsa-miR-30a-5p was already assured by using the CCR algorithm in their selection (this approach is based on constantly present miRNAs). The expression uniformity across samples of each tested normalizer was assessed by variance calculation: lower variance values indicate higher uniformity. To check expression uniformity between fertile and infertile populations, the mean expression value of each tested normalizer was compared using the Non-parametric Wilcoxon test; *P*-values <0.01 were considered significant after Benjamini–Hochberg False Discovery Rate (FDR) correction. For this purpose, The R Graphical User Interface Deducer (http://www.deducer.org/) was used.

Normalized data analyses

To identify DE-miRNAs in the infertile population, normalized Ct values of each miRNA were compared by non-parametric paired Wilcoxon rank sum test. *P*-values <0.01 were considered statistically significant after FDR correction. Analyses were performed separately for each normalization method.

DE-miRNAs target genes were predicted by DIANA micro-T CDS v.5.0 software (http://diana.imis.athena-innovation.gr/DianaTools/), considering a miRNA target gene score ≥0.8 as a threshold (highly restrictive). The enrichment of biological processes among targets was evaluated by a Gene Ontology (GO) search, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources v.6 (https://david.ncifcrf.gov/) considering as significant a *P*-value of <0.05 after Bonferroni correction. REViGO (*Reduce and Visualize Gene Ontology*) web server (http://revigo.irb.hr) (Supek *et al.*, 2011) was employed, using a 0.7 cutoff value (default), to semantically classify and summarize GO terms in Clusters Representatives and Superclusters.

To evaluate the similitude of the DE-miRNAs results when comparing the different tested strategies with MCR method, Receiver Operating Characteristic (ROC) curve analyses were performed. Areas Under the Curves (AUC) were calculated from the sensitivity and specificity values obtained from each data set. These two indices are based on classification of the results into four categories: True Positives (TPs: results obtained either using the tested normalizer and the reference MCR method), False Positives (FPs: results obtained from the tested normalizer but not from MCR), True Negatives (TNs: negative results using the tested normalizer and the reference MCR method) and False Negatives (FNs: results obtained from MCR but not from the tested normalizer). Sensitivity was calculated by the formula: $\text{Sentitivity} = \frac{TPs}{TPs + FNs}$. Specificity was calculated by the formula: $\text{Specificity} = \frac{TNs}{TNs + FPs}$. The proximity of the tested normalizing strategies to the reference MCR method was classified into the traditional academic point system: *Excellent* ($0.9 \leq AUC \leq 1$), *Good* ($0.8 \leq AUC < 0.9$), *Fair* ($0.7 \leq AUC < 0.8$), *Poor* ($0.6 \leq AUC < 0.7$) and *Failed* ($AUC < 0.6$).

Regarding target genes and enriched biological processes, the percentages of TPs, FPs and FNs were calculated for each normalizer with respect to MCR results.

Results

Ubiquity and uniformity across samples and populations

All tested normalizers (RNU6B and the four miRNAs) were confirmed as ubiquitous across samples. Regarding variance analysis, three miRNAs candidates showed the lowest values ($\text{variance}_{\text{miR-92a-3p}} = 0.84$; $\text{variance}_{\text{miR-100-5p}} = 1.03$; $\text{variance}_{\text{miR-30a-5p}} = 1.23$) indicating their high uniformity across samples. In contrast, RNU6B and hsa-miR-146b-5p showed higher variance ($\text{variance}_{\text{miR-146b-5p}} = 1.80$; $\text{variance}_{\text{RNU6B}} = 2.53$) and thus lower expression uniformity (Fig. 2). Concerning the evaluation of possible differential expressions of the tested candidates between the fertile and infertile group, none of the normalizing candidates showed a significant variation ($P_{\text{miR-100-5p}} = 0.18$; $P_{\text{miR-146b-5p}} = 0.82$; $P_{\text{miR-92a-3p}} = 0.99$; $P_{\text{miR-30a-5p}} = 0.33$; $P_{\text{RNU6B}} = 0.81$) (Fig. 3).

DE-miRNAs

Data normalized by the reference MCR method revealed 43 DE-miRNAs between the fertile and infertile populations. Table II presents the number of DE-miRNAs identified by the different normalization strategies ranging from 0 (when using RNU6B or hsa-miR-92a-3p) to 25 (when using the hsa-miR-100-5p and hsa-miR-30a-5p combination).

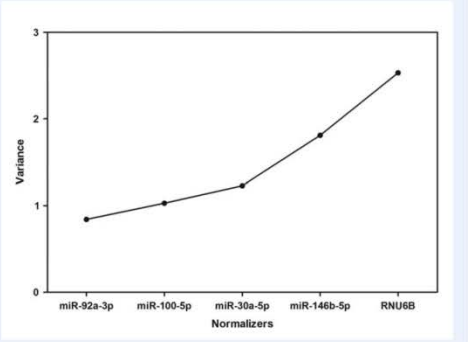


Figure 2 Normalizer expression uniformity across samples. Variance value is inversely proportional to uniformity.

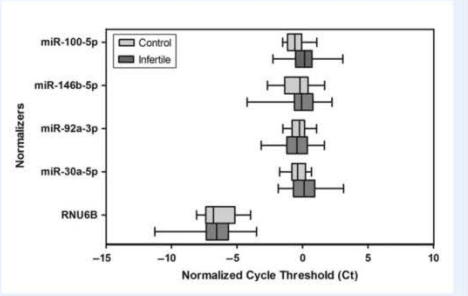


Figure 3 Expression levels of each normalizer in the control and infertile populations. The boxes mark the interquartile range with medians as vertical line; whiskers indicate the maximum and minimum values.

Table II The number of True Positives (TPs), False Negatives (FNs), True Negatives (TNs), False Positives (FPs) and the Area Under the Curve (AUC) for each normalizer combination evaluated with respect to Mean-Centering Restricted (MCR) differentially expressed miRNAs (DE-miRNAs) prediction.

Normalizer	DE-miRNAs	TPs	FNs	TNs	FPs	Sensitivity	Specificity	AUC
Hsa-miR-100-5p + hsa-miR-30a-5p	25	24	19	692	1	0.6	1.0	0.8
Hsa-miR-30a-5p	14	13	30	692	1	0.3	1.0	0.6
Hsa-miR-100-5p	11	11	32	693	0	0.3	1.0	0.6
Hsa-miR-146b-5p	2	2	41	693	0	0.0	1.0	0.5
Hsa-miR-92a-3p	0	0	43	693	0	0.0	1.0	0.5
RNU6B	0	0	43	693	0	0.0	1.0	0.5

The scores were sorted according to their AUC, classified as *Good* ($0.8 \leq AUC < 0.9$), *Poor* ($0.6 \leq AUC < 0.7$) and *Failed* ($AUC < 0.6$) proximity to MCR results.

The full list of DE-miRNAs for every normalizing method is available in Supplementary Table II. The number of TPs varied from 0 to 24 DE-miRNAs, while the number of FNs ranged from 19 to 43 DE-miRNAs (Table II). TNs and FPs values oscillated within a smaller rank ($n = 692–693$ and $n = 0–1$, respectively; Table II). The sensitivity displayed by each normalizer option showed a wide range of figures (sensitivity range = 0.0–0.6; Table II), whereas specificity showed high values in all cases (specificity values = 1.0; Table II). ROC curve analyses (Fig. 4) revealed a *Failed* proximity to MCR by RNU6B, hsa-miR-92a-3p and hsa-miR-146b-5p normalized data (AUC = 0.5; Table II). A *Poor* similarity was displayed by the hsa-miR-100-5p and the hsa-miR-30a-5p normalized data (AUC = 0.6; Table II). The normalization by the mean expression of these two last miRNAs combined (hsa-miR-100-5p + hsa-miR-30a-5p) led to a *Good* proximity rank (AUC = 0.8; Table II).

Predicted target genes and biological processes of the DE-miRNAs

About 7854 target genes were predicted from the DE-miRNAs obtained through MCR normalized data. For the compared normalizing strategies, no target genes were predicted in the cases of RNU6B and hsa-miR-92a-3p since no DE-miRNAs were previously detected. The number of target genes ranged from 235 (in the case of hsa-miR-146b-5p) to 5676 (when using the combination of hsa-miR-100-5p and hsa-miR-30a-5p) (full lists available in Supplementary Table III). Among the methods that use a single reference normalizer, the highest percentage of TPs was achieved by hsa-miR-100-5p normalized data (40.9%) followed by hsa-miR-30a-5p (34.2%) (Table III). The percentages of FNs obtained in these cases were 59.1% and 65.5%, respectively. The lowest TP percentages were displayed by hsa-miR-146b-5p (3.0%) normalized data. In this case, FN percentage was higher (97.0%). FP values remained ≤0.3 in all tested normalizers. The hsa-miR-100-5p and hsa-miR-30a-5p combination allowed obtaining better TP values than when using any single normalizer (72.3%). This strategy also showed the lowest FN rate (27.4%) and a low FP percentage (0.3%).

Moreover, a total of 71 biological processes were found to be significantly enriched among the DE-miRNAs targets obtained from MCR normalized data. Alternative methods showed different numbers of results, varying from 0 (in the case of hsa-miR-146b-5p, since only 235 target genes were predicted) to 53 (in the case of the hsa-miR-100-5p and hsa-miR-30a-5p combination) biological processes (full lists available in Supplementary Table VI). No biological processes were predicted in the cases of RNU6B and hsa-miR-92a-3p since no DE-miRNAs—and thus predicted target genes—were previously detected. Among the methods that use a single reference normalizer, the highest percentage of TPs was achieved by hsa-miR-100-5p (42.3%) and hsa-miR-30a-5p (33.8%) (Table III). These result candidates showed a low rank of FNs (57.8–61.8%). As expected, no TP

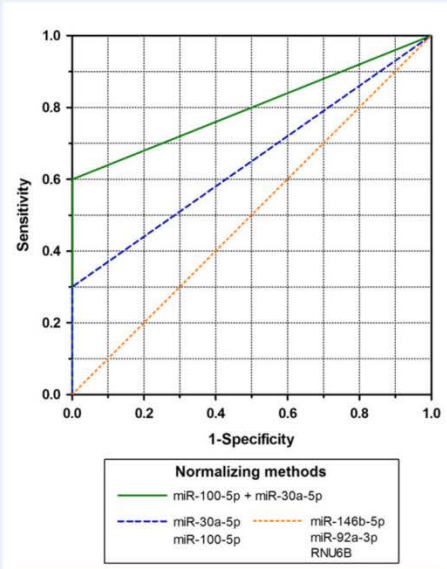


Figure 4 Receiver Operating Characteristic (ROC) curves corresponding to the several normalizing methods compared in this study. Areas Under the Curve (AUCs) displaying a *Good* proximity ($0.8 \leq AUC < 0.9$) to Mean-Centering Restricted (MCR) are represented by a continuous line; AUCs displaying a *Poor* proximity ($0.6 \leq AUC < 0.7$) to MCR are represented by a dashed line; AUCs displaying a *Failed* proximity ($AUC < 0.6$) to MCR are represented by a dotted line.

percentage was achieved in the case of RNU6B, hsa-miR-92a-3p and hsa-miR-146b-5p. In all strategies, FP percentages remained ≤4.2%. The hsa-miR-100-5p and hsa-miR-30a-5p combination led to the highest TP rate (70.4%) and the lowest FN percentage (25.4%).

REViGO classification summarized the GO terms associated to MCR normalized data in five Superclusters: 'Regulation of transcription from RNA polymerase II promoter', 'Cell part morphogenesis', 'Protein transport', 'Chromatin modification' and 'Movement of cell or subcellular component' (Fig. 5a). Three Superclusters were derived from hsa-miR-100-5p normalized data: 'Regulation of transcription from RNA polymerase II promoter', 'Movement of cell or subcellular component' and 'Embryonic morphogenesis' (Fig. 5b). Hsa-miR-30a-5p normalized data results were summarized uniquely in the Supercluster of 'Regulation of transcription from RNA polymerase II promoter' (Fig. 5c). The combination of hsa-miR-100-5p and hsa-miR-30a-5p revealed a distribution with five Superclusters: 'Regulation of

Table III True positive (TP), False Positive (FP) and False Negative (FN) percentages of target genes and enriched biological processes regarding the different normalizers evaluated in relation to Mean-Centering Restricted (MCR) results.

	Target genes				Enriched biological processes			
Normalizer	**Number**	**% TPs**	**% FNs**	**% FPs**	**Number**	**% TPs**	**% FNs**	**% FPs**
Hsa-miR-100-5p + hsa-miR-30a-5p	5676	72.3	27.4	0.3	53	70.4	25.4	4.2
Hsa-miR-100-5p	3214	40.9	59.1	0.0	30	42.3	57.8	0.0
Hsa-miR-30a-5p	2689	34.2	65.5	0.3	22	33.8	61.8	4.2
Hsa-miR-146b-5p	235	3.0	97.0	0.0	0			
Hsa-miR-92a-3p								
RNU6B								

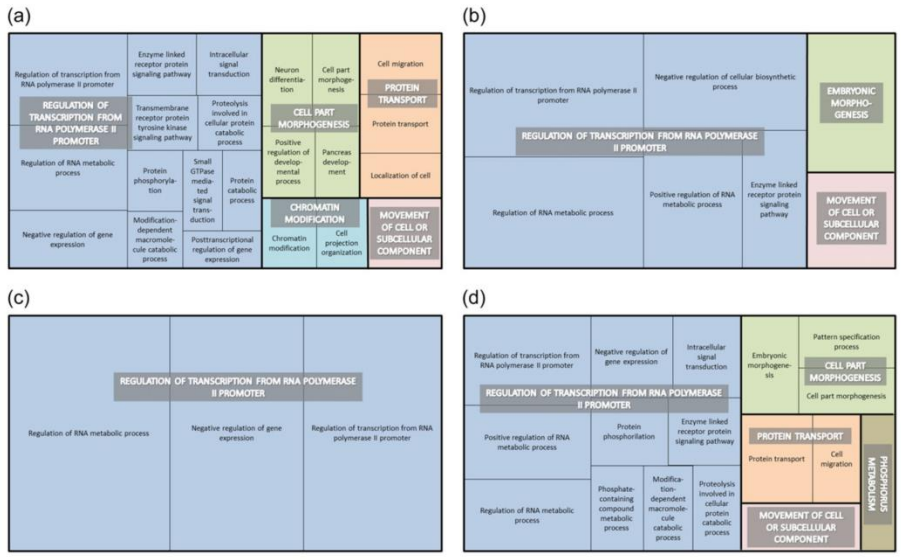


Figure 5 REViGO (Reduce and Visualize Gene Ontology) diagrams displaying the enriched biological processes obtained from data normalized using: (**a**) Mean-Centering Restricted (MCR); (**b**) Hsa-miR-100-5p; (**c**) Hsa-miR-30a-5p and (**d**) Hsa-miR-100-5p + hsa-miR-30a-5p. The predicted Gene Ontology (GO) terms are semantically classified in Cluster Representatives (lower-case letters) and Superclusters (capital letters). The size of each box is proportional to the number of GO terms related to the described biological function.

transcription from RNA polymerase II promoter', 'Cell part morphogenesis', 'Protein transport', 'Movement of cell or subcellular component' and 'Phosphorus metabolism' (Fig. 5d).

Discussion

In the search for a valid normalizer molecule, able to be used in sperm miRNA singleplex quantifications, the validation of their ubiquity and uniformity across samples and populations is mandatory to guarantee their trustworthy use as a normalizer. Our results indicated that the expression of RNU6B, although ubiquitous and uniform between fertile and infertile populations, did not completely fulfill the requirements of a reliable normalizer due to its high variance. Alternatively, the four miRNA candidates achieved better results in the assessment of these three parameters.

The limitations of RNU6B were further verified based on the results of the DE-miRNAs analysis, their predicted target genes and the associated enriched biological processes. Data normalized against the expression of this molecule barely resembled the reference MCR results, showing a scant or null proximity of the corresponding ROC curves and TP percentages. The differences between the enriched biological processes based on the DE-miRNAs predicted targets from the differentially normalized data panels exemplify the consequences of an incorrect normalization strategy. Specifically, no target genes were detected in the prediction performed from RNU6B normalized results compared to the 7854 target genes determined from MCR normalized data. Moreover, none of the 47 enriched biological functions associated to MCR normalization were either detected by using RNU6B normalized results. Therefore, completely different conclusions would come out from the same data. As stated in the introduction, RNU6B has been frequently used in the literature as an endogenous reference gene for sperm miRNA profiling normalization (Li *et al.*, 2012; Abu-Halima *et al.*, 2013, 2014; Zhou *et al.*, 2015). Our results advise against the election of this snRNA for this purpose. Actually, the use of this molecule in miRNA quantification has been also questioned by other authors in other cell types (Wotschofsky *et al.*, 2011; Benz *et al.*, 2013; Wang *et al.*, 2013; Lamba *et al.*, 2014; Das *et al.*, 2016).

Concerning the four tested miRNAs, results derived from the use of hsa-miR-100-5p and hsa-miR-30a-5p stood out compared to the hsa-miR-146b-5p and hsa-miR-92a-3p outputs. Although hsa-miR-92a-3p was proved to be the most uniformly expressed molecule, data normalized by this miRNA showed a poor similarity to MCR, not better than the obtained from RNU6B. Hsa-miR-146b-5p, besides maintaining a less uniform expression level, did not achieve a good proximity rank either. On the contrary, hsa-miR-100-5p and hsa-miR-30a-5p showed good results concerning ubiquity and uniformity across samples and populations besides displaying high proximity rates to MCR along all the performed analyses.

The possibility of selecting only one normalizer could be considered since normalization against the expression of a single reference control has been proved to be satisfactory in certain cases (Ohl *et al.*, 2005). Nevertheless, the selection of multiple reference molecules offers a more robust normalization (Vandesompele *et al.*, 2002) since it minimizes the effect of sporadic technical failures and reduces a potential risk of fluctuations in their expression across samples and populations. This fact is reflected in our study since the usage of the combination of hsa-miR-100-5p and hsa-miR-30a-5p improved the outputs obtained in comparison of data normalized using these two normalizers separately. This can be easily seen when ultimately comparing the DAVID functional annotation results derived from the different normalization strategies through REViGO: diagrams obtained from MCR and this miRNA combination normalized data display a close similarity, while the remaining normalizing methods give rise to fairly different Supercluster distributions. All these evidences confirm a higher reliability of the hsa-miR-100-5p and hsa-miR-30a-5p combination in the miRNA normalization of data from sperm singleplex qPCR studies.

The incorporation of additional molecules to this panel (i.e. hsa-miR-146b-5p and/or hsa-miR-92a-3p) was considered and tested. Nevertheless, no improvement was observed in the results obtained. For example, incorporating hsa-miR-92a-3p leads to an identical AUC value and reduces the percentage of TP target genes to 72.1% compared to the combination of hsa-miR-100-5p and hsa-miR-30a-5p, while adding hsa-miR-146b-5p to the panel reduces the AUC value to 0.7 and the percentage of TP target genes to 66.4% (data not shown). In the particular case of hsa-miR-146b-5p, although this miRNA had been previously taken as reference control in other expression studies (Torres *et al.*, 2013; Solayman *et al.*, 2016), no previous use or validation of this normalizer in sperm miRNA quantifications had been reported so far. Besides these considerations, the inclusion of additional assays in the proposed combination of normalizers would imply an increase of the allocated costs of the study, which according to the benefits, can be considered unnecessary. Altogether, these reasons converge in the inconvenience of widening the number of molecules included in this selection and point out the use of the two miRNAs hsa-miR-100-5p and hsa-miR-30a-5p combined as the best strategy to normalize sperm miRNA quantification data.

Considering other normalization strategies, besides the ones reported in this article, two previous studies have made use of specific miRNAs: miR-548q (Metzler-Guillemain *et al.*, 2015) and let-7b-5p (Belleannée *et al.*, 2013). Nevertheless, no supporting analyses testing the normalizing potential of these miRNAs have been provided so far. On the other hand, from a study performed by Applied Biosystems in 38 different tissues and 59 cell lines (Wong *et al.*, 2007), a set of six miRNAs (i.e. hsa-miR-26b, hsa-miR-92 and hsa-miR-92 N for tissues; hsa-miR-423, hsa-miR-374 and hsa-miR-16 for cell lines) are recommended as qRT-PCR data normalizers. However, these analyses did not comprise isolated spermatozoa among the analysed cells types, despite including testicular tissue.

In summary, evidence compiled in this study constitute a strong basis to propose the combination of hsa-miR-100-5p and hsa-miR-30a-5p as an optimal normalization strategy for sperm miRNAs quantification. The combination of these two normalizers is suggested as the best election among all the interrogated candidates since it constitutes a balanced option between accuracy and cost. In spite of the widespread use in the literature of RNU6B for sperm miRNA singleplex qRT-PCR data normalization, our findings strongly advise against this strategy.

Supplementary data

Supplementary data are available at *Molecular Human Reproduction* online.

Authors' roles

C.C.-V. execution of study, data collection, data analysis and interpretation, manuscript preparation and critical discussion. J.B. conception and design, data analysis and interpretation, manuscript preparation and critical discussion. F.V. manuscript preparation and critical discussion. A.S.-H. sample and data collection, 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 2014/SGR00524 project (Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya, Catalonia, Spain); the 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

None declared.

References

Abu-Halima M, Hammadeh M, Backes C, Fischer U, Leidinger P, Lubbad AM, Keller A, Meese E. Panel of five microRNAs as potential biomarkers for the diagnosis and assessment of male infertility. *Fertil Steril* 2014;**102**:989–997.

Abu-Halima M, Hammadeh M, Schmitt J, Leidinger P, Keller A, Meese E, Backes C. Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. *Fertil Steril* 2013;**99**:1249–1255.e16.

Ambros V. microRNAs: tiny regulators with great potential. *Cell* 2001;**107**: 823–826.

Belleannée C, Légaré C, Calvo É, Thimon V, Sullivan R. MicroRNA signature is altered in both human epididymis and seminal microvesicles following vasectomy. *Hum Reprod* 2013;**28**:1455–1467.

Benz F, Roderburg C, Vargas Cardenas D, Vucur M, Gautheron J, Koch A, Zimmermann H, Janssen J, Nieuwenhuijsen L, Luedde M *et al.* U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis. *Exp Mol Med* 2013;**45**:e42.

Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* 2002;**29**:23–39.

Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Ther* 2012;**22**:271–274.

Cui L, Fang L, Shi B, Qiu S, Ye Y. Spermatozoa micro ribonucleic acid-34c level is correlated with intracytoplasmic sperm injection outcomes. *Fertil Steril* 2015;**104**:312–317.e1.

Das MK, Andreassen R, Haugen TB, Furu K. Identification of endogenous controls for use in miRNA quantification in human cancer cell lines. *Cancer Genom Proteom* 2016;**68**:63–68.

Ferguson BS, Nam H, Hopkins RG, Morrison RF. Impact of reference gene selection for target gene normalization on experimental outcome using real-time qRT-PCR in adipocytes. *PLoS One* 2010;**5**:e15208.

Gangaraju V, Lin H. MicroRNAs: key regulators of stem cells. *Nat Rev Mol Cell Biol* 2009;**10**:116–125.

Goodrich R, Johnson G, Krawetz SA. The preparation of human spermatozoal RNA for clinical analysis. *Arch Androl* 2007;**53**:161–167.

Ji Z, Lu R, Mou L, Duan YG, Zhang Q, Wang Y, Gui Y, Cai Z. Expressions of miR-15a and its target gene HSPA1B in the spermatozoa of patients with varicocele. *Reproduction* 2014;**147**:693–701.

Kanakachari M, Solanke AU, Prabhakaran N, Ahmad I, Dhandapani G, Jayabalan N, Kumar PA. Evaluation of suitable reference genes for normalization of qPCR gene expression studies in brinjal (Solanum melongena L.) during fruit developmental stages. *Appl Biochem Biotechnol* 2015; **178**:433–450.

Lamba V, Ghodke-Puranik Y, Guan W, Lamba JK. Identification of suitable reference genes for hepatic microRNA quantitation. *BMC Res Notes* 2014;**7**:129.

Li Y, Li M, Liu Y, Song G, Liu N. A microarray for microRNA profiling in spermatozoa from adult men living in an environmentally polluted site. *Bull Environ Contam Toxicol* 2012;**89**:1111–1114.

Li C, Zhou X. Gene transcripts in spermatozoa: markers of male infertility. *Clin Chim Acta* 2012;**413**:1035–1038.

Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F. Altered microRNA expression in patients with non-obstructive azoospermia. *Reprod Biol Endocrinol* 2009;**7**:13.

Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;**45**:255–268.

Luo L-F, Hou C-C, Yang W-X. Small non-coding RNAs and their associated proteins in spermatogenesis. *Gene* 2015;**578**:141–157.

Marczylo EL, Amoako AA, Konje JC, Gant TW, Marczylo TH. Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? *Epigenetics* 2012;**7**:432–439.

Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, Vandesompele J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 2009;**10**:R64.

Metzler-Guillemain C, Victorero G, Lepoivre C, Bergon A, Yammine M, Perrin J, Sari-Minodier I, Boulanger N, Rihet P, Nguyen C. Sperm mRNAs and microRNAs as candidate markers for the impact of toxicants on human spermatogenesis: an application to tobacco smoking. *Syst Biol Reprod Med* 2015;**61**:139–149.

Muñoz X, Mata A, Bassas L, Larriba S. Altered miRNA signature of developing germ-cells in infertile patients relates to the severity of spermatogenic failure and persists in spermatozoa. *Sci Rep* 2015;**5**:17991.

Ohl F, Jung M, Xu C, Stephan C, Rabien A, Burkhardt M, Nitsche A, Kristiansen G, Loening SA, Radonic A *et al.* Gene expression studies in prostate cancer tissue: which reference gene should be selected for normalization? *J Mol Med (Berl)* 2005;**83**:1014–1024.

Peltier HJ, Latham GJ. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* 2008;**14**:844–852.

Pritchard CC, Cheng HH, Tewari M, Division B, Hutchinson F, Health P, Divisions S, Hutchinson F. MicroRNA profiling: approaches and considerations. *Nat Rev Genet* 2015;**13**:358–369.

Salas-Huetos A, Blanco J, Vidal F, Godo A, Grossmann M, Pons MC, F-Fernández S, Garrido N, Anton E. Spermatozoa from patients with seminal alterations exhibit a differential micro-ribonucleic acid profile. *Fertil Steril* 2015;**104**:591–601.

Salas-Huetos A, Blanco J, Vidal F, Grossmann M, Pons MC, Garrido N, Anton E. Spermatozoa from normozoospermic fertile and infertile individuals convey a distinct miRNA cargo. *Andrology* 2016;**4**:1028–1036.

Salas-Huetos A, Blanco J, Vidal F, Mercader JM, Garrido N, Anton E. New insights into the expression profile and function of micro-ribonucleic acid in human spermatozoa. *Fertil Steril* 2014;**102**:213–222.

Shenoy A, Blelloch RH. Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat Rev Mol Cell Biol* 2014;**15**:565–576.

Solayman MHM, Langaee T, Patel A, El-Wakeel L, El-Hamamsy M, Badary O, Johnson JA. Identification of suitable endogenous normalizers for qRT-PCR analysis of plasma microRNA expression in essential hypertension. *Mol Biotechnol* 2016;**58**:179–187.

Supek F, Bosnjak M, Skunca N, Smuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011;**6**:e21800.

Torres A, Torres K, Wdowiak P, Paszkowski T, Maclejewski R. Selection and validation of endogenous controls for microRNA expression studies in endometrioid endometrial cancer tissues. *Gynecol Oncol* 2013;**130**:588–594.

Van Rooij E. The art of microRNA research. *Circ Res* 2011;**108**:219–234.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:RESEARCH0034.

Wang Y, Tang N, Hui T, Wang S, Zeng X, Li H, Ma J. Identification of endogenous reference genes for RT-qPCR analysis of plasma microRNAs levels in rats with acetaminophen-induced hepatotoxicity. *J Appl Toxicol* 2013;**33**:1330–1336.

Wong L, Lee K, Russell I, Chen C. Endogenous controls for real-time quantitation of miRNA using TaqMan® microRNA assays. *Appl Biosyst Appl Note* 2007; Publication 127AP11-01.

Wotschofsky Z, Meyer HA, Jung M, Fendler A, Wagner I, Stephan C, Busch J, Erbersdobler A, Disch AC, Mollenkopf HJ *et al*. Reference genes for the relative quantification of microRNAs in renal cell carcinomas and their metastases. *Anal Biochem* 2011;**417**:233–241.

Wylie D, Shelton J, Choudhary A, Adai AT. A novel mean-centering method for normalizing microRNA expression from high-throughput RT-qPCR data. *BMC Res Notes* 2011;**4**:555.

Zhou J-H, Zhou QZ, Lyu X-M, Zhu T, Chen Z-J, Chen M-K, Xia H, Wang C-Y, Qi T, Li X *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**:28.

Zhou R, Wang R, Qin Y, Ji J, Xu M, Wu W, Chen M, Wu D, Song L, Shen H *et al*. Mitochondria-related miR-151a-5p reduces cellular ATP production by targeting CYTB in asthenozoospermia. *Sci Rep* 2015;**5**:17743.

Supplementary table I. Number of spermatozoa, total RNA amount and quality controls of each analyzed sample.

Sample code	Number of spz	Total RNA (ng)	RNA (fg)/spz.	260/280 nm ratio	*PRM1*	*GAPDH*	*CD45*	*Agilent Nano 6000*
S01	$5.78x10^7$	3,697	64	1.92	✓	✓	✓	✓
S02	$1.96x10^7$	1,229	63	1.94	✓	✓	✓	✓
S03	$1.73x10^7$	828	48	1.93	✓	✓	✓	✓
S04	$8.20x10^6$	696	85	1.84	✓	✓	✓	✓
S05	$7.45x10^6$	737	99	1.85	✓	✓	✓	✓
S06	$1.31x10^7$	712	54	1.84	✓	✓	✓	✓
S07	$1.09x10^7$	402	37	1.84	✓	✓	✓	✓
S08	$1.79x10^7$	458	26	1.97	✓	✓	✓	✓
S09	$1.14x10^7$	428	38	1.85	✓	✓	✓	✓
S10	$8.85x10^6$	373	42	1.96	✓	✓	✓	✓
S11	$2.58x10^7$	1,225	47	1.83	✓	✓	✓	✓
S12	$4.36x10^7$	477	11	1.79	✓	✓	✓	✓
S13	$7.78x10^7$	895	11	1.77	✓	✓	✓	✓
S14	$2.00x10^7$	207	10	1.62	✓	✓	✓	✓
S15	$2.62x10^7$	1,103	42	1.77	✓	✓	✓	✓
S16	$7.58x10^7$	2,066	27	1.78	✓	✓	✓	✓
S17	$6.65x10^6$	2,152	324	1.8	✓	✓	✓	✓
S18	$1.75x10^7$	5,313	304	1.83	✓	✓	✓	✓
S19	$2.95x10^6$	392	133	1.72	✓	✓	✓	✓
S20	$2.67x10^7$	9,532	357	1.8	✓	✓	✓	✓
S21	$5.39x10^7$	497	9	1.74	✓	✓	✓	✓
S22	$1.13x10^7$	765	68	1.75	✓	✓	✓	✓
S23	$6.29x10^7$	3,256	52	1.84	✓	✓	✓	✓
S24	$3.79x10^7$	3,085	81	1.8	✓	✓	✓	✓
S25	$8.89x10^7$	1,924	22	1.81	✓	✓	✓	✓
S26	$5.51x10^7$	1,923	35	1.75	✓	✓	✓	✓
S27	$1.29x10^7$	944	73	1.75	✓	✓	✓	✓
S28	$1.42x10^7$	3,165	223	1.77	✓	✓	✓	✓
S29	$2.06x10^7$	414	20	1.6	✓	✓	✓	✓
S30	$3.28x10^7$	769	23	1.8	✓	✓	✓	✓
S31	$7.00x10^6$	352	50	1.67	✓	✓	✓	✓
S32	$2.80x10^6$	463	165	1.61	✓	✓	✓	✓
S33	$8.00x10^6$	811	101	1.74	✓	✓	✓	✓
S34	$3.60x10^6$	698	194	1.64	✓	✓	✓	✓
S35	$1.05x10^7$	382	36	1.77	✓	✓	✓	✓
S36	$7.80x10^6$	446	57	1.7	✓	✓	✓	✓
S37	$8.60x10^6$	482	56	1.75	✓	✓	✓	✓
S38	$4.40x10^6$	493	112	1.36	✓	✓	✓	✓
S39	$4.60x10^6$	586	127	0.93	✓	✓	✓	✓
S40	$5.40x10^6$	1,260	350	1.43	✓	✓	✓	✓
S41	$51.0x10^6$	937	18.39	1.72	✓	✓	✓	✓
S42	$46.2x10^6$	2540	55.04	1.74	✓	✓	✓	✓
S43	$82.6x10^6$	2868	34.74	1.73	✓	✓	✓	✓
S44	$52.7x10^6$	311	5.9	1.77	✓	✓	✓	✓
S45	$35.0x10^6$	3632	103.92	1.74	✓	✓	✓	✓
S46	$54.7x10^6$	660	12.07	1.6	✓	✓	✓	✓
S47	$30.3x10^6$	2213	73.04	1.73	✓	✓	✓	✓
S48	$68.6x10^6$	1310	19.11	1.72	✓	✓	✓	✓

Supplementary table II. List of differentially expressed miRNAs (DE-miRNAs) predicted from data normalized by the different normalization strategies.

Reference method	**Evaluated methods**					
MCR	RNU6B	miR-100-5p	miR-146b-5p	miR-92a-3p	miR-30a-5p	miR-100-5p + miR-30a-5p
hsa.miR.15b.000390		hsa.miR.324.3p.002161	hsa.miR.941.002183		hsa.miR.520D.3P.002743	hsa.miR.324.3p.002161
mmu.miR.491.001630		hsa.miR.941.002183	hsa.miR.455.3p.002244		hsa.miR.941.002183	hsa.miR.15b.000390
hsa.miR.324.3p.002161		hsa.miR.455.3p.002244			hsa.miR.132..002132	hsa.miR.616.002414
hsa.miR.34b.002102		mmu.miR.491.001630			hsa.miR.192..002272	hsa.miR.941.002183
hsa.miR.942.002187		hsa.miR.616.002414			hsa.miR.1236.002761	hsa.miR.455.3p.002244
hsa.miR.28.000411		hsa.miR.636.002088			hsa.miR.380.5p.000570	hsa.miR.10b..002315
hsa.miR.520h.001170		hsa.miR.34a.000426			hsa.miR.10b..002315	mmu.miR.491.001630
hsa.miR.616.002414		hsa.miR.192..002272			hsa.miR.616.002414	hsa.miR.192..002272
hsa.miR.483.5p.002338		hsa.miR.1236.002761			hsa.miR.552.001520	hsa.miR.1236.002761
hsa.miR.455.3p.002244		hsa.miR.483.5p.002338			hsa.miR.520h.001170	hsa.miR.520D.3P.002743
hsa.miR.520D.3P.002743		hsa.miR.548c.5p.002429			hsa.miR.30d..002305	hsa.miR.636.002088
hsa.miR.770.5p.002002					hsa.miR.34a.000426	hsa.miR.212.000515
hsa.miR.34a.000426					hsa.miR.604.001567	hsa.miR.939.002182
hsa.miR.30d..002305					hsa.miR.151.5P.002642	hsa.miR.520h.001170
hsa.miR.139.5p.002289						hsa.miR.34a.000426
hsa.miR.320B.002844						hsa.miR.483.5p.002338
hsa.miR.192..002272						hsa.miR.30d..002305
hsa.miR.1236.002761						hsa.miR.604.001567
hsa.miR.941.002183						hsa.miR.552.001520
hsa.miR.212.000515						hsa.miR.509.5p.002235
hsa.miR.30e.3p.000422						hsa.miR.380.5p.000570
hsa.miR.939.002182						hsa.miR.548c.5p.002429
hsa.miR.151.5P.002642						hsa.miR.1254.002818
hsa.miR.509.5p.002235						hsa.miR.708.002341
hsa.miR.132..002132						hsa.miR.132..002132
hsa.miR.604.001567						

hsa.miR.380.5p.000570						
hsa.miR.130b..002114						
hsa.miR.636.002088						
hsa.miR.548c.5p.002429						
hsa.miR.340..002259						
hsa.miR.935.002178						
hsa.miR.30b.000602						
hsa.miR.605.001568						
hsa.miR.708.002341						
hsa.miR.552.001520						
hsa.miR.10b.002218						
hsa.miR.591.001545						
hsa.miR.30a.3p.000416						
hsa.miR.1227.002769						
hsa.miR.518e..002371						
hsa.miR.654.001611						
hsa.miR.1254.002818						

Supplementary table III. List of the target genes predicted from data normalized by the different normalization strategies.

Scan the QR code for downloading the complete table.

Supplementary IV. List of the enriched biological processes predicted from data normalized by the different normalization strategies.

Scan the QR code for downloading the complete table.

Publication 2

Sperm microRNA pairs: new perspectives in the search for male fertility biomarkers

Corral-Vazquez, C., Salas-Huetos, A., Blanco, J., Vidal, F., Sarrate, Z., Anton, E. (2019)

Fertil Steril (in press)

Quartile: Q1 (Reproductive Biology)

Impact Factor (2018): 5.411

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 means of TaqMan quantitative polymerase chain reaction (PCR) were reexamined. A set of microRNA pairs with the best biomarker potential were selected and validated by means of quantitative real-time (qRT) PCR in an independent cohort.

Setting: University laboratory.

Patient(s): Semen samples were obtained from fertile (n=10) and infertile (asthenozoospermia, n=10; teratozoospermia, n=10; oligozoospermia, n=10; unexplained male infertility [UMI], n=8) individuals. The validation cohort included 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 noncorrelated expression in each infertile group. The biomarker potential of these pairs was determined by the use of receiver operating characteristic curves. The differential relative expression of each pair in fertile and infertile populations was verified (Mann-Whitney test). Those pairs with best results were validated by qRT-PCR.

Result(s): Forty-eight pairs showed significant correlations in the fertile group. The pairs that were uncorrelated in the infertile populations and displayed the best 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 showed the greatest potential for detecting seminal alterations in the validation cohort (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.

DOI: https://doi.org/10.1016/j.fertnstert.2019.07.006

The content of this research article has been omitted due to privacy politics of the journal.

Publication 3

RNA espermático: ¿huella de eventos pasados o dote para el embrión?

Corral-Vazquez, C., Blanco, J., Vidal, F., Anton, E., (2017)

Med Reprod y Embriol Clin 4, 59–71

Abstract

The biological relevance of sperm contribution to the embryo has currently been shown to go beyond a mere transmission of the paternal genome. New findings have revealed that spermatozoa carry a complex and multifunctional population of RNA with a wide range of implications in spermatogenesis, fertilisation, and embryogenesis. Accordingly, the conside- ration of these molecules as a simple residual pool of earlier processes has been left behind. Besides the coding RNAs, the role of non-coding RNAs has started to be untangled and integra- ted within the global cell processes, demonstrating their relevance in cellular regulation and epigenetics. Comparative high-throughput expression studies have provided a general impres- sion on the differences in the sperm transcriptome between populations. Hence, a new diagnostic value related to these RNAs has emerged. In this review, an overview of biosynt-hesis, functions and biological mechanisms of different subtypes of sperm RNAs is presented. The relevance of these molecules during spermatogenesis, their role after fertilisation, and their potential value as fertility biomarkers are also discussed.

DOI: http://dx.doi.org/10.1016/j.medre.2017.03.002

The content of this literature review has been omitted due to privacy politics of the journal.

Publication 4

Sperm ncRNA

Corral-Vazquez, C., Anton, E., (2018)

In: Camprubí, C., Blanco, J. (Eds.), **Epigenetics and Assisted Reproduction: An Introductory Guide**. *CRC Press*, Boca Raton

The content of this book chapter has been omitted due to privacy politics of the editors.

4 Discussion

4.1. Methodological considerations for sperm transcriptomic profiling

4.1.1. Sperm-associated limitations

RNA extraction for transcriptomic purposes is a practice that entails several limitations associated to the fragile nature of this kind of samples. Among them, the propensity to degradation is highlighted, due to the frequent presence of ribonucleases in RNA samples, both endogenous and external (i.e. contamination from the experimental material or processing). This risk can be partially avoided by taking precautions during the extraction procedure, like changing gloves frequently, using RNase-free material and thoroughly cleaning the working area (Peimbert and Alcazar, 2016). Moreover, there is a basal risk of DNA contamination derived from the experimental procedure. The origin can be external laboratory material (which can be avoided by taking the above mentioned precautions), and also endogenous cross-contamination during the RNA separation, which should be prevented by using a reliable extraction method (WG. Guder, S. Narayanan, H. Wisser, 2009).

Apart from those handicaps, the application of RNA extraction procedures in ejaculated spermatozoa has several specific limitations. Therefore, the implementation of an appropriate methodology is required to ensure unbiased and reliable results. Some of these limitations are related to the low RNA quantity contained in spermatozoa in comparison with other cells. In consequence, if the protocols of sperm cell isolation do not work properly, there is a risk of masking the results by non-sperm cells. Moreover, a low RNA recovery rate increases the possibility of deleterious effects produced by the presence of remnants of organic reagents (used in the extraction methodology) that will hinder the subsequent RNA analysis.

4.1.1.1. Sperm cell isolation

Semen samples may contain a fraction of non-sperm cells including leukocytes, epithelial cells, and a certain amount of germ cells (Fedder, 1996). Therefore, an initial sperm isolation step is essential to remove this

material in order to start the RNA extraction protocol with a purified fraction of spermatozoa.

The Somatic Cell Lysis Buffer (SCLB) method, adapted from Goodrich et al. (2007), was implemented in previous sperm transcriptomic studies performed by our research group (Salas-Huetos et al., 2016, 2015, 2014), and has been applied with satisfactory results in several studies (Abhari et al., 2014; Bansal et al., 2015; Goodrich et al., 2007; Jodar et al., 2012; Lalancette et al., 2009; Pacheco et al., 2012). It is based on the lysis of the plasma membrane of non-sperm cells, caused by its exposure to a mixture of detergents. Sperm cells present a higher resistance to detergents compared to non-sperm cells due to their high concentration of lipid rafts, also known as Detergent Resistant Membranes (DRMs). These domains of the plasma membrane are enriched in proteins, cholesterol and sphingolipids, and are highly resistant to solubilization by non-ionic detergents (Nixon et al., 2011).

The principal strength of SCLB, in comparison to other sperm isolation methods, is that the obtained cell fraction maintains the original proportion of sperm cells. That is, no sperm selection is performed based on their characteristics. Therefore, this method enables the analysis of the whole sperm fraction and provides deep and full profiling of the sample. Other methods, like PureSperm or Swim-Up techniques, are based on the selective isolation of the high-motility fraction or the morphologically normal sperm cells. Although these strategies have displayed a high maintenance of RNA integrity in comparison to SCLB (Mao et al., 2013), their application in transcriptomic profiling studies implies the loss of a wide fraction of spermatozoa (more than 50%) with low motility and/or abnormal morphology (Jayaraman et al., 2012; Malvezzi et al., 2014). In this sense, sperm selection is mandatory when the cells are destined to assisted reproduction techniques, since it enables a selection of optimal spermatozoa. Nevertheless, when aiming to characterize the sperm transcriptome profile of infertile males in association with their altered seminal parameters (e.g. including abnormal morphology and low motility), the RNA cargo of the sperm that were discarded by these methods would be highly informative. Besides, sperm loss implies a lowering of the already scarce RNA recovery rate.

Independently of the pursued methodology of sperm cell isolation, it should be taken into account that the exclusion of 100% non-sperm cells cannot be achieved. Therefore, to ensure the effectiveness of SCLB, a microscopic assessment of the non-sperm cell ratio over sperm fraction was included in the RNA extraction protocol of all samples. In this step, a maximum threshold of 1:10,000 non-sperm cells versus spermatozoa was allowed. Since somatic cells have an RNA content of 100-1,000 times fold higher than spermatozoa, the use of this ratio ensured that more than 99% of the RNA molecules obtained after the application of RNA extraction methods belong to sperm cells.

Moreover, additional quality controls were included in our experimental design to assess the correctness of the results of the SCLB. These controls included several RT-PCR analyses of marker genes specifically expressed in different types of non-sperm cells (which are potentially present in semen samples). One of them was Protein Tyrosine Phosphatase Receptor Type C (*PTPRC* or *CD45*), which was aimed to detect possible leukocyte contamination since it encodes a transmembrane glycoprotein expressed only in nucleated hematopoietic cells (Donovan and Koretzky, 1993). The amplification of the proto-oncogene receptor *KIT* was selected as a marker of pre-meiotic germ cells (Von Schönfeldt et al., 2004), and the gene encoding the cell-cell adhesion glycoprotein Cadherin-1 *(CDH1*) was used as an epithelial marker (Van Roy and Berx, 2008). As an additional check to corroborate the absence of non-sperm RNA, and taking advantage of the characteristic fragmented status of sperm rRNA, the lack of signal associated to non-degraded 18S and 28S ribosomal subunits was verified in all samples by nanoelectrophoresis.

Regarding the works developed in this thesis, all the studied samples presented a ratio of non-sperm cells under 1:10,000, displayed no RT-PCR amplification of *CD45*, *KIT* and *GAPDH*, and showed no signal corresponding to 18S or 28S subunits. These results corroborated the suitability of the used methodology and enabled further analyses.

4.1.1.2. Sperm RNA extraction methods

As mentioned above, the main handicap of sperm transcriptome isolation is the scarce amount of RNA that is generally yielded, due to the naturally

low RNA content of sperm cells (10-100 fg) (Krawetz, 2005). Besides, the purity of the obtained samples must be high in order to ensure the absence of possible remnants of DNA or organic substances that would eventually difficult further steps of the protocol. Therefore, a robust RNA extraction methodology needs to be implemented in order to reach an optimal RNA yield and purity that will enable its further characterization.

RNA extraction strategies can be categorized in three groups depending on their isolation mechanism and used resources (Vomelová et al., 2009): 1) Methods based on the different solubility of the homogenized cell sample components in organic solvents; 2) Methods relying on RNA adsorption to specific surfaces in the presence of chaotropic salts; and 3) Methods based on RNA separation on isopycnic gradient centrifugation. Along the development of this thesis, three RNA extraction methodologies were employed (see Appendix II for detailed protocols): one of them belonging to the first group (i.e. TRIzol), another one to the second group (i.e. Maxwell kit), and the third one being a mixture of group 1 and 2 (i.e. mirVana kit).

TRIzol is commonly considered as the golden standard for total RNA isolation. It allows the recovery of both large and small RNAs, maintaining high transcript stability (Mraz et al., 2009; Rio et al., 2010). This method was employed for the RNA extraction of samples S1-S48 and V1-V23 (Publications 1 and 2). It allowed good quality results concerning RNA yield, with a mean of 64.80±79.49 fg/spermatozoa, which is coherent with the theoretical RNA content of sperm cells. Nevertheless, methods based on organic solvents imply a risk of contamination with TRIzol reagent during the phase-separation step or with ethanol during the washing process. This is an especially important handicap when it concerns low quantities of RNA, since any organic remnant can significantly lower the purity of the sample. In fact, although the achieved mean range of 260/280 nm ratio was 1.74±0.17 (which is close to the 1.8 threshold), the minimum value of the purity range was 0.93. Although TRIzol has proved to be a suitable method for qPCR studies, organic contamination could be a problem for an RNA-seq protocol (Sultan et al., 2014).

Therefore, alternative RNA extraction strategies were settled for the RNA-seq experiments of Appendix I. Firstly, mirVana kit was employed in samples P1-P10 as an intermediate option between organic solvent

extraction and surface adsorption isolation method. This protocol aimed to reach the high final RNA yield provided by phenol-chloroform phase separation, but also reducing residual organic contaminations with the attachment of the RNA to the solid surface of the column filter. Additionally, centrifugation steps facilitate the elimination of the washing buffer, avoiding final ethanol remnants. The resulting ultimate RNA quantity was lower than the samples extracted by TRIzol (mean yield of 26.64±54.53 fg per cell). Regarding the achieved purity, although both the minimum and the mean 260/280 nm ratio (1.43 and 1.80±0.26 respectively) was higher in the samples obtained by mirVana in comparison with TRIzol, there was no statistical difference between both groups (data not shown). Therefore, although a possible reduction of organic or DNA contamination might be achieved by using mirVana in comparison with TRIzol, there is no statistical evidence of this fact.

Finally, a surface-adsorption method based on the use of paramagnetic beads (i.e. Maxwell kit) was selected for the samples P11-P12 of Appendix I. The main advantage offered by this system is its automation, which reduces possible variations related to sample manipulation throughout the experimental protocol. However, lower RNA yields were achieved (4.22±1.10 fg per cell) in comparison with TRIzol and MirVana. Regarding purity, a high 260/280 nm ratio was achieved in the samples obtained by Maxwell (1.80±0.07), similarly to the outcomes of MirVana and higher than the scores provided by TRIzol. However, it should be taken into account that only two RNA samples were extracted by Maxwell, so a higher number of cases would be necessary for establishing a statistical comparative. Even so, sufficient quantities of total RNA with an appropriate quality were achieved by Maxwell.

Therefore, although TRIzol allowed the highest RNA yield, a slightly higher purity was reached by the use of both MirVana and Maxwell kit, which makes them more appropriate for RNA-seq studies. The possible influence of using two different extraction methodologies in the samples analyzed by RNA-seq was assessed by the Principal Component Analysis (PCA) and heatmap clustering performed in Appendix I. As observed in these analyses, no sample segregation associated to the used extraction methodology was produced, agreeing with previous results obtained by other authors (Mao et al., 2014). Therefore, this indicated that no

significant bias in the expression profile of the samples was introduced by the different kits, ensuring solid and homogeneous results.

4.1.1.3. Verifying the lack of DNA contamination and transcript integrity

As mentioned above, a possible cause of a drop in RNA purity could be the presence of DNA. In order to discard this potential contamination, qRT-PCRs of two genes (*PRM1* and *GAPDH*) were performed using primers that were designed following an exon-exon strategy. This method ensures discrimination per size between genomic amplicons (with introns) and cDNA amplicons (no introns). The exon-exon strategy is commonly followed for quality controls in sperm RNA studies (Goodrich et al., 2007). All the analyzed samples showed a *PRM1* amplification of 331 bp and a *GAPDH* amplification of 228 bp, which corresponds to the cDNA amplicon size, while none of the samples showed the 422 bp and 332 bp bands corresponding to genomic contamination.

At last, verifying the integrity of relevant transcripts is an important step for obtaining conclusive expression results. Routinely, RNA integrity controls rely on estimating the RNA Integrity Number (RIN) (Schroeder et al., 2006). Nevertheless, this value cannot be calculated in sperm cells since it is based on the presence of intact rRNA 18S and 28S transcripts and, as stated before, no functional ribosomes can be found in sperm cells. Therefore, the amplification of mRNA transcripts that are known to be present and preserved in sperm (i.e. *PRM1* and *GAPDH*) was also intended to verify RNA integrity. The amplification of both genes in all the analyzed samples confirmed the presence of full-length transcripts.

In summary, the absence of DNA in all the samples, together with the isolation of RNA transcripts with proper integrity, indicate that the methodology used for RNA extraction was suitable for achieving good quality samples.

4.1.2. Data normalization

Normalization of raw RNA expression data is essential for removing variations caused by technical deviations in sample preparation and

processing, thus allowing comparisons between different experiments or samples. In these cases, adapted normalizing strategies need to be implemented depending on the employed technology. Therefore, several correction approaches were selected in this thesis according to the used methodologies: MCR for miRNA qRT-PCR arrays, endogenous reference miRNAs for singleplex miRNA qRT-PCR assays, and FPKM for RNA-seq.

When correcting raw expression results from qRT-PCR high-throughput studies, mean-based methods suppose a reliable option. This is the case of the dataset used in Publications 1 and 2, in which MCR method provided the robustness of a mean-centering normalization, which took advantage of the high amount of used values. Additionally, this method offered an additional specificity factor since it was specially designed for miRNA TaqMan Megaplex qRT-PCR experiments (Wylie et al., 2011).

When working with a low set of qRT-PCR reactions, reference normalizing molecules need to be incorporated to the experimental design. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines related to qRT-PCR experiments advised against the use of standard pre-established reference molecules without validation due to the high risk of misleading effects (Bustin et al., 2009). Nevertheless, these suggestions are frequently disregarded in scientific production (Gutierrez et al., 2008). In human sperm miRNA expression studies, U6 is one of the most recurrently used reference gene (Abu-Halima et al., 2014b, 2013; Y. Li et al., 2012; Zhou et al., 2015). As it was observed in Publication 1, U6 expression showed high variability between samples, and data correction by this endogenous control caused important alterations in the outcomes when compared to MCR normalization. Therefore, our findings constituted an advice against the use or this snRNA for normalizing miRNA quantification data. A similar conclusion has been reached by other authors working in different cell types (Benz et al., 2013; Das et al., 2016; Lamba et al., 2014; Wang et al., 2013; Wotschofsky et al., 2011). Accordingly, a systematic selection and validation of normalizers for particular tissues, cell types, and experimental designs should be mandatory in expression studies that lack previously implemented normalizers (Bustin et al., 2009). In this sense, the main purpose of Publication 1 was to provide efficiently tested reference controls not only for our sperm miRNA expression studies, but also for

any study based on sperm miRNA singleplex qRT-PCR. To find out the best normalizing candidates, the study aimed to identify miRNAs that extrapolated MCR effectiveness, thus providing a similar quality in the final outcomes. Our results proved that the mean expression value of hsa-miR-100-5p and hsa-miR-30a-5p maintained a high resemblance to MCR normalization along the different steps of the bioinformatics analysis pipeline (identification of differentially-expressed miRNAs, target genes prediction, and GO analyses). The results of this study were successfully employed for data normalization in Publication 2, and also in other published studies (Riesco et al., 2019; Salas-Huetos et al., 2018).

Similarly to qRT-PCR studies, data from RNA-seq analyses performed in Appendix I also required normalization by a specific method. The raw alignment of RNA-seq reads to the genome sequence presents biases caused by several aspects. On the one hand, longer gene sequences accumulate a higher number of reads in comparison with shorter ones, even when there is not a real difference in their expression levels. Additionally, a different number of total reads (library size) could be produced by different samples, impeding their comparison. In this work, FPKM normalization was used to remove effects produced by these differences in gene length and library sizes, allowing gene-gene and sample-sample expression comparisons.

4.2. Human sperm RNA profile

The results derived from this thesis revealed a wide population of coding and non-coding sperm transcripts. Besides the small quantity of RNA contained in these cells, the fragmentation tendency of the sperm transcriptome, as described by other authors (Boerke et al., 2007; Jodar et al., 2013), was evidenced by the low TIN score obtained from the samples analyzed in Appendix I. In spite of this degradation, half of the identified mRNAs and lncRNAs were categorized as *Highly Expressed*. Additionally, hosts of ubiquitous and stable large transcripts were identified, just as other authors have also described in previously published studies (G. C. Ostermeier et al., 2005; Ostermeier et al., 2002). Besides, the identified sets of mRNAs described in Appendix I show a certain similarity with sperm RNA-seq datasets previously obtained by other authors (Sendler et

al., 2013), indicating inter-individual preservation of sperm transcripts. Altogether, these results reinforce the existence of a certain selective preservation of transcripts in spite of overall fragmentation of sperm RNA, and suggest a functionality of these transcripts that may be relevant for sperm function.

4.2.1. Biological functions of sperm mRNA

Some authors had defended that sperm RNA were randomly retained transcripts from the residual cytoplasmic droplets that failed in being absorbed by the Sertoli cells (Miller et al., 2005). Nevertheless, other findings have suggested that a set of functional transcripts are specifically retained within mature spermatozoa. Among these findings, the discovery of RNA stored in the nucleus of mature sperm (Pessot et al., 1989; Wykes et al., 1997) was a key to consider potential roles of these transcripts during fertilization and male pronucleus formation. In fact, the presence of stable sperm transcripts in the zygote before ZGA was later identified (Ostermeier et al., 2004).(Ostermeier et al., 2004b). Diverse potential functions in early embryo stages were attributed to several of these sperm transcripts, mainly in cell signaling, morphogenesis, stress response, implantation, and transcription modulation (Boerke et al., 2007). Among this whole set of sperm transcripts, several studies have revealed the presence of a consistent fingerprint of common transcripts that are maintained in sperm samples of different individuals, suggesting that they hold a specific functionality. This involvement reinforces the idea of selective preservation of necessary transcripts for previous and further events (G. C. Ostermeier et al., 2005; Ostermeier et al., 2002).

The results of this thesis also support this new perspective of sperm transcriptome. The identified 13,126 *Expressed* mRNAs in Appendix I showed a major involvement in several fertility-related processes. Most of the enriched functions revealed by GO analyses showed a deep implication in spermatogenesis pathways, including mitosis and cell cycle regulation. It has been suggested that mRNAs involved in these functions would have probably been transcribed during the transcriptional waves of spermatogenesis and stored for further stages (Krawetz, 2005). Additionally, *Expressed* mRNAs also displayed an involvement in processes

related to reproduction, morphogenesis, and cellular response to stress. Concurring with the role described by other authors, these functions could be related with a post-fertilization activity. In fact, some of the transcripts detected in the analyzed samples have been previously related to early embryo development: *PLC-Z1* (which is produced during late spermatogenesis and directly contributes to oocyte activation) (Saunders et al., 2007), *APAK4* (essential for sperm motility and potentially involved in oocyte activation) (Miki et al., 2002), *WNT5A* (related to morphometric patterning) (T.Moon et al., 1997), and *STAT4* (with a potential regulatory role in the male pronucleus) (Herrada and Wolgemuth, 1997).

Furthermore, among the whole set of *Expressed* genes, our results revealed a host of common (*Ubiquitously Expressed*) and full-length conserved (*Highly Stable*) mRNAs throughout all analyzed individuals. This is coherent with the previous identification of common fingerprints of transcripts among different fertile individuals, indicating a selective preservation of these molecules. Indeed, our results showed that the functional characterization of these transcripts was involved in cell cycle regulation, spermatogenesis and reproduction. Many of the detected transcripts have been previously found to be highly expressed in testis tissue (Fagerberg et al., 2014), and involved in different aspects of spermatogenesis and fertility, like *PRM2*, *TNP1*, *SPATA18*, *SPATA3,* Testis Specific Serine Kinase 6 (*TSSK6*), and Pyruvate Kinase M1/2 (*PKM*) (Jodar et al., 2013; Xu et al., 2008).

4.2.2. Biological functions of sperm non-coding RNAs

Although the role of sperm ncRNAs is a recent research field, increasing studies keep revealing an important involvement of these transcripts in the regulation of spermatogenesis and reproduction (An et al., 2019; Fullston et al., 2016; Gapp et al., 2014; Kotaja, 2014; Salas-Huetos et al., 2014; Yuan et al., 2016; Zhang et al., 2015; X. Zhang et al., 2019). Among them, the participation of sperm lncRNA is a matter of research that still needs to be deepened.

The regulatory role of lncRNAs is based on the modulation of gene expression in *cis* (regulation of close genomic sequences) or *trans* (regulation of distant sequences). Although they have not been fully

characterized yet, it is known that *cis* regulation can produce gene inactivation by different interactions with promoter sequences, transcription factors, DNA-binding proteins or epigenetic modifiers (Ponting et al., 2009). Alternatively to gene silencing, they can also perform a gene activation by affecting chromatin remodeling and interacting with specific transcription factors or activator proteins (Ponting et al., 2009).

A recent RNA-seq study has found an implication of human sperm lncRNAs in the regulation of sperm functionality and spermatogenesis (X. Zhang et al., 2019). Our results about the predicted target genes of sperm lncRNAs presented in Appendix I agree with this statement. We identified 7,521 *Expressed* lncRNAs that were located at a short distance (<10 kb) of several coding genes. Although the associated GO functionality of these genes was quite diverse, it included a strong relation with cell-to-cell adhesion processes and embryogenesis. Furthermore, when the target genes were divided into *Present* (genes that belong to the identified sperm *Expressed* mRNAs) and *Absent* (genes that are not included among the sperm *Expressed* mRNAs), GO analyses showed clear differences in their functionality: while *Absent* targets presented an exclusive enrichment of embryo-related and cell adhesion GO processes, the analysis of the *Present* targets showed an overall representation of cytoskeleton organization functions, suggesting an implication in cell division processes.

This differential functionality might suggest that, while *Present* target genes could have been regulated in spermatozoa by lncRNAs during spermatogenesis, the regulation of *Absent* genes could take place in further embryonic stages. That is, some sperm lncRNAs could be involved in the regulation of genes implicated on the first stages of zygote activation, between sperm RNA release to the oocyte and parental RNA degradation produced at ZGA. For this reason, sperm lncRNAs would be preserved to perform selective gene activation in the embryo. Although this role of sperm lncRNAs has not been previously described, some oocyte lncRNAs have been found to be involved in the regulation of early embryo stages, like Neat AK124742 and RNAsGtl2 (involved in the regulation of oocyte maturation, early embryo development, and ZGA) (Liu et al., 2018). Therefore, it may be suggested that this embryo

regulation by maternal lncRNAs could occur similarly with paternally originated transcripts. Indeed, the regulatory influence of paternal ncRNA over embryo gene expression has been proved in mice by several authors. Specifically, some studies have found that obese mice presented altered sperm miRNA profiles that could have been transmitted to the offspring, which displayed a similar metabolic phenotype (Fullston et al., 2016; Gapp et al., 2014). Recently, a relation between sperm lncRNAs and epigenetic inheritance of obesity in mice has also been discovered (An et al., 2019). Therefore, although the results of our study should be cautiously interpreted and would still need validation, they open the possibility of new roles of paternal lncRNA after fertilization.

Similarly to sperm lncRNAs, several studies have described an involvement of sperm miRNAs in spermatogenesis and embryogenesis. Among them, the study on which Publications 1 and 2 are partially based (Salas-Huetos et al., 2014) identified a set of ubiquitous sperm miRNAs that regulated the expression of genes involved in cell differentiation, development, morphogenesis, and embryogenesis. These results agreed with the previously described importance that paternal miRNAs exert in spermatogenesis (Kotaja, 2014), and early embryo stages (e.g. miR-34c has been described to be crucial for first embryonic cleavage divisions in mice) (Liu et al., 2012). The implication of paternally originated miRNAs in embryogenesis was further corroborated by embryo monitoring both in presence and absence of sperm miRNAs, showing zygote development failing in *Drosha* knock-out mice (Yuan et al., 2016).

4.3. Sperm RNAs as fertility biomarkers

Results from Publication 2 and Appendix I contribute to the search for novel methods in which expression levels of certain sperm RNAs could be indicative of the male fertility status. This kind of approach could represent an additional tool to complement the current diagnosis methods of male infertility from a molecular insight. These assessments would be especially informative when facing cases in which seminogram is inconclusive. A clear example would be those individuals whose seminal parameters are close to the threshold values that delimitate fertility (World Health Organization, 2010). It has been observed that seminal

parameters are highly variable between different samples of the same patient (Keel, 2006), so minimal changes in seminal measurements of these individuals could be critical for the final diagnosis. Also, this type of molecular approach could be a crucial tool for tracing cases of UMI given the fact that current methods do not provide a reliable diagnosis.

4.3.1. A common fingerprint of transcripts

Ubiquitous sets of mRNAs and lncRNAs have been described in the fertile populations analyzed in Appendix I. Specifically, a total of 784 mRNAs and 116 lncRNAs were identified among all the studied samples. Moreover, our group described in a previous work a set of miRNAs (221) that are also ubiquitously expressed in human spermatozoa (Salas-Huetos et al., 2014). As discussed above, the constitutive presence of transcripts in all studied sperm samples must be indicative of their importance for human fertility. In addition, they also display the property of raising the probability of being detected in any further human sperm transcriptome study. Therefore, their expression levels could be used as a reference value for comparative studies, in order to detect transcripts with altered expression associated to fertility disorders.

In fact, among the 784 ubiquitous mRNAs, an altered expression of some of them has been previously described in different infertility populations (Table 7). For example, *PRM1* and *PRM2* alterations have been related to teratozoospermia and asthenozoospermia (Aoki et al., 2006; Aslani et al., 2011; Avendaño et al., 2009; Kempisty et al., 2007; Lambard et al., 2004; Rogenhofer et al., 2013; Savadi-Shiraz et al., 2015; Zalata et al., 2016); *ODF1* and *TNP1* abnormal levels have been associated to teratozoospermia (Platts et al., 2007; Savadi-Shiraz et al., 2015); and altered *HSPA2* expression has been described in oligozoospermia (Cedenho et al., 2006; Motiei et al., 2013). Regarding sperm miRNAs, several of the 221 identified ubiquitous transcripts have been previously found to display a differential expression related to different seminal anomalies (asthenozoospermia, teratozoospermia, oligozoospermia and UMI) (Salas-Huetos et al., 2016, 2015). With respect to the 116 ubiquitous lncRNAs identified in Appendix I, no previous publications have compared the expression of any of these molecules in fertile and infertile cohorts.

Even so, together with the sets of mRNAs and miRNAs, they could be taken as a reference point in further comparative studies, serving to identify new transcripts with biomarker potential.

4.3.2. Biomarker pairs of transcripts

Alternatively to the use of single-molecule methods as biomarkers for infertility diagnosis, a novel molecular strategy based on evaluating the relative expression of transcript pairs was developed in Publication 2 and Appendix I. This approach was based on the fact that fertile samples contain several transcripts with a pairwise stable and correlated expression that is not maintained in specific infertility profiles. Therefore, a relative expression between molecules is considered instead than a single expression reference value. In this sense, this kind of strategy represents a more sensitive approach, since a variation in the expression of any of the two molecules that compose each pair (or changes in both molecules in an uncorrelated way) will be reflected in the resulting value.

In Publication 2, the comparison between profiles of fertile and infertile populations allowed the selection of two miRNA pairs as potential biomarkers: hsa-miR-942-5p/hsa-miR-1208, and hsa-miR-34b-3p/hsa-miR-93-3p. These molecules displayed a ubiquitous and stable correlation in the fertile population. In the validation stage included in this study, the hsa-miR-942-5p/hsa-miR-1208 pair stood out due to its high potential for discerning infertile patients with alterations in sperm motility, morphology and count. Therefore, this pair was postulated as the best biomarker candidate for detecting infertile individuals with any seminal parameter altered. In fact, the versatility of this biomarker is a convenient characteristic since more than one type of sperm parameters tend to be simultaneously affected in infertile males with abnormal seminogram (e.g. the oligo-astheno-teratozoospermia syndrome has a frequency of ~26% of male infertility cases) (Dohle et al., 2005). Nevertheless, the fact that no UMI patients could be included in the validation cohort supposes a meaningful limitation of the study that must be taken into account. The main reason why no UMI individuals could be included in the validation cohort relies on the low incidence of these cases among infertile patients. Besides that, there is also an intrinsic difficulty along the selection of this

kind of patients in which any other possible cause of infertility (e.g. infertility originated by female factor) must be discarded. In spite of this handicap, it is important to note that one of the four selected miRNA pairs (i.e. hsa-miR-34b-3p/hsa-miR-93-3p) displayed a promising potential for the diagnosis of this specific infertile profile in the initial study population. The validation of this pair in UMI patients supposes a promising premise for future studies.

Regarding sperm mRNAs, some other authors have already described several potential biomarker pairs that would allow detecting diverse types of infertility disorders (see section 1.4.3. Stable pairs of transcripts as infertility biomarkers). In fact, some of the *Ubiquitously Expressed* mRNAs identified in Appendix I have been previously found by other authors to have a stable correlated expression with other transcripts. This correlation became altered in infertility patients such is the case of DFS (Burl et al., 2018; Lima-Souza et al., 2012). In our study, among the whole set of *Ubiquitously Expressed* mRNAs, four pairs of transcripts were observed to display a correlated expression: i) Nuclear Mitotic Apparatus Protein 1 (*NUMA1*), and Chromodomain Helicase DNA Binding Protein 2 (*CHD2*); ii) Cerebellin 1 Precursor (*CBLN1*), and Required for Meiotic Nuclear Division 5 Homolog B (*RMND5B*); iii) Tousled Like Kinase 2 (*TLK2*), and Neurofibromin 1 (*NF1*); and iv) Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1*), and Mitochondrially Encoded Cytochrome C Oxidase II (*MT-CO2*). These pairs have not been identified as potential biomarkers by any other author yet, so the validation of their suitability as biomarker pairs still requires future studies. In the same way, 22 pairs of lncRNAs showed a stable correlated expression constituting the first pairwise-correlated fingerprint of lncRNAs characterized in a fertile population, which provides a novel perspective for ncRNA biomarker studies.

The discussed set of stable pairs of transcripts lays the groundwork for new and powerful alternatives in the field of infertility biomarker research, beyond the recurrent and widely explored single-molecule methods. This provides novel perspectives for a future implementing of molecular approaches as additional tools for reproductive medicine, enabling an improvement in clinical counseling and diagnosis.

5 Conclusions

1. The optimized methodologies related to the analysis of the human sperm transcriptome have allowed obtaining sperm RNA samples with appropriate quality for qRT-PCR and RNA-seq studies.

2. The combination of hsa-miR-100-5p and hsa-miR-30a-5p displays a high normalizer potential for sperm miRNA singleplex qRT-PCR studies, since they are ubiquitously and stably expressed, share the same biological nature than the analyzed molecules, and the results derived from data normalized by their mean expression value are equivalent to those derived from the mean-centering normalization method.

3. Spermatozoa from fertile men present a complex network of mRNAs and lncRNAs with a high fragmentation status. Among the whole transcriptome cargo, a homogeneous subset of highly expressed RNAs is found, as well as transcripts with ubiquitous expression, and RNAs with high integrity.

4. The set of sperm mRNAs from fertile men is ontologically associated with biological functions related with spermatogenesis. The relation with these processes is reinforced in the ubiquitous and most stable subset of mRNAs, which indicates their very specific role.

5. The sperm lncRNAs from fertile men are predicted to regulate in *cis* a set of targets genes that are ontologically related to cell adhesion and embryo development processes. The significance of this association is stronger in those *cis*-target genes that are not present among the sperm mRNA cargo.

6. A novel biomarker strategy based on the stable and correlated expression of two transcripts has been developed. Specifically, the sperm miRNA pairs hsa-miR-942-5p/hsa-miR-1208 allow discerning infertility cases with seminal alterations, while the pair hsa-miR-34b-3p/hsa-miR-93-3p is suitable for the detection of UMI patients. Moreover, several pairs of sperm mRNAs and lncRNAs that display a stable correlated expression in fertile individuals have also been observed, supposing a starting point for further clinical research regarding the biomarker potential of these molecules.

6 References

A

Aarabi, M., Ousati-Ashtiani, Z., Nazarian, A., Modarressi, M., Heidari, M., 2008. Association of TGIFLX/Y mRNA expression with Azoospermia in infertile men. Mol. Reprod. Dev. 75, 1761–1766.

Abhari, A., Zarghami, N., Shahnazi, V., Barzegar, A., Farzadi, L., Karami, H., Vahed, S.Z., Nouri, M., 2014. Significance of microRNA targeted estrogen receptor in male fertility. Iran. J. Basic Med. Sci. 17, 81–86.

Abu-Halima, M., Backes, C., Leidinger, P., Keller, A., Lubbad, A.M., Hammadeh, M., Meese, E., 2014a. MicroRNA expression profiles in human testicular tissues of infertile men with different histopathologic patterns. Fertil. Steril. 101, 78-86.e2.

Abu-Halima, M., Hammadeh, M., Backes, C., Fischer, U., Leidinger, P., Lubbad, A.M., Keller, A., Meese, E., 2014b. Panel of five microRNAs as potential biomarkers for the diagnosis and assessment of male infertility. Fertil. Steril. 102, 989–997.

Abu-Halima, M., Hammadeh, M., Schmitt, J., Leidinger, P., Keller, A., Meese, E., Backes, C., 2013. Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. Fertil. Steril. 99, 1249–1255.

Agarwal, A., Mulgund, A., Hamada, A., Chyatte, M.R., 2015. A unique view on male infertility around the globe. Reprod. Biol. Endocrinol. 13, 1–9.

Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U. S. A. 74, 5350–5354.

Amanai, M., Brahmajosyula, M., Perry, A., 2006. A restricted role for sperm-borne microRNAs in mammalian fertilization. Biol Reprod 75, 877–884.

American College of Physicians, 2007. MKSAP 14 : Medical Knowledge Self-Assessment Program. Philadelphia, USA.

An, T., Zhang, T., Teng, F., Zuo, J.-C., Pan, Y.-Y., Liu, Y.-F., Miao, J.-N., Gu, Y.-J., Yu, N., Zhao, D.-D., Mo, F.-F., Gao, S.-H., Jiang, G., 2019. Long non-coding RNAs could act as vectors for paternal heredity of high fat diet-induced obesity. Oncotarget 10, 47876–47889.

Andersen, C.L., Jensen, J.L., Orntoft, T.F., 2004. Normalization of real-time

quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res 64, 5245–5250.

Anton, E., Krawetz, S., 2012. Spermatozoa as biomarkers for the assessment of human male infertility and genotoxicity. Syst. Biol. Reprod. Med. 58, 41–50.

Aoki, V., Liu, L., Carrell, D., 2006. A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. Mol. Hum. Reprod. 12, 41–50.

Aslani, F., Modarresi, M.H., Soltanghoraee, H., Akhondi, M.M., Shabani, A., Lakpour, N., Sadeghi, M.R., 2011. Seminal molecular markers as a non-invasive diagnostic tool for the evaluation of spermatogenesis in non-obstructive azoospermia. Syst. Biol. Reprod. Med. 57, 190–6.

Aston, K.I., Carrell, D.T., 2009. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. J. Androl. 30, 711–725.

Aston, K.I., Krausz, C., Laface, I., Ruiz-Castané, E., Carrell, D.T., 2010. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. Hum. Reprod. 25, 1383–1397.

Avendaño, C., Franchi, A., Jones, E., Oehninger, S., 2009. Pregnancy-specific β-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: Differential expression in fertile and infertile men and evidence of a possible functional role during early development. Hum. Reprod. 24, 270–277.

B

Bansal, S.K., Gupta, N., Sankhwar, S.N., Rajender, S., 2015. Differential genes expression between fertile and infertile spermatozoa revealed by transcriptome analysis. PLoS One 10, 1–21.

Bao, J., Wu, J., Schuster, A.S., Hennig, G.W., Yan, W., 2013. Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. Biol. Reprod. 89, 1–12.

Barber, R.D., Harmer, D.W., Coleman, R.A., Clark, B.J., 2005. GAPDH as a

housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. Physiol. Genomics 21, 389–395.

Barbu, V., Dautry, F., 1989. Northern blot normalization with a 28S rRNA oligonucleotide probe. Nucleic Acids Res. 17, 7115.

Bartlett, J.M.S., Stirling, D., Oien, K.A., 2003. Serial Analysis of Gene Expression. PCR Protoc. 270, 271–284.

Becker-André, M., Hahlbrock, K., 1989. Nucleic Acids Research Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATITY) Nucleic Acids Research Olikonucleotides. Nucleic Acids Res. 17, 9437–9446.

Benchaib, M., Braun, V., Ressnikof, D., Lornage, J., Durand, P., Niveleau, A., Guérin, J.F., 2005. Influence of global sperm DNA methylation on IVF results. Hum. Reprod. 20, 768–773.

Benz, F., Roderburg, C., Vargas Cardenas, D., Vucur, M., Gautheron, J., Koch, A., Zimmermann, H., Janssen, J., Nieuwenhuijsen, L., Luedde, M., Frey, N., Tacke, F., Trautwein, C., Luedde, T., 2013. U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis. Exp. Mol. Med. 45, e42.

Boerke, A., Dieleman, S., Gadella, B., 2007. A possible role for sperm RNA in early embryo development. Theriogenology 68S, 147–155.

Bracke, A., Peeters, K., Punjabi, U., Hoogewijs, D., Dewilde, S., 2018. A search for molecular mechanisms underlying male idiopathic infertility. Reprod. Biomed. Online 36, 327–339.

Burl, R.B., Clough, S., Sendler, E., Estill, M., Krawetz, S.A., 2018. Sperm RNA elements as markers of health. Syst. Biol. Reprod. Med. 64, 25–38.

Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mole 25, 169–193.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin. Chem. 55, 611–622.

Bustin, S.A., Mueller, R., 2005. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. Clin. Sci. 109, 365–379.

C

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927.

Camprubí, C., Salas-Huetos, A., Aiese-Cigliano, R., Godo, A., Pons, M.C., Castellano, G., Grossmann, M., Sanseverino, W., Martin-Subero, J.I., Garrido, N., Blanco, J., 2016. Spermatozoa from infertile patients exhibit differences of DNA methylation associated with spermatogenesis-related processes: an array-based analysis. Reprod. Biomed. Online 33, 709–719.

Cappallo-Obermann, H., Schulze, W., Jastrow, H., Baukloh, V., Spiess, A., 2011. Highly purified spermatozoal RNA obtained by a novel method indicates an un- usual 28S/18S rRNA ratio and suggests impaired ribosome assembly. Mol. Hum. Reprod. 17, 669–678.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Lomax, J., Mungall, C., Hitz, B., Balakrishnan, R., Dolan, M., Wood, V., Hong, E., Gaudet, P., 2009. AmiGO: Online access to ontology and annotation data. Bioinformatics 25, 288–289.

Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., de Rooij, D.G., Hannon, G.J., 2007. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. Dev. Cell 12, 503–514.

Cedenho, A., Lima, S., Cenedeze, M., Spaine, D., Ortiz, V., Oehninger, S., 2006. Oligozoospermia and heat-shock protein expression in ejaculated spermatozoa. Hum. Reprod. 21, 1791–1794.

Centers for Disease Control and Prevention, 2002. Vital and Health Statistics Series 23, No. 26, CDC.

Chapman, J.R., Waldenström, J., 2015. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. PLoS One 10, 1–18.

Chen, K., Mai, Z., Zhou, Y., Gao, X., Yu, B., 2012. Low NRF2 mRNA expression in spermatozoa from men with low sperm motility. Tohoku J Exp Med 228, 259–266.

Chervoneva, I., Li, Y., Schulz, S., Croker, S., Wilson, C., Waldman, S.A., Hyslop, T., 2010. Selection of optimal reference genes for normalization in quantitative RT-PCR. BMC Bioinformatics 11.

Chuma, S., Nakano, T., 2013. piRNA and spermatogenesis in mice. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368, 20110338.

Cleveland, W.S., Devlin, S.J., 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J. ofthe Am. Stat. Assoc. 83, 596–610.

Cleveland, W.S., Grosse, E., 1991. Computational Methods for Local Regression. Stat. Comput. 1, 47–62.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 1–19.

Cooper, T.G., Noonan, E., von Eckardstein, S., Auger, J., Baker, H.W.G., Behre, H.M., Haugen, T.B., Kruger, T., Wang, C., Mbizvo, M.T., Vogelsong, K.M., 2009. World Health Organization reference values for human semen characteristics. Hum. Reprod. Update 16, 231–245.

Corlan, A.D., 2004. Medline trend: automated yearly statistics of PubMed results for any query. Web Resour. URLhttp//dan.corlan.net/medline-trend.htm Accessed:

Corral-Vazquez, C., Anton, E., 2019. Sperm ncRNAs. In: Camprubí, C., Blanco, J. (Eds.), Epigenetics and Assisted Reproduction: An Introductory Guide. CRC Press, Boca Raton.

Corral-Vazquez, C., Blanco, J., Salas-Huetos, A., Vidal, F., Anton, E., 2017a. Normalization matters: tracking the best strategy for sperm miRNA quantification. Mol. Hum. Reprod. 23, 45–53.

Corral-Vazquez, C., Blanco, J., Vidal, F., Anton, E., 2017b. RNA espermático: ¿huella de eventos pasados o dote para el embrión? Med. Reprod. y Embriol. Clínica 4, 59–71.

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.

Cummins, J., 2001. Cytoplasmic inheritance and its implications for animal biotechnology. Theriogenology 55, 1381–1399.

D

Dadoune, J., 2009. Spermatozoal RNAs: What about their functions? Microsc. Res. Tech. 72, 536–551.

Dadoune, J., Pawlak, A., Alfonsi, M., Siffroi, J., 2005. Identification of transcripts by macroarrays, RT-PCR and in situ hybridization in human ejaculate spermatozoa. Mol. Hum. Reprod. 11, 133–140.

Das, M.K., Andreassen, R., Haugen, T.B., Furu, K., 2016. Identification of Endogenous Controls for Use in miRNA Quantification in Human Cancer Cell Lines 68, 63–68.

De Mateo, S., Sassone-corsi, P., 2014. Regulation of spermatogenesis by small non-coding RNAs: role of the Germ Granule. Semin. Cell Dev. Biol. 29, 84–92.

Del Giudice, P.T., Lima, S.B., Cenedeze, M.A., Pacheco-Silva, A., Bertolla, R.P., Cedenho, A.P., 2010. Expression of the Fas-ligand gene in ejaculated sperm from adolescents with and without varicocele. J. Assist. Reprod. Genet. 27, 103–109.

Depa-Martynow, M., Kempisty, B., Jagodziński, P.P., Pawelczyk, L., Jedrzejczak, P., 2012. Impact of protamine transcripts and their proteins on the quality and fertilization ability of sperm and the development of preimplantation embryos. Reprod. Biol. 12, 57–72.

Depa-Martynów, M., Kempisty, B., Lianeri, M., Jagodziński, P., Jedrzejczak, P., 2007. Association between fertilin beta, 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.

Dheda, K., Huggett, J.F., Chang, J.S., Kim, L.U., Bustin, S.A., Johnson, M.A., Rook, G.A.W., Zumla, A., 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal. Biochem. 344, 141–143.

Dohle, G., Colpi, G., Hargreave, T., Papp, G., Jungwirth, A., Weidner, W., 2005. EAU guidelines on male infertility. Eur. Urol. 48, 703–711.

Dong, W.W., Li, H.M., Qing, X.R., Huang, D.H., Li, H.G., 2016. Identification and characterization of human testis derived circular RNAs and their existence in seminal plasma. Sci. Rep. 6, 1–11.

Donovan, J.A., Koretzky, G.A., 1993. CD45 and the immune response. J. Am. Soc. Nephrol. 4, 976–985.

Du, Y., Li, M., Chen, J., Duan, Y., Wang, X., Qiu, Y., Cai, Z., Gui, Y., Jiang, H., 2016. Promoter targeted bisulfite sequencing reveals DNA methylation profiles associated with low sperm motility in asthenozoospermia. Hum. Reprod. 31, 24–33.

E

Eggers, S., DeBoer, K.D., Van Den Bergen, J., Gordon, L., White, S.J., Jamsai, D., McLachlan, R.I., Sinclair, A.H., O'Bryan, M.K., 2015. Copy number variation associated with meiotic arrest in idiopathic male infertility. Fertil. Steril. 103, 214–219.

Esteves, S.C., Miyaoka, R., Agarwal, A., 2011. An update on the clinical assessment of the infertile male. Clinics 66, 691–700.

Evans, C., Hardin, J., Stoebel, D.M., 2018. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief. Bioinform. 19, 776–792.

F

Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigyarto, C.A.-K., Skogs, M., Takanen, J.O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlén, M., 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics 13, 397–406.

Fedder, J., 1996. Nonsperm cells in human semen: With special reference to seminal leukocytes and their possible influence on fertility. Arch. Androl. 36, 41–65.

Fei, Z. hua, Yu, X. juan, Zhou, M., Su, H. fang, Zheng, Z., Xie, C. ying, 2016. Upregulated expression of long non-coding RNA LINC00982 regulates cell proliferation and its clinical relevance in patients with gastric cancer. Tumor Biol. 37, 1983–1993.

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.

Ferlin, A., Speltra, E., Patassini, C., Pati, M., Garolla, A., Caretta, N., Foresta, C., 2010. Heat shock protein and heat shock factor expression in

sperm: relation to oligozoospermia and varicocele. J. Urol. 183, 1248–1252.

Fernandes, J., Acuña, S., Aoki, J., Floeter-Winter, L., Muxel, S., 2019. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. Non-Coding RNA 5, 17.

Flegel, C., Vogel, F., Hofreuter, A., Wojcik, S., Schoeder, C., Kieć-Kononowicz, K., Brockmeyer, N.H., Müller, C.E., Becker, C., Altmüller, J., Hatt, H., Gisselmann, G., 2016. Characterization of non-olfactory GPCRs in human sperm with a focus on GPR18. Sci. Rep. 6, 1–10.

Fleige, S., Pfaffl, M.W., 2006. RNA integrity and the effect on the real-time qRT-PCR performance. Mol. Aspects Med. 27, 126–139.

Fullston, T., Ohlsson-Teague, E.M.C., Print, C.G., Sandeman, L.Y., Lane, M., 2016. Sperm microRNA content is altered in a mouse model of male obesity, but the same suite of microRNAs are not altered in offspring's sperm. PLoS One 11, 1–18.

G

Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Farinelli, L., Miska, E., Mansuy, I.M., 2014. Feeding the Next Generation of Corn Ethanol Byproducts To. Nat. Neurosci. 17, 667–669.

Garcia-Herrero, S., Garrido, N., Martinez-Conejero, J.A., Remohi, J., Pellicer, A., Meseguer, M., 2011. Differential transcriptomic profile in spermatozoa achieving pregnancy or not via ICSI. Reprod. Biomed. Online 22, 25–36.

Garrido, N., Martínez-Conejero, J.A., Jauregui, J., Horcajadas, J.A., Simón, C., Remohí, J., Meseguer, M., 2009. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. Fertil. Steril. 91, 1307–1310.

Giebler, M., Greither, T., Müller, L., Mösinger, C., Behre, H.M., 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.

Gilbert, S., 2000. Developmental Biology, 6th ed. Sunderland (USA).

Goodrich, R., Anton, E., Krawetz, S., 2013. Isolating mRNA and small noncoding RNAs from human sperm. Methods Mol. Biol. 927, 385–

96.

Goodrich, R.J., Johnson, G.D., Krawetz, S., 2007. The preparation of human spermatozoal RNA for clinical analysis. Arch. Androl. 53, 161–167.

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.

Gutierrez, L., Mauriat, M., Guénin, S., Pelloux, J., Lefebvre, J.F., Louvet, R., Rusterucci, C., Moritz, T., Guerineau, F., Bellini, C., Van Wuytswinkel, O., 2008. The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnol. J. 6, 609–618.

Guzick, D., Overstreet, J., Factor-Litvak, P., Brazil, C., Nakajima, S., Coutifaris, C., Carson, S., Cisneros, P., Steinkampf, M., Hill, J., Xu, D., Vogel, D., Network., N.C.R.M., 2001. Sperm morphology, motility, and concentration in fertile and infertile men. N. Engl. J. Med. 345, 1388–1393.

H

Hamada, A., Esteves, S.C., Agarwal, A., 2011. Unexplained male infertility: potential causes and management. Hum. Androl. 1, 2–16.

Hamatani, T., 2012. Human spermatozoal RNAs. Fertil. Steril. 97, 275–281.

Haraguchi, T., Ishikawa, T., Yamaguchi, K., Fujisawa, M., 2009. Cyclin and protamine as prognostic molecular marker for testicular sperm extraction in patients with azoospermia. Fertil. Steril. 91, 1424–1426.

Hardin, B., Kim, E., 2015. Potential Male Etiologies of Unexplained Infertility. In: Unexplained Infertility. Springer US, New York, pp. 53–56.

Hashemi, M.S., Mozdarani, H., Ghaedi, K., Nasr-Esfahani, M.H., 2018. Expression of ZMYND15 in Testes of Azoospermic Men and Association With Sperm Retrieval. Urology 114, 99–104.

Hashemitabar, M., Sabbagh, S., Orazizadeh, M., Ghadiri, A., Bahmanzadeh, M., 2015. A proteomic analysis on human sperm tail: Comparison between normozoospermia and asthenozoospermia. J. Assist. Reprod. Genet. 32, 853–863.

Hashimshony, T., Wagner, F., Sher, N., Yanai, I., 2012. CEL-Seq: Single-Cell

RNA-Seq by Multiplexed Linear Amplification. Cell Rep. 2, 666–673.

Herrada, G., Wolgemuth, D.J., 1997. The mouse transcription factor Stat4 is expressed in haploid male germ cells and is present in the perinuclear theca of spermatozoa. J. Cell Sci. 110, 1543–1553.

Heyn, H., Ferreira, H.J., Bassas, L., Bonache, S., Sayols, S., Sandoval, J., Esteller, M., Larriba, S., 2012. Epigenetic Disruption of the PIWI Pathway in Human Spermatogenic Disorders. PLoS One 7.

Horgan, R.P., Clancy, O.H., Myers, J.E., Baker, P.N., 2009. An overview of proteomic and metabolomic technologies and their application to pregnancy research. BJOG An Int. J. Obstet. Gynaecol. 116, 173–181.

Hosken, D.J., Hodgson, D.J., 2014. Why do sperm carry RNA? Relatedness, conflict, and control. Trends Ecol. Evol. 29, 451–455.

Host, E., Mikkelsen, A.L., Lindenberg, S., Smidt-Jensen, S., 2000. Apoptosis in human cumulus cells in relation to maturation stage and cleavage of the corresponding oocyte. Acta Obs. Gynecol Scand 79, 936–940.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Huggett, J., Dheda, K., Bustin, S., Zumla, A., 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes Immun. 6, 279–84.

Huo, F., Li, Y., Yang, X., Wang, Q., Liu, J., Wang, L., Su, Y., Sun, L., 2017. Expressions of SLC22A14 and SPAG6 proteins in the ejaculated sperm of idiopathic asthenozoospermia patients. Zhonghua Nan Ke Xue 23, 703–707.

J

Janghorban-Laricheh, E., Ghazavi-Khorasgani, N., Tavalaee, M., Zohrabi, D., Abbasi, H., Nasr- Esfahani, M.H., 2016. An association between sperm PLCζ levels and varicocele? J. Assist. Reprod. Genet. 33, 1649–1655.

Jayaraman, V., Upadhya, D., Narayan, P.K., Adiga, S.K., 2012. Sperm processing by swim-up and density gradient is effective in elimination of sperm with DNA damage. J. Assist. Reprod. Genet. 29,

557–563.

Jedrzejczak, P., Kempisty, B., Bryja, A., Mostowska, M., Depa-Martynow, M., Pawelczyk, L., Jagodzinski, 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.

Jensen, T.K., Henriksen, T.B., Ernst, E., Scheike, T., Bonde, J.P.E., Olsen, J., Giwercman, A., Hjollund, N.H.I., Kolstad, H., Skakkebæk, N.E., 2002. Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. Lancet 352, 1172–1177.

Ji, Z., Lu, R., Mou, L., Duan, Y.G., Zhang, Q., Wang, Y., Gui, Y., Cai, Z., 2014. Expressions of miR-15a and its target gene HSPA1B in the spermatozoa of patients with varicocele. Reproduction 147, 693–701.

Jodar, M., Kalko, S., Castillo, J., Ballescà, J.L., Oliva, R., 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Hum. Reprod. 27, 1431–1438.

Jodar, M., Selvaraju, S., Sendler, E., Diamond, M.P., Krawetz, S.A., 2013. The presence, role and clinical use of spermatozoal RNAs. Hum. Reprod. Update 19, 604–624.

K

Kanduri, C., 2016. Long noncoding RNAs: Lessons from genomic imprinting. Biochim. Biophys. Acta - Gene Regul. Mech. 1859, 102–111.

Karlic, R., Ganesh, S., Franke, V., Svobodova, E., Urbanova, J., Suzuki, Y., Aoki, F., Vlahovicek, K., Svoboda, P., 2017. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. DNA Res. 24, 129–141.

Keel, B.A., 2006. Within- and between-subject variation in semen parameters in infertile men and normal semen donors. Fertil. Steril. 85, 128–134.

Kempisty, B., Depa-Martynow, M., Lianeri, M., Jedrzejczak, P., Darul-Wasowicz, A., Jagodzinski, P., 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. Folia Histochem. Cytobiol. 45, 109–113.

Kleiman, S.E., Lehavi, O., Hauser, R., Botchan, A., Paz, G., Yavetz, H., Yogev,

L., 2011. CDY1 and BOULE transcripts assessed in the same biopsy as predictive markers for successful testicular sperm retrieval. Fertil. Steril. 95, 2297-2302.e1.

Kocabas, A.M., Crosby, J., Ross, P.J., Otu, H.H., Beyhan, Z., Can, H., Tam, W.-L., Rosa, G.J.M., Halgren, R.G., Lim, B., Fernandez, E., Cibelli, J.B., 2006. The transcriptome of human oocytes. Proc. Natl. Acad. Sci. U. S. A. 103, 14027–32.

Kosova, G., Scott, N.M., Niederberger, C., Prins, G.S., Ober, C., 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. Am. J. Hum. Genet. 90, 950–961.

Kotaja, N., 2014. MicroRNAs and spermatogenesis. Fertil. Steril. 101, 1552–1562.

Kozera, B., Rapacz, M., 2013. Reference genes in real-time PCR. J. Appl. Genet. 54, 391–406.

Krausz, C., Giachini, C., Lo Giacco, D., Daguin, F., Chianese, C., Ars, E., Ruiz-Castane, E., Forti, G., Rossi, E., 2012. High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males. PLoS One 7.

Krawetz, S., 2005. Paternal contribution: new insights and future challenges. Nat. Rev. Genet. 6, 633–642.

Krawetz, S.A., Kruger, A., Lalancette, C., Tagett, R., Anton, E., Draghici, S., Diamond, M.P., 2011. A survey of small RNAs in human sperm. Hum. Reprod. 26, 3401–3412.

Kruger, T.F., Acosta, A.A., Simmons, K.F., Swanson, R.J., Matta, J.F., Oehninger, S., 1988. Predictive value of abnormal sperm morphology in in vitro fertilization. Fertil. Steril. 49, 112–117.

Kuo, P.L., Lin, Y.H., Teng, Y.N., Hsu, C.C., Lin, J.S.N., Lin, Y.M., 2004. Transcriptional levels of four Y chromosome-linked AZF genes in azoospermic men and their association with successful sperm retrieval. Urology 63, 131–136.

Kuramochi-Miyagawa, S., 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131, 839–849.

L

Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., Krawetz, S.A., 2009. Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. 87, 735–748.

Lamba, V., Ghodke-Puranik, Y., Guan, W., Lamba, J.K., 2014. Identification of suitable reference genes for hepatic microRNA quantitation. BMC Res. Notes 7, 129.

Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A., Carreau, S., 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: Relationship to sperm motility and capacitation. Mol. Hum. Reprod. 10, 535–541.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Ferrante, C., Yang, J.L., Terry, R., Jeanty, S.S.F., Li, C., Peters, D.T., Turczyk, B.M., Marblestone, A.H., Samuel, A., Bernard, A., Rios, X., Aach, J., George, M., 2014. Highly multiplexed subcellular RNA sequencing in situ. Science (80-.). 343, 1360–1363.

Li, C., Zheng, L., Wang, C., Zhou, X., 2010. Absence of nerve growth factor and comparison of tyrosine kinase receptor A levels in mature spermatozoa from oligoasthenozoospermic, asthenozoospermic and fertile men. Clin. Chim. Acta 411, 1482–1486.

Li, H., Huang, S., Guo, C., Guan, H., Xiong, C., 2012. Cell-free seminal mRNA and microRNA exist in different forms. PLoS One 7.

Li, Y., Li, M., Liu, Y., Song, G., Liu, N., 2012. A Microarray for MicroRNA Profiling in Spermatozoa from Adult Men Living in an Environmentally Polluted Site. Bull. Environ. Contam. Toxicol. 89, 1111–1114.

Lima-Souza, A., Anton, E., Mao, S., Ho, W.J., Krawetz, S.A., 2012. A platform for evaluating sperm RNA biomarkers: Dysplasia of the fibrous sheath - Testing the concept. Fertil. Steril. 97, 1061-1066.e3.

Liu, K.S., Li, T.P., Ton, H., Mao, X.D., Chen, Y.J., 2018. Advances of Long Noncoding RNAs-mediated Regulation in Reproduction. Chin. Med. J. (Engl). 131, 226–234.

Liu, W.-M., Pang, R.T.K., Chiu, P.C.N., Wong, B.P.C., Lao, K., Lee, K.-F., Yeung, W.S.B., 2012. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. Proc. Natl. Acad. Sci. U. S. A. 109, 490–4.

Lopes, A.M., Aston, K.I., Thompson, E., Carvalho, F., Gonçalves, J., Huang, N., Matthiesen, R., Noordam, M.J., Quintela, I., Ramu, A., Seabra, C., Wilfert, A.B., Dai, J., Downie, J.M., Fernandes, S., Guo, X., Sha, J., Amorim, A., Barros, A., Carracedo, A., Hu, Z., Hurles, M.E., Moskovtsev, S., Ober, C., Paduch, D.A., Schiffman, J.D., Schlegel, P.N., Sousa, M., Carrell, D.T., Conrad, D.F., 2013. Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and

Both Sex Chromosomes, including the Gene DMRT1. PLoS Genet. 9.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. PLOS Comput. Biol. 13, e1005457.

Luk, A.C.S., Chan, W.Y., Rennert, O.M., Lee, T.L., 2014. Long noncoding RNAs in spermatogenesis: Insights from recent high-throughput transcriptome studies. Reproduction 147, R131–R141.

Luo, B., He, W., Wang, S., Xiong, H., Liang, X., Shan, X., Lü, Z., 2017. Expressions of ODF2 mRNA and protein are down-regulated in the sperm of asthenospermia patients. Natl. J. Androl. 23, 1002–1006.

Luo, L.-F., Hou, C.-C., Yang, W.-X., 2015. Small non-coding RNAs and their associated proteins in spermatogenesis. Gene 578, 141–157.

M

Malcher, A., Rozwadowska, N., Stokowy, T., Kolanowski, T., Jedrzejczak, P., Zietkowiak, W., Kurpisz, M., 2013. Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. Fertil. Steril. 100, 1686-1694.e7.

Malvezzi, H., Sharma, R., Agarwal, A., Abuzenadah, A.M., Abu-Elmagd, M., 2014. Sperm quality after density gradient centrifugation with three commercially available media: A controlled trial. Reprod. Biol. Endocrinol. 12, 1–7.

Mao, S., Goodrich, R.J., Hauser, R., Schrader, S.M., Chen, Z., Krawetz, S.A., 2013. Evaluation of the effectiveness of semen storage and sperm purification methods for spermatozoa transcript profiling. Syst. Biol. Reprod. Med. 59, 287–295.

Mao, S., Sendler, E., Goodrich, R.J., Hauser, R., Krawetz, S.A., 2014. A Comparison of Sperm RNA-Seq Methods. Syst Biol Reprod Med 60, 308–315.

Marra, M., Hillier, L., Waterston, R., 1998. Expressed sequence tags--ESTablishing bridges between genomes. Trends Genet. 14, 4–7.

Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., Vandesompele, J., 2009. A novel and universal method for microRNA RT-qPCR data normalization. Genome Biol. 10, R64.

Miki, K., Willis, W.D., Brown, P.R., Goulding, E.H., Fulcher, K.D., Eddy, E.M., 2002. Targeted Disruption of the Akap4 Gene Causes Defects in Sperm Flagellum and Motility. Dev. Biol. 248, 331–342.

Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R., Krawetz, S.A., 1999. A complex population of RNAs exists in human ejaculate spermatozoa: Implications for understanding molecular aspects of spermiogenesis. Gene 237, 385–392.

Miller, D., Ostermeier, G.C., Krawetz, S.A., 2005. The controversy, potential and roles of spermatozoal RNA. Trends Mol. Med. 11, 156–163.

Modi, D., Shah, C., Sachdeva, G., Gadkar, S., Bhartiya, D., Puri, C., 2005. Ontogeny and cellular localization of SRY transcripts in the human testes and its detection in spermatozoa. Reproduction 130, 603–613.

Moghissi, K.S., Wallach, E.E., 1983. Unexplained infertility. Fertil. Steril. 39, 5–21.

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.

Montjean, D., Zini, A., Ravel, C., Belloc, S., Dalleac, A., Copin, H., Boyer, P., Mcelreavey, K., Benkhalifa, M., 2015. Sperm global DNA methylation level: Association with semen parameters and genome integrity. Andrology 3, 235–240.

Motiei, M., Tavalaee, M., Rabiei, F., Hajihosseini, R., Nasr-Esfahani, M.H., 2013. Evaluation of HSPA2 in fertile and infertile individuals. Andrologia 45, 66–72.

Mraz, M., Malinova, K., Mayer, J., Pospisilova, S., 2009. MicroRNA isolation and stability in stored RNA samples. Biochem. Biophys. Res. Commun. 390, 1–4.

N

Naz, R., Ahmad, K., Kumar, G., 1991. Presence and role of c-myc proto-oncogene product in mammalian sperm cell function. Biol. Reprod. 44, 842–850.

Nelson, N.J., 2001. Microarrays Have Arrived: Gene Expression Tool Matures. J. Natl. Cancer Inst. 93, 492–494.

Nguyen, M.T., Delaney, D.P., Kolon, T.F., 2009. Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. Fertil. Steril. 92, 182–187.

Ni, K., Steger, K., Yang, H., Wang, H., Hu, K., Chen, B., 2014. Sperm Protamine mRNA Ratio and DNA Fragmentation Index Represent

Reliable Clinical Biomarkers for Men with Varicocele after Microsurgical Varicocele Ligation. J. Urol. 192, 170–176.

Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., Pera, R.A.R., 2012. Human pre-implantation embryo development. Development 139, 829–841.

Nixon, B., Mitchell, L.A., Anderson, A.L., Mclaughlin, E.A., O'bryan, M.K., Aitken, R.J., 2011. Proteomic and functional analysis of human sperm detergent resistant membranes. J. Cell. Physiol. 226, 2651–2665.

O

Ostermeier, G., Goodrich, R., Moldenhauer, J., Diamond, M., Krawetz, S., 2005. A suite of novel human spermatozoal RNAs. J. Androl. 26, 70–74.

Ostermeier, G., Miller, D., Huntriss, J., Diamond, M., Krawetz, S., 2004. Delivering spermatozoan RNA to the oocyte. Nature 429, 154.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., Krawetz, S.A., 2002. Spermatozoal RNA profiles of normal fertile men. Mech. Dis. 360, 772–777.

Ostermeier, G.C., Goodrich, R.J., Diamond, M.P., Dix, D.J., Krawetz, S.A., 2005. Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. Fertil. Steril. 83, 1687–1694.

P

Pacheco, S.E., Anderson, L.M., Sandrof, M.A., Vantangoli, M.M., Hall, S.J., Boekelheide, K., 2012. Sperm mRNA Transcripts Are Indicators of Sub-Chronic Low Dose Testicular Injury in the Fischer 344 Rat. PLoS One 7.

Pacheco, S.E., Houseman, E.A., Christensen, B.C., Marsit, C.J., Kelsey, K.T., Sigman, M., Boekelheide, K., 2011. Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. PLoS One 6, 1–10.

Pansa, A., Sirchia, S.M., Melis, S., Giacchetta, D., Castiglioni, M., Colapietro, P., Fiori, S., Falcone, R., Paganini, L., Bonaparte, E., Colpi, G., Miozzo, M., Tabano, S., 2014. ESX1 mRNA expression in seminal fluid is an

indicator of residual spermatogenesis in non-obstructive azoospermic men. Hum. Reprod. 29, 2620–2627.

Pantano, L., Jodar, M., Bak, M., Ballescà, J.L., Tommerup, N., Oliva, R., Vavouri, T., 2015. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. RNA 21, 1085–1095.

Park, T., Yi, S.G., Kang, S.H., Lee, S.Y., Lee, Y.S., Simon, R., 2003. Evaluation of normalization methods for microarray data. BMC Bioinformatics 4, 1–13.

Parker, R., Barnes, N., 1999. mRNA: detection by in Situ and northern hybridization. Methods Mol. Biol. 106, 247–283.

Parte, P.P., Rao, P., Redij, S., Lobo, V., D'Souza, S.J., Gajbhiye, R., Kulkarni, V., 2012. Sperm phosphoproteome profiling by ultra performance liquid chromatography followed by data independent analysis (LC-MSE) reveals altered proteomic signatures in asthenozoospermia. J. Proteomics 75, 5861–5871.

Peimbert, M., Alcazar, L., 2016. A Hitchhiker's Guide to Metatranscriptomics. In: AM.Aransay, Trueba, J.L. (Eds.), Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing. Springer International Publishing, pp. 313–342.

Pelloni, M., Paoli, D., Majoli, M., Pallotti, F., Carlini, T., Lenzi, A., Lombardo, F., 2018. Molecular study of human sperm RNA: Ropporin and CABYR in asthenozoospermia. J. Endocrinol. Invest. 41, 781–787.

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, 844–52.

Pessot, C., Brito, M., Figueroa, J., Concha, I., Yanez, A., Burzio, L., 1989. Presence of RNA in the sperm nucleus. Biochem. Biophys. Res. Commun. 158, 272–278.

Pillai, R.S., Chuma, S., 2012. piRNAs and their involvement in male germline development in mice. Dev. Growth Differ. 54, 78–92.

Pilu, R., 2011. Paramutation: Just a Curiosity or Fine Tuning of Gene Expression in the Next Generation? Curr. Genomics 12, 298–306.

Place, R., Li, L., Pookot, D., Noonan, E., Dahiya, R., 2008. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc. Natl. Acad. Sci. USA 105, 1608–1613.

Platts, A.E., Dix, D.J., Chemes, H.E., Thompson, K.E., Goodrich, R., Rockett, J.C., Rawe, V.Y., Quintana, S., Diamond, M.P., Strader, L.F., Krawetz, S.A., 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum. Mol. Genet. 16, 763–773.

Ponting, C.P., Oliver, P.L., Reik, W., 2009. Evolution and Functions of Long Noncoding RNAs. Cell 136, 629–641.

Q

Qiu, J., Ren, Z., Yan, J., 2016. Identification and functional analysis of long non-coding RNAs in human and mouse early embryos based on single-cell transcriptome data. Oncotarget 7.

Quinn, J.J., Chang, H.Y., 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17, 47–62.

R

R Core Team, 2014. R: A language and environment for statistical computing. [WWW Document]. R Found. Stat. Comput. Vienna, Austria.

Rahman, M.S., Lee, J.-S., Kwon, W.-S., Pang, M.-G., 2013. Sperm Proteomics: Road to Male Fertility and Contraception. Int. J. Endocrinol. 2013, 1–11.

Raposo, G., Stoorvogel, W., 2013. Extracellular vesicles : Exosomes, microvesicles, and friends. J. Cell Biol. 200, 373–383.

Riesco, M.F., Valcarce, D.G., Martínez-Vázquez, J.M., Martín, I., Calderón-García, A.Á., Gonzalez-Nunez, V., Robles, V., 2019. Male reproductive dysfunction in Solea senegalensis: New insights into an unsolved question. Reprod. Fertil. Dev. 1104–1115.

Rio, D.C., Ares, M., Hannon, G.J., Nilsen, T.W., 2010. Purification of RNA using TRIzol (TRI Reagent). Cold Spring Harb. Protoc. 5, 1–4.

Rogenhofer, N., Dansranjavin, T., Schorsch, M., Spiess, A., Wang, H., Von Schönfeldt, V., Cappallo-Obermann, H., Baukloh, V., Yang, H., Paradowska, A., Chen, B., Thaler, C.J., Weidner, W., Schuppe, H.C., Steger, K., 2013. The sperm protamine mRNA ratio as a clinical parameter to estimate the fertilizing potential of men taking part in

an ART programme. Hum. Reprod. 28, 969–978.

Röther, S., Meister, G., 2011. Small RNAs derived from longer non-coding RNAs. Biochimie 93, 1905–1915.

RStudio Team, 2015. RStudio: Integrated Development for R.

S

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

Salas-Huetos, A., Blanco, J., Vidal, F., Grossmann, M., Pons, M.C., Garrido, N., Anton, E., 2016. Sperm from normozoospermic fertile and infertile individuals convey a distinct miRNA cargo. Andrology 4, 1028–1036.

Salas-Huetos, A., Blanco, J., Vidal, F., Mercader, J.M., Garrido, N., Anton, E., 2014. New insights into the expression profile and function of micro-ribonucleic acid in human spermatozoa. Fertil. Steril. 102, 213–222.

Salas-Huetos, A., Moraleda, R., Giardina, S., Anton, E., Blanco, J., Salas-Salvadó, J., Bulló, M., 2018. Effect of nut consumption on semen quality and functionality in healthy males: A randomized controlled trial. Am. Hournal Clin. Nutr. 108, 953–962.

Saleh, R.A., Agarwal, A., Nada, E.A., El-Tonsy, M.H., Sharma, R.K., Meyer, A., Nelson, D.R., Thomas, A.J., 2003. Negative effects of increased sperm DNA damage in relation to seminal oxidative stress in men with idiopathic and male factor infertility. Fertil. Steril. 79, 1597–1605.

Salone, V., Rederstorff, M., 2015. Stem-Loop RT-PCR Based Quantification of Small Non-Coding RNAs 1296, 103–108.

Sato, Y., Tajima, A., Tsunematsu, K., Nozawa, S., Yoshiike, M., Koh, E., Kanaya, J., Namiki, M., Matsumiya, K., Tsujimura, A., Komatsu, K., Itoh, N., Eguchi, J., Imoto, I., Yamauchi, A., Iwamoto, T., 2015. An association study of four candidate loci for human male fertility traits with male infertility. Hum. Reprod. 30, 1510–1514.

Saunders, C., Swann, K., Lai, F., 2007. PLCzeta, a sperm-specific PLC and its potential role in fertilization. Biochem. Soc. Simp. 74, 23–36.

Savadi-Shiraz, E., Edalatkhah, H., Talebi, S., Heidari-Vala, H., Zandemami, M., Pahlavan, S., Modarressi, M.H., Akhondi, M.M., Paradowska-

Dogan, A., Sadeghi, M.R., 2015. Quantification of sperm specific mRNA transcripts (PRM1, PRM2, and TNP2) in teratozoospermia and normozoospermia: New correlations between mRNA content and morphology of sperm. Mol. Reprod. Dev. 82, 26–35.

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, 1–14.

Schulz, K.N., Harrison, M.M., 2019. Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. 20, 221–234.

Schütte, B., El Hajj, N., Kuhtz, J., Nanda, I., Gromoll, J., Hahn, T., Dittrich, M., Schorsch, M., Müller, T., Haaf, T., 2013. Broad DNA methylation changes of spermatogenesis, inflammation and immune response-related genes in a subgroup of sperm samples for assisted reproduction. Andrology 1, 822–829.

Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., Krawetz, S.A., 2013. Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. 41, 4104–4117.

Sharlip, I.D., Jarow, J.P., Belker, A.M., Lipshultz, L.I., Sigman, M., Thomas, A.J., Schlegel, P.N., Howards, S.S., Nehra, A., Damewood, M.D., Overstreet, J.W., Sadovsky, R., 2002. Best practice policies for male infertility. Fertil. Steril. 77, 873–882.

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.

Sigman, M., Lipshultz, L., Howards, S., 2009. Office evaluation of the subfertile male. In: Lipshultz, L., Howards, S., Niederberger, C. (Eds.), Infertility in the Male. Cambridge University Press, Cambridge, pp. 153–176.

Sim, G.K., Kafatos, F.C., Jones, C.W., Koehler, M.D., Efstratiadis, A., Maniatis, T., 1979. Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families. Cell 18, 1303–1316.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: Key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132.

Song, G.J., Lee, H.S., Park, Y.S., Lee, H.J., Lee, Y.S., Seo, J.T., Kang, I.S., 2000. Expression pattern of germ cell-specific genes in the testis of

patients with nonobstructive azoospermia: Usefulness as a molecular marker to predict the presence of testicular sperm. Fertil. Steril. 73, 1104–1108.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H., Yan, W., 2011. Male germ cells express abundant endogenous siRNAs. Proc. Natl. Acad. Sci. U. S. A. 108, 13159–13164.

Stahl, P.J., Mielnik, A., Schlegel, P.N., Paduch, D.A., 2011. Heat shock factor y chromosome (HSFY) mRNA level predicts the presence of retrievable testicular sperm in men with nonobstructive azoospermia. Fertil. Steril. 96, 303–308.

Steger, K., Wilhelm, J., Konrad, L., Stalf, T., Greb, R., Diemer, T., Kliesch, S., Bergmann, M., Weidner, W., 2008. 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. 23, 11–16.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Chen, J., Blelloch, R., 2010. MicroRNA Function Is Globally Suppressed in Mouse Oocytes and Early Embryos. Curr. Biol. 20, 271–277.

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., Yaspo, M.-L., 2014. Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics 15, 675.

Supek, F., Bosnjak, M., Skunca, N., Smuc, T., 2011. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6.

T

T.Moon, R., Brown, J.D., Torres, M., 1997. WNTs modulate cell fate and behavior during vertebrate development. Trends Genet. 13, 157–162.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., Heinen, E., 1999. Housekeeping genes as internal standards: Use and limits. J. Biotechnol. 75, 291–295.

Tomlinson, M.J., 2016. Uncertainty of measurement and clinical value of semen analysis: has standardisation through professional guidelines helped or hindered progress? Andrology 4, 763–770.

Tong, M.-H., Mitchell, D., Evanoff, R., Griswold, M.D., 2011. Expression of Mirlet7 family microRNAs in response to retinoic acid-induced

spermatogonial differentiation in mice. Biol. Reprod. 85, 189–197.

Tricarico, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M., Bustin, S.A., Orlando, C., 2002. Quantitative real-time reverse transcription polymerase chain reaction: Normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal. Biochem. 309, 293–300.

Tüttelmann, F., Simoni, M., Kliesch, S., Ledig, S., Dworniczak, B., Wieacker, P., Röpke, A., 2011. Copy number variants in patients with severe oligozoospermia and sertoli-cell-only syndrome. PLoS One 6.

U

Urdinguio, R.G., Bayón, G.F., Dmitrijeva, M., Toraño, E.G., Bravo, C., Fraga, M.F., Bassas, L., Larriba, S., Fernández, A.F., 2015. Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility. Hum. Reprod. 30, 1014–1028.

V

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., Lötvall, J.O., 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat. Cell Biol. 9, 654–659.

Van Der Steeg, J.W., Steures, P., Eijkemans, M.J.C., Habbema, J.D.F., Hompes, P.G.A., Kremer, J.A.M., Van Der Leeuw-Harmsen, L., Bossuyt, P.M.M., Repping, S., Silber, S.J., Mol, B.W.J., Van Der Veen, F., 2011. Role of semen analysis in subfertile couples. Fertil. Steril. 95, 1013–1019.

Van Roy, F., Berx, G., 2008. The cell-cell adhesion molecule E-cadherin. Cell. Mol. Life Sci. 65, 3756–3788.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.

Vomelová, I., Vanícková, Z., Sedo, A., 2009. Methods of RNA purification. All ways (should) lead to Rome. Folia Biol. (Praha). 55, 243–51.

Von Schönfeldt, V., Wistuba, J., Schlatt, S., 2004. Notch-1, c-kit and GFRα-1 are developmentally regulated markers for premeiotic germ cells. Cytogenet. Genome Res. 105, 235–239.

W

Wang, H., Zhou, Z., Xu, M., Li, J., Xiao, J., Xu, Z.Y., Sha, J., 2004. A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. J. Mol. Med. 82, 317–324.

Wang, Liguo, Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J.E., Dasari, S., Vedell, P.T., Barman, P., Wang, Liewei, Weinshiboum, R., Jen, J., Huang, H., Kohli, M., Kocher, J.P.A., 2016. Measure transcript integrity using RNA-seq data. BMC Bioinformatics 17, 1–16.

Wang, Y., Tang, N., Hui, T., Wang, S., Zeng, X., Li, H., Ma, J., 2013. Identification of endogenous reference genes for RT-qPCR analysis of plasma microRNAs levels in rats with acetaminophen-induced hepatotoxicity. J. Appl. Toxicol. 33, 1330–1336.

Wanowska, E., Kubiak, M.R., Rosikiewicz, W., Makałowska, I., Szcześniak, M.W., 2018. Natural antisense transcripts in diseases: From modes of action to targeted therapies. Wiley Interdiscip. Rev. RNA 9, 1–16.

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W., Zhang, J., Bi, X., Dai, J., Zhang, Q., Lu, Z.J., Gao, G., 2016. Critical roles of long noncoding RNAs in Drosophila Spermatogenesis. Genome Res. 26, 1233–1244.

WG. Guder, S. Narayanan, H. Wisser, B.Z., 2009. Diagnostic Samples: From the Patient to the Laboratory: The Impact of Preanalytical Variables on the Quality of Laboratory Results, 4th ed. Wiley-VCH Verlag GmbH, Weinheim, Germany.

Wichman, L., Somasundaram, S., Breindel, C., Valerio, D.M., McCarrey, J.R., Hodges, C.A., Khalil, A.M., 2017. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. Biol. Reprod. 97, 313–323.

World Health Organization, 1992. WHO Laboratory Manual for the Examination of Human Semen and Sperm-cervical Mucus Interaction, 3rd ed. Cambridge University Press, Cambridge.

World Health Organization, 1999. WHO Laboratory Manual for the

Examination of Human Semen and Sperm-cervical Mucus Interaction, 4th ed. Cambridge University Press, Cambridge.

World Health Organization, 2000. WHO manual for the standardized investigation and diagnosis of the infertile couple, 4th ed. Cambridge University Press, Cambridge UK.

World Health Organization, 2010. WHO laboratory manual for the examination and processing of human semen, 5th ed. ed. Geneva: World Health Organization.

Wotschofsky, Z., Meyer, H.A., Jung, M., Fendler, A., Wagner, I., Stephan, C., Busch, J., Erbersdobler, A., Disch, A.C., Mollenkopf, H.J., Jung, K., 2011. Reference genes for the relative quantification of microRNAs in renal cell carcinomas and their metastases. Anal. Biochem. 417, 233–241.

Wykes, S., Visscher, D., Krawetz, S., 1997. Haploid transcripts persist in mature human spermatozoa. Mol. Hum. Reprod. 3, 15–19.

Wylie, D., Shelton, J., Choudhary, A., Adai, A.T., 2011. A novel mean-centering method for normalizing microRNA expression from high-throughput RT-qPCR data. BMC Res. Notes 4, 555.

X

Xu, B., Hao, Z., Jha, K., Zhang, Z., Urekar, C., Digilio, L., Pulido, S., Strauss, J. 3rd, Flickinger, C., Herr, J., 2008. Targeted Deletion of Tssk1 & 2 Causes Male Infertility Due to Haploinsufficiency Bingfang. Dev. Biol. 319, 211–222.

Y

Yadav, S., 2017. Multi-omics and male infertility status integration and future prospects. Front. Biosci. 9, 493.

Yan, Q., Ma, Y., Chen, R., Zhou, X., Qiao, J., Xian, Y., Feng, L., Chen, C., 2018. Expression of DKKL1 in spermatozoa of men with asthenospermia. Nan Fang Yi Ke Da Xue Xue Bao 38, 324–328.

Yang, Y., Cheng, L., Wang, Y., Han, Y., Liu, J., Deng, X., Chao, L., 2017. Expression of NDUFA13 in asthenozoospermia and possible pathogenesis. Reprod. Biomed. Online 34, 66–74.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., Speed, T.P., 2002.

Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. Sel. Work. Terry Speed 30, 591–600.

Yatsenko, A.N., Georgiadis, A.P., Murthy, L.J., Lamb, D.J., Matzuk, M.M., 2013. UBE2B mRNA alterations are associated with severe oligozoospermia in infertile men. Mol. Hum. Reprod. 19, 388–394.

Yuan, S., Schuster, A., Tang, C., Yu, T., Ortogero, N., Bao, J., Zheng, H., Yan, W., 2016. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development 143, 635–647.

Z

Zalata, A.A., Mokhtar, N., Atwa, A., Khaled, M., Shaker, O.G., 2016. The Role of Protamine 2 Gene Expression and Caspase 9 Activity in Male Infertility. J. Urol. 195, 796–800.

Zangbar, M.-S.S., Keshtgar, S., Zolghadri, J., Gharesi-Fard, B., 2016. Antisperm protein targets in azoospermia men. J. Hum. Reprod. Sci. 9, 47.

Zegers-Hochschild, F., Adamson, G.D., de Mouzon, J., Ishihara, O., Mansour, R., Nygren, K., Sullivan, E., Vanderpoel, S., 2009. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009*. Fertil. Steril. 92, 1520–1524.

Zhang, C., Li, X.Y., Luo, Z.Z., Wu, T.W., Hu, H., 2019. Upregulation of LINC00982 inhibits cell proliferation and promotes cell apoptosis by regulating the activity of PI3K/AKT signaling pathway in renal cancer. Eur. Rev. Med. Pharmacol. Sci. 23, 1443–1450.

Zhang, L., Liu, Z., Li, X., Zhang, P., Wang, J., Zhu, D., Chen, X., Ye, L., 2015. Low long non-coding RNA HOTAIR expression is associated with down-regulation of Nrf2 in the spermatozoa of patients with asthenozoospermia or oligoasthenozoospermia. Int. J. Clin. Exp. Pathol. 8, 14198–14205.

Zhang, X., Zhang, P., Song, D., Xiong, S., Zhang, H., Fu, J., Gao, F., Chen, H., Zeng, X., 2019. Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm. Biol. Reprod. 100, 982–993.

Zhao, H., Xu, J., Zhang, H., Sun, J., Sun, Y., Wang, Z., Liu, J., Ding, Q., Lu, S., Shi, R., You, L., Qin, Y., Zhao, X., Lin, X., Li, X., Feng, J., Wang, L., Trent, J.M., Xu, C., Gao, Y., Zhang, B., Gao, X., Hu, J., Chen, H., Li, G., Zhao, J., Zou, S., Jiang, H., Hao, C., Zhao, Y., Ma, J., Zheng, S.L., Chen, Z.J., 2012. A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. Am. J. Hum. Genet. 90, 900–906.

Zhao, Y., Li, Q., Yao, C., Wang, Z., Zhou, Y., Wang, Y., Liu, L., Wang, L., Qiao, Z., 2006. Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. Hum. Reprod. 21, 1583–1590.

Zheng, L., Li, C., Sun, Y., Liu, Z., Zhou, X., 2011. Expression of brain-derived neurotrophic factor in mature spermatozoa from fertile and infertile men. Clin. Chim. Acta 412, 44–47.

Zhong Wang, M.G., Snyder, M., 2001. RNA-Seq: a revolutionary tool for transcriptomics Zhong. Bioinforma. A Pract. Guid. to Anal. Genes Proteins 10, 57–63.

Zhou, J.-H., Zhou, Q.Z., Lyu, X.-M., Zhu, T., Chen, Z.-J., Chen, M.-K., Xia, H., Wang, C.-Y., Qi, T., Li, X., Liu, C.-D., 2015. The expression of cysteine-rich secretory protein 2 (CRISP2) and its specific regulator miR-27b in the spermatozoa of patients with asthenozoospermia. Biol Reprod 92, 1–9.

Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S., Li, Z., 2016. Dynamics of the Transcriptome during Human Spermatogenesis: Predicting the Potential Key Genes Regulating Male Gametes Generation. Sci. Rep. 6, 19069.

Ziyyat, A., Lefèvre, A., 2001. Differential gene expression in pre-implantation embryos from mouse oocytes injected with round spermatids or spermatozoa. Hum. Reprod. 16, 1449–1456.

Zou, S., Li, Zheng, Wang, Y., Chen, T., Song, P., Chen, J., He, X., Xu, P., Liang, M., Luo, K., Zhu, X., Tian, E., Du, Q., Wen, Z., Li, Zhiqiang, Wang, M., Sha, Y., Cao, Y., Shi, Y., Hu, H., 2014. Association Study Between Polymorphisms of PRMT6, PEX10, SOX5, and Nonobstructive Azoospermia in the Han Chinese Population1. Biol. Reprod. 90, 1–4.

7 Appendix

Appendix I

Characterization of the mRNA and lncRNA cargo of human spermatozoa by RNA-seq

Corral-Vazquez, C., Blanco, J., Anton, E. (2019)

(In preparation)

Characterization of the mRNA and lncRNA cargo of human spermatozoa by RNA-seq

Celia Corral-Vazquez, Joan Blanco, Ester Anton

RNA-seq combines high-throughput sequencing with deep computational reconstruction analyses to identify and quantify RNA molecules from a given biological sample. Its application in human sperm cells will improve the characterization of the complex transcriptome contained in this highly specialized cell. In this study, we provide a deep characterization of the sperm transcriptome by RNA-seq technologies, focusing on data integration and functional description of both mRNAs and lncRNAs. For this purpose, 12 semen samples from fertile donors were collected and processed for RNA extraction. The obtained RNA samples were subjected to strict quality controls to verify the absence of RNA from non-sperm cells and DNA. cDNA libraries of poly(A) RNA molecules were constructed by means of the Universal Plus mRNAseq with NuQuant kit (NuGEN). Paired-end sequencing of 125 bp fragments was performed using the HiSeq 2500 platform (Illumina). The obtained reads were aligned by STAR algorithm to the human genome (GRCh38 alignment), and expression levels of mRNAs and lncRNAs were calculated with FeatureCounts. Transcripts were classified according to their mean expression values (Fragments per Kilobase per Million) and their stability (Transcript Integrity Number). The results revealed the presence of a complex population of RNAs with a high degradation state, but also a common cargo of preserved transcripts that is retained in different individuals, thus suggesting a relevant functionality of these transcripts. Gene Ontology of the 13,126 *Expressed* mRNAs (FPKM≥2) and the 5,253 *Highly Expressed* mRNAs (FPKM≥10) showed a significant involvement in processes that included spermatogenesis and reproduction. The relation with these processes is reinforced when analyzing the 784 *Ubiquitously Expressed* and the 200 *Highly Stable* mRNAs, which indicates their very specific role. The 7,521 *Expressed* lncRNAs (FPKM≥0.5) were predicted to *cis*-regulate a set of targets genes ontologically related to biological processes that included cell adhesion and embryo development. This enrichment was even more evident when analyzing those *cis*-regulated targets that were not present among the sperm mRNA cargo, which would be indicative of the further role of these lncRNAs upon their release into the oocyte. Finally, several pairs of *Ubiquitously Expressed* transcripts displayed a correlated expression throughout all samples. These RNAs emerge as potential candidates for further analysis focused on the search for molecular biomarker tools.

Introduction

With the discovery of RNA in the nucleus of mature spermatozoa during the late 80s (Pessot et al., 1989), the transcriptome of these transcriptionally inert cells became a new research point of interest. Throughout the last decades, a complex sperm RNA population has been discovered, composed both by a coding fraction (mRNA) and a non-coding set of small and large transcripts (known as non-coding RNAs or ncRNAs) (Miller et al., 1999). In spite of the loss of the majority of the cytoplasm produced during spermiogenesis, some RNAs are retained at the head and midpiece of mature sperm cells (Dadoune et al., 2005; Modi et al., 2005). Although a partial degradation of these molecules has been described (Boerke et al., 2007), several authors have identified hosts of full-length sperm transcripts

that remain stable (G. C. Ostermeier et al., 2005; Ostermeier et al., 2002).

Due to the inactive translational activity caused by rRNA degradation in mature spermatozoa (Ostermeier et al., 2002), no active functionality was attributed to stored mRNAs. Some authors have revealed that the coding sperm RNA fraction mainly represent a fingerprint of previous spermatogenic events (Ostermeier et al., 2002; Wang et al., 2004). Nevertheless, other studies have pointed out that part of the stable sperm mRNAs are delivered to the oocyte at fertilization. Moreover, some of these transcripts have been related to biological functions that are necessary for Zygote Genome Activation (ZGA) and embryo development (Boerke et al., 2007; Ostermeier et al., 2004a).

Besides coding RNA, the non-coding fraction of sperm RNA also entails strong biological relevance. Specifically, long non-coding RNAs (lncRNAs) constitute an emerging field in Reproduction research. These transcripts can be divided into five different classes depending on their location with respect to coding genes: sense, antisense, bidirectional, intronic, and intergenic (Ponting et al., 2009). The functionality of lncRNAs is fundamentally regulatory, although their regulation mechanisms are still in the process of full characterization. However, it is known that their role in transcriptional regulation can be performed in close proximity (regulation in *cis*) or targeting distant sequences (regulation in *trans*). Therefore, they can produce gene expression activation or repression by different pathways involving their binding to DNA regions or proteins (Ponting et al., 2009). By these mechanisms, a potential intervention of sperm lncRNAs in the control of spermatogenesis and sperm functionality has been identified, e.g., the potential involvement of lnc32058, lnc09522 and lnc98487 in the regulation of sperm motility (Zhang et al., 2019), or the control of sperm defects related to ROS effects (Zhang et al., 2015). Nevertheless, the whole spectrum of functions related to sperm-retained lncRNAs still remains unclear, as well as their possible implication in the control of embryonic gene expression in early stages.

In order to fully understand the biological relevance of mature sperm RNAs, deep data analyses are needed. From the first sperm transcriptomic characterization (Miller et al., 1999), several microarray results aiming to profile the expression patterns of sperm RNA have been published (Dadoune et al., 2005; Lalancette et al., 2009; G. Ostermeier et al., 2005; G. C. Ostermeier et al., 2005; Ostermeier et al., 2002). Nevertheless, the application of this technique on sperm RNA implies several limitations. One of them is the scarce amount of RNA contained by these cells (10-100 femtograms per spermatozoa) when ~1 µg of RNA is needed for microarray analysis (Krawetz, 2005; Lowe et al., 2017). To solve this handicap, the practice of pooling different samples is recurrent, which impedes individual comparisons. Another limitation of microarray technology is the moderate sensitivity provided by fluorescence detection in comparison with other techniques. Also, the possibilities of transcript identification by this technique are limited to the sequences included in the array (Lowe et al., 2017).

The development of state-of-the-art transcriptomic approaches like Next

Generation Sequencing (NGS) has provided a more sensitive technology, a wider dynamic range and higher sequence resolution using a lower initial input quantity (Lowe et al., 2017). Also, the characterized transcripts are not restricted to a set of predesigned sequences, but can be profiled from updated massive databases of coding and non-coding RNAs (Conesa et al., 2016). For these reasons, the use of RNA sequencing (RNA-seq) in human sperm RNAs has begun to proliferate with intend to expand the current knowledge. Sendler et al. took three single and three pooled samples from healthy donors to perform the first sperm RNA-seq characterization study of coding and non-coding RNAs. This study allowed identifying known and novel sperm transcripts, and revealed specific regulation patterns and functionality of RNAs both in previous and subsequent fertilization stages (Sendler et al., 2013). More recently, Flegel et al. performed RNA-seq to analyze the expression of the different types of G protein-coupled receptors (GPRCs) in sperm cells of 10 individuals, finding 223 different GPRC transcripts and highlighting the involvement of the newly detected GPR18 in sperm physiology (Flegel et al., 2016). Additionally, some authors performed comparative analyses in order to optimize the methodological process of sperm RNA-seq profiling. The results showed a lack of biases caused by different semen storage methods, RNA purification strategies, and amount of input RNA library preparation protocols. Nevertheless, the outcomes did differ depending on the sperm isolation methodology and amplification protocol used previously to library construction (Mao et al., 2014, 2013). Concerning lncRNAs, the use of RNA-seq for profiling this human sperm transcriptome fraction is still very new. Recently, Zhang et al. (Zhang et al., 2019) made a comparative study between lncRNA and mRNAs of 48 fertile and 36 asthenozoospermic patients, indicating an association between the expression of certain lncRNAs and sperm motility.

Although the mentioned studies are appropriately developed, upgrading the current knowledge about human sperm mRNA and lncRNA is still dependent of further research. Thus, additional in-depth RNA-seq characterization studies based on wide and well-characterized fertile populations are needed for increasing and integrating the sperm transcriptomic network.

In this context, the main objective of the present work is to provide a deep characterization of the sperm transcriptome by RNA-seq technologies, focusing on the data integration and functional description of both sperm mRNAs and lncRNAs. To achieve this goal, the methodology was designed to solidly sustain the obtained outcomes, regarding both the population included in the study and the followed protocols. Accordingly, a large population of fertile donors was characterized under strict selection criteria, and their semen samples were carefully processed by an optimized methodology for sperm transcriptome obtainment and processing, with strict quality controls. By means of this design, this study offers a set of reliable sperm NGS data that enables deepening both into the total human sperm RNA cargo and the host of common stable and preserved transcripts. These include a novel set of RNAs that can be considered as

candidates for further analysis in the field of biomarker research and fertility diagnosis.

Materials and methods

Sample collection

A total of 12 semen samples from donors were collected by masturbation after 3-7 days of sexual abstinence and cryopreserved with EggYolk Freezing Medium (IRVINE) in 1:1 proportion. The inclusion criteria of the donors comprised proven fertility, normal karyotype, no previous exposure to any genotoxic agent, and no history of chemotherapy, radiotherapy or chronic illness. All donors displayed a normozoospermic profile according to the standard thresholds established by World Health Organization (World Health Organization, 2010). That is, total number of spermatozoa per ejaculate with progressive motility above 90 · 10^6, more than 4% of normal sperm forms, and more than 10 · 10^6 spermatozoa/mL with progressive motility after post-thawing cryosurvival. The average age of the donors was 24.08, ranging from 19 to 31 years (**Supplemental Table 1**).

Biological samples were provided by the Instituto Valenciano de Infertilidad center (IVI Valencia, Spain). Written informed consent was obtained from all donors. The study was approved by the Ethics Committees of the collaborative center and of the Universitat Autònoma de Barcelona (ref. CEEAH1884).

Sperm RNA extraction

Semen samples were thawed and centrifuged at 2500 x g to eliminate the seminal plasma and freezing medium. The cell fraction was subjected to selective lysis of non-sperm cells by the method described by Goodrich et al. (Goodrich et al., 2007). Briefly, samples were incubated in a lysis buffer (0.1% Sodium Dodecyl Sulfate and 0.5% Triton X-100 in milliQ water) for 30 minutes. Non-sperm cell elimination was verified by optical microscopic examination. The treatment was repeated until obtaining less than one non-sperm cell in 10,000 spermatozoa.

Sperm samples were subjected to total RNA extraction by either *mir*Vana™ PARIS™ RNA and Native Protein Purification Kit (Thermo Fisher), or Maxwell® RSC simplyRNA Cells Kit, (Promega) following the manufacturer's protocol (for a detailed protocol see Appendix II).

Quality controls of RNA samples

RNA quantification was performed by a Quantus® Fluorometer (Promega). The full-spectrum UV-Vis spectro-photometer NanoDrop© 2000 (Thermo Fisher) was used to determine RNA purity by measuring their 260/280 nm absorbance ratio.

RNA size distribution profiles were analyzed using the RNA 6000 Pico chip with the Agilent DNA High Sensitivity Bioanalyzer (Thermo Fisher). The absence of the peaks corresponding to 28S and 18S rRNAs was confirmed as a control to verify the absence of non-sperm cells in the samples.

Additionally, the absence of non-sperm RNAs and DNA contamination was corroborated by RT-PCR (High-

Capacity cDNA Reverse Transcription kit, Thermo Fisher) using primers for *PRM1*, *GAPDH*, (both expressed in sperm cells), *PTPRC*/*CD45* (expressed in leukocytes), *KIT* (expressed in germ cells), and *CDH1* (expressed in epithelial cells). The primers were designed following an exon-exon strategy. Therefore, amplicons from genomic sequences (including introns) and cDNA (without introns) could be discerned by their size (**Supplemental Table 2**). Primer3 v.3.0.0 software (http://primer3.wi.mit.edu) and the Genome Browser of UCSC) (https://genome.ucsc.edu/) were used for primer design.

Library construction and sequencing

The cDNA libraries were constructed using the Universal Plus mRNAseq with NuQuant kit (NuGEN) following the manufacturer's protocol. Briefly, poly(A) RNA selection was performed by Oligo(dT) beads binding and elution, followed by RNA fragmentation and double-strand cDNA synthesis using a mixture of random and oligo(dT) primering. End-repairs were performed from purified cDNA to generate blunt ends, and adaptors were ligated to the fragments. Finally, strand selection and PCR amplification were made in order to obtain the final libraries, which were quantified by Qubit Fluorometer (Thermo Fisher).

Libraries were sequenced on paired-end 125 bp mode using the HiSeq 2500 platform (Illumina) with library sizes of 30M reads per sample. The Bcl2Fastq 2.0.2 version of the Illumina pipeline was used to produce raw data and to perform de-multiplexing.

Data analysis

The pipeline analysis design is summarized in **Figure 1**. Firstly, reads were aligned to the GRCh38 assembly of the human genome using STAR v.2.5.3a. Expression counts were estimated by featureCounts (Subread package v.1.5.2). Genome annotation of mRNAs and lncRNAs were obtained from USCS Table Browser (https://genome.ucsc.edu/) and GENCODE (https://www.gencodegenes.org/), respectively.

Raw data were normalized by Fragments Per Kilobase Million (FPKM) that allows correcting raw data by gene length and library sizes.

Subsequent bioinformatic and statistical analyses were made using R statistical programming in RStudio (http://www.rstudio.com/) (R Core Team, 2014; RStudio Team, 2015). The packages FactoMineR v.1.41, Factoextra v.1.0.5, Hmisc v.4.2.0, pheatmap v.1.0.10, ggplot2 v.3.1.0, RCircos v.1.2.1, ggsci v.2.9, and RColorBrewer v.1.1.2 were employed.

A Principal Component Analysis (PCA) was performed with total normalized expression data to identify variations and similarities across samples, and to discard possible outliers.

The mean expression levels and chromosomal distribution corresponding to the sequences of mRNAs and lncRNAs were assessed.

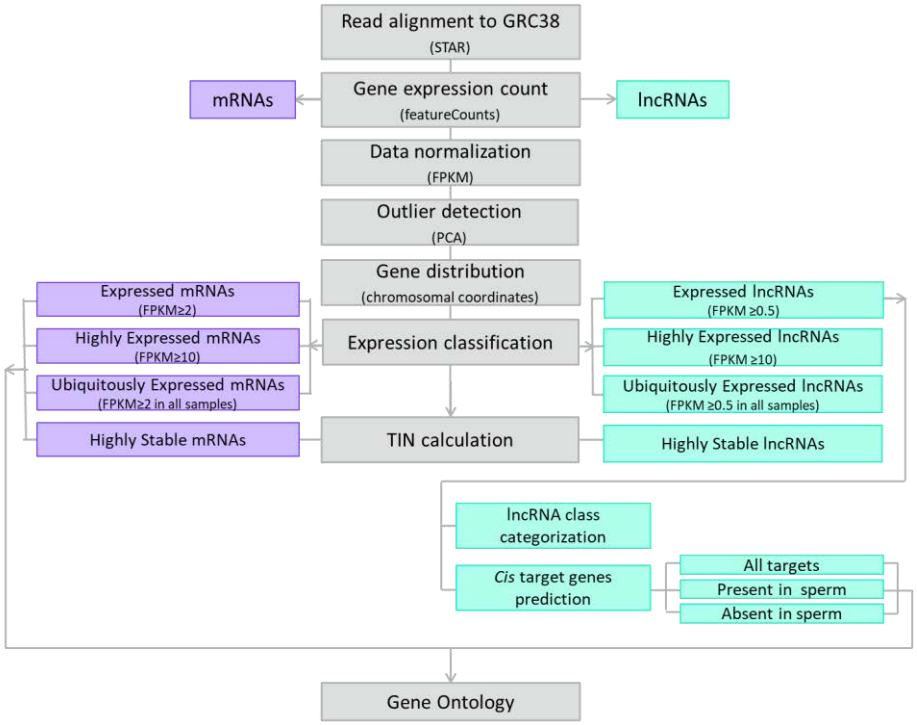


Figure 1. Schematic summary of the bioinformatics pipeline.

FPKM = Fragments Per Kilobase Million; PCA = Principal Component Analysis; TIN = Transcript Integrity Number.

The total set of identified mRNAs was classified according to their FPKM values into: *Expressed* (mean FPKM≥2), *Highly Expressed* (mean FPKM≥10), and *Ubiquitously Expressed* (mean FPKM≥2 in all samples).

Concerning lncRNAs, they were classified into: *Expressed* (mean FPKM≥0.5), *Highly Expressed* (mean FPKM≥10), and *Ubiquitously Expressed* (mean FPKM≥0.5 in all samples).

Among the *Highly Expressed* set of mRNAs, a heatmap clustering analysis of the 500 mRNAs with the greatest mean FPKM value was used to assess sample classification and determine the expression homogeneity of the study population.

In order to identify the most stable and preserved transcripts among the set of *Expressed* mRNAs and lncRNAs, the Transcript Integrity Number (TIN) was calculated, which can range from 0 to 100 for each analyzed transcript (Wang et al., 2016). TIN score was estimated using the RSeQC v.2.6.4 package. The 200 transcripts with higher TIN were classified as *Highly Stable* transcripts.

All *Expressed*, *Highly Expressed*, *Ubiquitously Expressed* and *Highly Stable* mRNAs were submitted separately to a Gene Ontology (GO) functional analysis in DAVID v.6.8 (Database for Annotation, Visualization and Integrated Discovery) (https://david.ncifcrf.gov/home.jsp)

(Huang et al., 2009a, 2009b). Significantly enriched GO terms (P-values<0.05 after Bonferroni correction) were clustered with REViGO (Reduce and Visualize Gene Ontology) (http://revigo.irb.hr/) (Supek et al., 2011).

The *Expressed* lncRNAs were categorized according to their class (i.e. sense, antisense, intergenic or others) using GENCODE annotation data (https://www.gencodegenes.org). A prediction of target genes in *cis* of these lncRNAs was performed by identifying those genes located within a 10 kb window upstream and downstream the annotated lncRNA sequences using *BEDtools* v.2.27.1. These targets were classified as *Present* (predicted genes that were also identified among the *Expressed* mRNAs) and *Absent* (predicted genes that were not listed among the category of *Expressed* mRNAs). GO analysis was performed as described above over all predicted lncRNA targets, the *Present* targets, and the *Absent* targets, separately.

Finally, the presence of pairs of transcripts which maintained a correlated expression level in all samples was assessed by means of a Spearman test over every *Ubiquitously Expressed* mRNA-mRNA and lncRNA-lncRNA combination. P-values<0.05 after Bonferroni correction were considered significant.

Results

Sperm RNA quality

Results of the total human sperm RNA isolation from the 12 samples are summarized in **Supplemental Table 1**. The total amount of sperm RNA obtained in the analyzed samples indicated the presence of 2.78-181 femtograms (fg) of RNA per cell, with an average value of 22.91±50.09 fg/cell. The 260/280 nm ratio of the samples reflected a range from 1.43 to 2.26, with a mean purity of 1.80±0.24. No signal corresponding to 18S or 28S was detected in the nano-electrophoresis analysis of any sample.

RT-PCR results of the marker genes displayed cDNA amplicons for *PRM1* (331 bp) and *GAPDH* (228 bp), with no amplification of *CD45*, *KIT* or *CDH1*. This pattern was homogeneous for all the samples included in the study.

Altogether, these results confirmed a sufficient yield of good quality sperm RNA, discarding traces of DNA or RNA from non-sperm cell

Sample homogeneity

PCA sample distribution displayed a gathered distribution of all samples (except for S10), showing no segregation associated to the extraction protocol. The analysis showed a separation of sample S10 with respect to the rest of the samples. Consequently, it was considered as an outlier and discarded in further analyses (**Supplemental figure 1A**). The non-biased homogeneity of the remaining 11 samples was also reinforced by the results of the heatmap referred to the 500 most expressed mRNAs, which showed a homogeneous clusterization (**Supplemental figure 1B**). In consequence, the influence of external factors such are demographic variables or methodological or technical issues over the expression profiles of the different samples was disregarded. This homogeneity verification allowed proceeding with the subsequent data analyses.

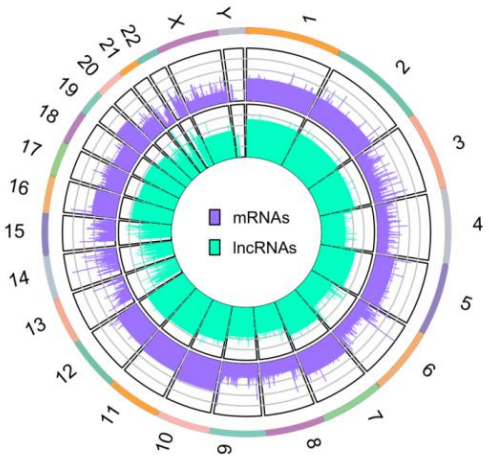


Figure 2. Chromosome distribution and mean expression value of the *Expressed* sperm mRNAs and lncRNAs. Log10 mean FPKM values are represented at the chromosomal coordinates of each sequence.

FPKM = Fragments Per Kilobase Million.

mRNA and lncRNA expression and distribution

A total of 13,126 mRNAs were identified as *Expressed* in the analyzed sperm samples. From them, 40.02% were classified as *Highly Expressed* (5,253 mRNAs), and 5.97% were categorized as *Ubiquitously Expressed* (784 mRNAs).

Regarding lncRNAs, 7,521 sequences were identified as *Expressed*, from which 56.56% were classified as *Highly Expressed* (4,254 lncRNAs), and 1.54% were selected as *Ubiquitously Expressed* (116 lncRNAs).

Concerning the *Highly Stable* transcripts, the list of the 200 mRNAs and lncRNAs with the highest TIN score is summarized in **Supplemental Table 3**. The average TIN value of all samples was 10.54 for mRNAs and 0.34 for lncRNAs.

Chromosomal distribution of the *Expressed* mRNAs and lncRNAs showed an overall homogeneous coverage of all genome, with notorious gaps corresponding to the p-arm of acrocentric chromosomes, and the Yqh region (**Figure 2**).

Functionality profiling of sperm mRNA

The functional annotation of the *Expressed* mRNAs was revealed by GO. The REViGO clustering of the results displayed six significantly enriched biological processes (**Figure 3A**). These main superclustered processes referred to biological functions related to *Single-organism Organelle Organization* (GO:1902589; this supercluster encompassed 32 biological processes), *Protein Modification by Small Protein Conjugation or Removal* (GO:0070647, with 16 related processes), *Intracellular Transport* (GO:0046907, involving eight processes), *Spermatogenesis*

(GO:0007283), *Cellular Response to Stress* (GO:0033554, comprising two processes), and *Reproduction* (GO:0000003).

Regarding the fraction of *Highly Expressed* sperm mRNAs, seven REViGO superclusters of enriched processes were identified (**Figure 3B)**: *Chromosome Organization* (GO:0051276, comprising 13 associated biological processes), *Spermatogenesis*, *Microtubule-based Process* (GO:0007017, with three related processes), *Protein Modification by Small Protein Conjugation or Removal* (related to six processes), *Reproduction*, *Ribonucleoprotein Complex Localization* (GO:0071166, associated with four biological processes), and *Cellular Response to Stress*.

The *Ubiquitously Expressed* genes fraction showed an involvement in a more reduced spectrum of biological functions, displaying three main REViGO superclusters referring to *Chromosome Organization* (GO:0051276, containing 13 related biological processes), *Spermatogenesis* and *Reproduction* (**Figure 3C)**.

Lastly, regarding the *Highly Stable* mRNAs, the REViGO clustering of their related functionality showed a significant enrichment of a single supercluster: *Spermatogenesis* (**Figure 3D)**.

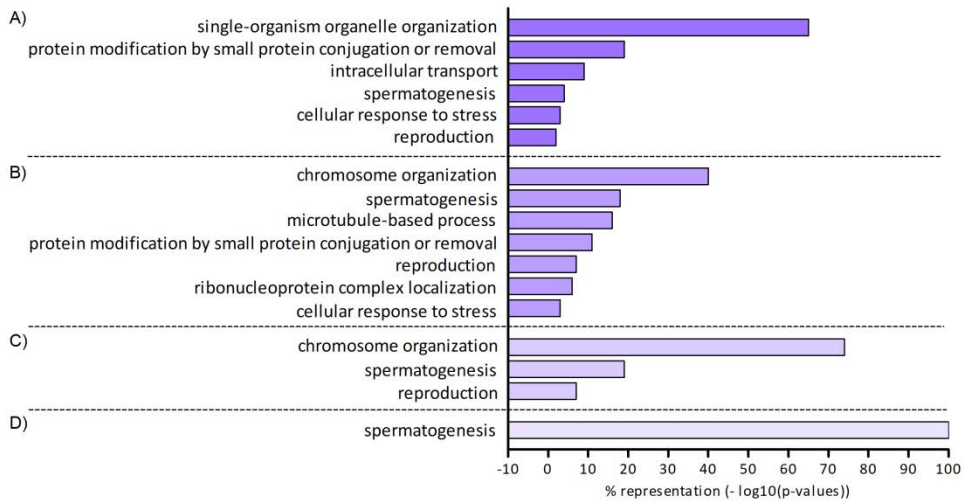


Figure 3. REViGO clustering of Gene Ontology analysis of A) *Expressed* mRNAs (mean FPKM≥2), B) *Highly Expressed* mRNAs (mean FPKM≥10), C) *Ubiquitously Expressed* mRNAs (FPKM≥2 in all samples), and D) *Highly Stable* mRNAs (the 200 genes with the highest Transcript Integrity Number). The X axis indicates the percentual representation of the significance of each supercluster.
FPKM = Fragments Per Kilobase Million.

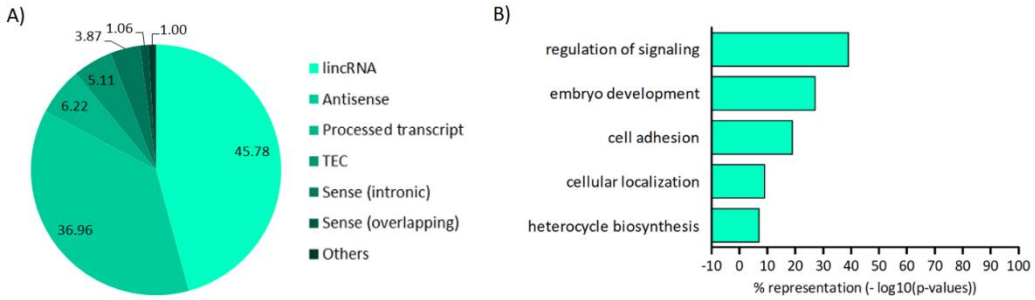


Figure 4. Characterization of *Expressed* sperm lncRNAs: A) Percentage of *Expressed* lncRNAs according to their class; B) REViGO clustering of Gene Ontology analysis of the predicted target genes.
LincRNA = long intergenic non-coding RNA; TEC = To be Experimentally Confirmed.

Sperm lncRNA characterization and potential regulatory role

The classification of the *Expressed* lncRNAs revealed that most of them corresponded to the class of long intergenic non-coding RNA (lincRNAs, 45.78%), followed by antisense lncRNAs (36.96%). Sense lncRNA sequences, both intronic and overlapping, accounted for 4.93% while <1% of the sequences belonged to other categories like bidirectional promoter lncRNAs or 3'-overlapping ncRNAs (**Figure 4A**).
18,480 genes were predicted as potential *cis* target genes of the *Expressed* lncRNAs. The *Highly Expressed* lncRNAs led to the identification of 10,922 targets, 467 were predicted from *Ubiquitously Expressed* lncRNAs, and 721 from the *Highly Stable* lncRNAs.
The REViGO analysis of the *Expressed* lncRNAs targets showed an enrichment of 5 superclusters: *Regulation of Signaling* (GO:0023051, comprehending eight related processes), *Embryo Development* (GO:0009790, with three associated processes), *Homophilic Cell Adhesion Via Plasma Membrane Adhesion Molecules* (GO:0007156, with two related processes), *Cellular Location* (GO:0051641), and *Heterocycle Biosynthetic Process* (GO:0018130) (**Figure 4B**). No significant enriched processes were detected in the GO analysis of the target genes predicted for the *Highly Expressed*, *Ubiquitously Expressed*, or *Highly Stable* lncRNAs.
From the 18,480 predicted targets of the *Expressed* lncRNAs, 6,181 were classified as *Present* among the *Expressed* mRNAs (**Figure 5A**). This set of genes showed a significant enrichment in *Cytoskeleton Organization* in the REViGO clustering (GO:0007010, containing 12 associated biological processes), as well as in *Negative Regulation of Biosynthetic Process* (GO:0009890, related to nine processes), *Cellular Location* (with four related processes), *Microtubule-based Process* (related to three processes), and *Protein Modification by Small Protein Conjugation* (**Figure 5B**). The remaining 12,299 targets of the *Expressed* lncRNAs were categorized as *Absent* (not included among the *Expressed* mRNAs). The REViGO summarization showed a significant enrichment in processes related to *Embryo Organ Development*

(GO:0048568), and *Homophilic Cell Adhesion via Plasma Membrane Adhesion Molecules* (**Figure 5B).**

Pairs of transcripts with correlated expression

The Spearman test revealed four pairs of *Ubiquitously Expressed* mRNA with a significant correlated expression throughout all samples. These eight genes were: i) Nuclear Mitotic Apparatus Protein 1 (*NUMA1*), and Chromodomain Helicase DNA Binding Protein 2 (*CHD2*); ii) Cerebellin 1 Precursor (*CBLN1*), and Required for Meiotic Nuclear Division 5 Homolog B (*RMND5B*); iii) Tousled Like Kinase 2 (*TLK2*), and Neurofibromin 1 (*NF1*); and iv) Mitochondrially Encoded Cytochrome C Oxidase I (*MT-CO1*), and Mitochondrially Encoded Cytochrome C Oxidase II (*MT-CO2*) (**Table 1**).

Regarding lncRNAs, a correlated expression of 22 pairs of *Ubiquitously Expressed* transcripts showed a significant p-value in the test (**Table 1**).

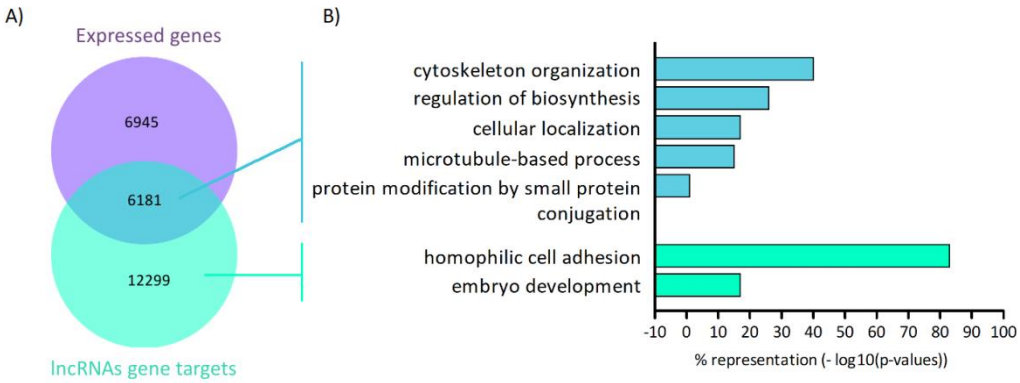


Figure 5. Representation and functionality of the *cis*-predicted targets of the *Expressed* lncRNAs.

A) In purple: Count of *Expressed* mRNAs; In green: Count of lncRNA gene targets categorized as *Absent* (not present among the Expressed mRNAs); Overlaping blue area: Count of lncRNA gene targets *Present* among the *Expressed* mRNAs.

B) REViGO clustering of the enriched processes detected in the Gene Ontology analysis of *Present* and *Absent* target genes of the *Expressed* lncRNAs.

Table 1. Transcripts with significantly correlated expression levels (p-values of the Spearman test are specified in the third column).

mRNA 1	**mRNA 2**	**P-value**
NUMA1	*CHD2*	3.76E-09
CBLN1	*RMND5B*	3.76E-09
TLK2	*NF1*	3.76E-09
MT-CO1	*MT-CO2*	8.40E-08
lncRNA 1	**lncRNA 2**	**P-value**
FP236383.2	FP671120.4	<E-10
FP236383.3	FP236383.2	<E-10
FP236383.3	FP671120.4	<E-10
FP236383.2	FP671120.5	3.76E-09
FP671120.4	FP671120.5	3.76E-09
FP236383.3	FP671120.5	3.76E-09
HMGA2-AS1	SLC26A4-AS1	8.40E-08
AIRN	AC002467.1	8.40E-08
FP236383.1	FP671120.2	8.40E-08
FP671120.9	FP236383.7	5.14E-07
HMGA2-AS1	AC002467.1	1.85E-06
HMGA2-AS1	AC020978.7	1.85E-06
LINC01320	AC002467.1	1.85E-06
LINC02050	AL627309.1	1.85E-06
AIRN	HMGA2-AS1	1.85E-06
AIRN	SLC26A4-AS1	1.85E-06
HMGA2-AS1	AC007098.1	4.99E-06
LINC01320	AC020978.7	4.99E-06
LINC01320	AIRN	4.99E-06
LINC01320	AC090984.1	4.99E-06
AC007098.1	SLC26A4-AS1	4.99E-06
FP236383.7	FP236383.1	4.99E-06

Discussion

Characteristics of the sperm transcriptome

This study supports the view of the sperm transcriptome as a complex and heterogeneous network of coding and non-coding large RNA molecules annotated throughout all the genome. Besides their complexity, a substantial fragmentation status was also revealed. In addition, the classification of the identified mRNAs and lncRNAs according to their expression level indicated that around half of them were *Highly Expressed*, and a reduced host of transcripts was maintained *Ubiquitously Expressed*. Therefore, a common fingerprint of coding and non-coding stable transcripts was preserved throughout all samples. These results reinforce the existence of a certain degree of homogeneity between sperm samples of different

individuals, which suggests a relevant functionality of these transcripts.

Sperm mRNA: stability and biological role

The identified set of mRNAs supposes a representation of the overall coding transcriptome cargo of sperm cells from fertile men. When contrasting these results with the only previous sperm RNA-seq study (Sendler et al., 2013), a certain homogeneity is observed. Thus, if the dataset collected by Sendler et al. is filtered according to our criteria (FPKM≥2 for *Expressed* mRNAs and FPKM≥10 for *Highly Expressed* mRNAs), a 62% of the *Expressed* mRNAs and 55% of the *Highly Expressed* mRNAs are coincident with the transcripts identified in our analysis. This issue suggests that a similar RNA pattern can be found in fertile donors from different cohorts of individuals independently of demographic differences, reinforcing the idea that preserved sperm transcripts could present a fertility-related functionality.

In order to understand the biological relevance of these transcripts, functional annotation analyses were performed with the set of *Expressed*, *Highly Expressed*, *Ubiquitously Expressed*, and *Highly Stable* mRNAs. GO results of the *Expressed* mRNAs showed a wide spectrum of biological functions with a predominant representation of processes that can be associated to the development of spermatogenesis. This agrees with the fact that most mRNAs retained in the mature sperm cells were produced during early and mid-spermatogenesis with the purpose of regulating representative processes of these stages such as cell division and differentiation. Among the obtained superclusters, the GO term *Reproduction* includes a wide conjunct of processes beyond gametogenesis, such as those related with fertilization and post-fertilization functions. In fact, the role of some of the identified *Expressed* mRNAs in early embryo development was experimentally validated in previous sperm RNA studies: *PLC-Z1* (produced during late spermatogenesis and directly contributing to oocyte activation) (Saunders et al., 2007), *APAK4* (essential for sperm motility and potentially involved in oocyte activation) (Miki et al., 2002), *WNT5A* (related to morphometric patterning) (T.Moon et al., 1997), and *STAT4* (with a potential regulatory role in the male pronucleus) (Herrada and Wolgemuth, 1997).

When focusing on the GO results from the *Highly Expressed* mRNAs, the functional role also appeared involved in processes related to cell cycle, spermatogenesis, reproduction, and cellular response to stress. Nevertheless, spermatogenesis and reproduction gain importance among the represented GO terms in comparison with results from *Expressed* mRNAs, implying a deeper role of these abundant transcripts in the mentioned processes. Given the relevance of these events in sperm functionality, a higher abundance of transcripts that are involved in those pathways is expected.

Concerning the 784 *Ubiquitously Expressed* mRNAs, GO results displayed a more reduced variety of enriched processes, involved in cell division (chromosome organization), spermatogenesis, and reproduction. This specialization still becomes more

evident when analyzing the *Highly Stable* transcripts since their GO analysis showed a unique involvement in spermatogenesis. Altogether, these results indicate that the host of sperm mRNAs that are ubiquitously expressed throughout fertile men and preserved from degradation are implicated in a narrow variety of biological processes and thus display a very specific role. It can be suggested that these transcripts were selectively preserved for the development of these precise and definite roles, which are mainly focused in sperm development.

Actually, comparing these outcomes with previous findings of other authors, it stands out that numerous mRNAs belonging to the *Highly Stable* and/or *Ubiquitously Expressed* transcripts have been previously described to display a significantly enriched expression in human testis or sperm cells (Fagerberg et al., 2014). Several of these transcripts have also been related by other authors to enriched GO processes that were similar to our results. It is the case of Protamine 1 and 2 (*PRM1* and *PRM2*) and Transition Protein 1 (*TNP1*), which are directly involved in protamine packaging of sperm chromatin. Spermatogenesis Associated 18 (*SPATA18*) and 3 (*SPATA3*) are also present among these genes. Another example is Testis-specific Serine Kinase 6 (*TSSK6*), since it has been described that proteins from TSSK family have a crucial role in spermatid development (Xu et al., 2008).

Regulatory role of sperm lncRNA

The results of the present study also provide insights into the functional profile of lncRNAs in sperm cells. The biological implications of these transcripts in sperm-specific processes were previously suggested by other authors basing on the especially high number of expressed lncRNAs found in testis (Cabili et al., 2011). Additionally, these ncRNAs have been associated with fertility in other species (Wen et al., 2016; Wichman et al., 2017). Nevertheless, only a small number of RNA-seq studies have focused on this non-coding fraction of human sperm cargo and its interaction network with mRNA expression. In the context of this knowledge gap, the present characterization of lncRNAs has allowed assessing their potential functionality and biological role.

Regarding their regulatory mechanism, most of the detected lncRNAs were categorized as antisense and lincRNAs. These two classes have been described to frequently regulate gene expression in *cis* (Cabili et al., 2011; Wanowska et al., 2018). The GO analysis of the 18,480 potentially *cis*-regulated target genes of the *Expressed* lncRNAs showed a significant enrichment of processes related to general cell functioning (like cellular location, heterocycle bio-synthesis, and regulation of signaling). Also, an enriched representation of embryo development and cell adhesion processes was observed, implying that the genomic location of sperm lncRNA sequences has a predominant distribution near genes involved in embryogenesis.

With respect to the target genes of the *Highly Expressed*, *Ubiquitously Expressed*, and *Highly Stable* lncRNAs, no significantly enriched processes were obtained in the analysis. However, if the strictness of the

analysis is reduced (e.g. by suppressing the Bonferroni correction), several enriched biological processes related with spermatogenesis and embryogenesis are observed from these sets of target genes (data not shown). In the case of the targets of the *Highly Expressed* lncRNA, we found processes related to cell cycle, regulation of signaling, and cell-cell adhesion. Regarding *Ubiquitously Expressed* lncRNAs, their targets showed a predominance of processes involved in DNA repair and embryogenesis-related processes. The targets of the *Highly Stable* lncRNAs displayed a representation of processes related with DNA repair and meiotic chromosome separation. Therefore, despite these results were not statistically significant by applying strict statistical criteria, they support the implication of sperm lncRNAs in spermatogenesis and embryogenesis. In fact, some of the identified *Highly Stable* lncRNAs have been related in previous studies to biological functions produced in germ cells: a regulatory role have been attributed to UBOX5-AS1 during pre-meiotic development of germ (Zhu et al., 2016), and high expression levels of LINC00982 have been found to be involved in inhibiting cell proliferation and promoting cell-cycle arrest (Fei et al., 2016; C. Zhang et al., 2019).

The functionality of sperm lncRNAs diverges when analyzing the *Present* and *Absent* target genes. The set of *Present* target genes in sperm cells showed a significant implication in GO terms related to cytoskeleton organization pathways that can reflect cell division processes. On the other hand, those *Absent* target genes displayed an exclusive implication in cell adhesion and embryo development. The fact that embryogenesis-related targets showed no expression in sperm cells might suggest a possible biological role of sperm lncRNA after fertilization, during the early stages of the zygote. It is known that ZGA is produced at day 3 of human embryo development (at the 4- to 8-cell stage), and parental RNAs are degraded at that step (Niakan et al., 2012). In this sense, it has been stated that sperm lncRNAs remain stable in the embryo until that point (Karlic et al., 2017). Therefore, it is possible that sperm lncRNAs are preserved after their release to the oocyte, being involved in early embryogenesis gene regulation. In fact, a study performed by Qiu et al. in human preimplantation embryos has recently shown that lncRNAs participated in ZGA, oocyte maturation, and mitochondrial functions. Moreover, the network of lncRNAs involved in ZGA was found to be highly conserved between human and mouse embryos, revealing a conserved regulatory fingerprint (Qiu et al., 2016). Therefore, our results and other outcomes found in literature bring significant hints towards the implication of paternal-originated lncRNAs in the embryogenesis regulation network. However, an experimental identification of sperm lncRNAs in early zygotic stages is a pending key. This step would allow a deeper understanding of the role of sperm non-coding RNA in reproduction.

At this point, and regarding the significance of our results about lncRNA profiling, it is important to highlight that the followed sample preparation methodology for RNA-seq did not follow a lncRNA-specific selection protocol. Instead, an overall

characterization design was pursued in order to allow the inclusion of coding and non-coding transcripts with a thorough selection of poly(A) RNAs. For that reason, oligo(dT) beads were used in the initial transcript selection process prior to sequencing. This step enabled the inclusion of both mRNAs and the fraction of lncRNAs that present polyadenylation (Fernandes et al., 2019). The presence of absence of polyadenilation in lncRNA depends mostly on their biogenesis process, so poly(A) is more frequent in antisense sequences (Quinn and Chang, 2016). This fact is consistent with the high percentage of antisense transcripts identified among *Expressed* lncRNAs in comparison with sense sequences. Therefore, the interpretation of the quantification and biological implications of these lncRNAs should be considered in a much careful manner as they did not constitute the specific target molecules of the initial poly(A) selection step. Consequently, it is important to bear in mind that our RNA-seq results of sperm lncRNAs do not necessarily represent the whole lncRNA sperm cargo but only a fraction of them. For these reasons, it should be considered that the complete spectrum of lncRNA-related biological processes could have not been entirely detected in this study. Nevertheless, the biological functions associated to the identified set of sperm lncNAs reinforce a significant involvement of these transcripts in sperm-specific processes and embryogenesis regulation, suggesting novel aspects of their regulatory implications.

Sperm transcripts as fertility biomarkers

Although the diagnosis of male infertility is mainly based on the seminogram, the diagnostic potential of this technique has been often considered as doubtful (Tomlinson, 2016). Therefore, the search for new molecular approaches that improve the actual clinical diagnosis have been pursued. Specifically, the seeking of differentially expressed sperm RNAs as biomarkers of infertility has been a topic of interest (Jodar et al., 2013). These assessments would be especially informative when facing patients with an intricate diagnosis like those whose seminal parameters are close to the threshold values (World Health Organization, 2010) or in cases of unexplained male infertility (infertile patients with discarded female factor, no physical, genetic or endocrine alterations, and normal semen parameters), in which current methods cannot provide a reliable diagnosis.

The described sets of *Ubiquitously Expressed* mRNAs and lncRNAs represent a host of molecules with a constitutive presence in the human sperm. The fact that these RNAs were detected in all analyzed samples raises the guaranties of successful detection in future prospective analyses. This condition is an essential requirement for any molecule suitable for being used as a biomarker. A constitutive expression of a molecule ensures that its measure can be used for detecting alterations associated to a certain condition (e.g. infertility), reducing the options of obtaining false positives due to a null expression level.

On this basis, the expression of the established *Ubiquitously Expressed*

mRNAs in our study could be analyzed in future comparative studies involving fertile and infertile populations, with the purpose of evaluating possible differential patterns and identify the best candidates of fertility biomarkers. In fact, several of these identified mRNAs have been previously suggested by other authors as biomarker candidates of different infertility profiles like asthenozoospermia (*BRD2*, *CABYR*, *CRISP2*, *PRM1* and *2*, and *TNP1*) (Aslani et al., 2011; Jedrzejczak et al., 2007; Jodar et al., 2012; Kempisty et al., 2007; Lambard et al., 2004; Pelloni et al., 2018; Zhou et al., 2015), oligozoospermia (*CREM*, *DDX3X*) (Montjean et al., 2012), teratozoospermia (*ODF1*, *PRM1* and *2*) (Platts et al., 2007; Savadi-Shiraz et al., 2015), non-obstructive azoospermia (*AKAP4*, *FAM71F1*, *GTSF1L*, *PRM1* and *2*, *SPATA3*, *UBQLN3*, and *ZMYND15*) (Aslani et al., 2011; Haraguchi et al., 2009; Hashemi et al., 2018; Malcher et al., 2013), and cryptorchidism (*ODF1*, *PRM1*, *TCP11*, *UBN1*, and *USP25*) (Nguyen et al., 2009).

Concerning lncRNAs, although other studies have identified some sperm lncRNAs with altered expression in asthenozoospermic patients (as HOTAIR, lnc32058, lnc09522 and lnc98487) (Zhang et al., 2015; Zhang et al., 2019), none of them were identified as *Ubiquitously Expressed* lncRNAs in our study. Therefore, our results provide a new set of lncRNAs with a consistent expression in fertile individuals, which could be evaluated in other infertile populations to determine their biomarker potential.

Additionally to single-transcripts biomarker approaches, the use of stable pairs of transcripts with a disrupted correlation in cases of different types of infertility represents a novel perspective in biomarker research, since it provides a wider insight of the multifactorial origins of infertility. As of today, only a few authors have set out a multiple-molecule biomarker strategy from expression data of sperm mRNAs (Aoki et al., 2006; Lalancette et al., 2009; Lima-Souza et al., 2012; Ni et al., 2014; Steger et al., 2008) and miRNAs (Corral-Vazquez et al., 2019; Salas-Huetos et al., 2014). In the present study, four pairs of mRNAs and 22 pairs of lncRNAs that maintain a stable expression correlation in fertile individuals were detected throughout the analyzed population. Although their establishment as potential biomarkers would still need an additional comparative study with infertile cohorts, these pairs of transcripts could suppose a reference point in the search for additional diagnostic tools.

Conclusion

The present study provides a detailed profiling of the large fraction of the sperm transcriptome of fertile men. This characterization has been obtained by state-of-the-art *omic* technologies, providing a high coverage and deep analysis of the studied samples. Our results demonstrate the presence of a complex population of sperm RNAs with a high degradation status, and a common fingerprint of transcripts that is preserved between different individuals. Among them, the set of coding transcripts showed a significant implication in major previous spermatogenesis processes and in reproduction. The existence of a common fingerprint of coding transcripts suggests a selective

preservation of functional mRNAs among sperm cells. Additionally, our results also have indicated that the lncRNAs cargo present in spermatozoa regulate a set of potential target genes involved not only in the regulation of spermatogenesis, but also in early embryo development. At last, the detection of *Ubiquitously Expressed* transcripts and pairs of transcripts with correlated expression suggests a potential use of these molecules in the fertility biomarker field. Altogether, these results offer an integrative overview of human sperm coding and non-coding transcripts and suppose an additional step towards the full understanding of their biological relevance and potential clinical application.

References

Aoki, V., Liu, L., Carrell, D., 2006. A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. Mol. Hum. Reprod. 12, 41–50.

Aslani, F., Modarresi, M.H., Soltanghoraee, H., Akhondi, M.M., Shabani, A., Lakpour, N., Sadeghi, M.R., 2011. Seminal molecular markers as a non-invasive diagnostic tool for the evaluation of spermatogenesis in non-obstructive azoospermia. Syst. Biol. Reprod. Med. 57, 190–6.

Boerke, A., Dieleman, S., Gadella, B., 2007. A possible role for sperm RNA in early embryo development. Theriogenology 68S, 147–155.

Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915–1927.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 1–19.

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.

Dadoune, J., Pawlak, A., Alfonsi, M., Siffroi, J., 2005. Identification of transcripts by macroarrays, RT-PCR and in situ hybridization in human ejaculate spermatozoa. Mol. Hum. Reprod. 11, 133–140.

Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjöstedt, E., Lundberg, E., Szigyarto, C.A.-K., Skogs, M., Takanen, J.O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlén, M., 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol. Cell. Proteomics 13, 397–406.

Fei, Z. hua, Yu, X. juan, Zhou, M., Su, H. fang, Zheng, Z., Xie, C. ying, 2016. Upregulated expression of long non-coding RNA LINC00982 regulates cell proliferation and its clinical relevance in patients with gastric cancer. Tumor Biol. 37, 1983–1993.

Fernandes, J., Acuña, S., Aoki, J., Floeter-Winter, L., Muxel, S., 2019. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and

Disease. Non-Coding RNA 5, 17.

Flegel, C., Vogel, F., Hofreuter, A., Wojcik, S., Schoeder, C., Kieć-Kononowicz, K., Brockmeyer, N.H., Müller, C.E., Becker, C., Altmüller, J., Hatt, H., Gisselmann, G., 2016. Characterization of non-olfactory GPCRs in human sperm with a focus on GPR18. Sci. Rep. 6, 1–10.

Goodrich, R.J., Johnson, G.D., Krawetz, S., 2007. The preparation of human spermatozoal RNA for clinical analysis. Arch. Androl. 53, 161–167.

Haraguchi, T., Ishikawa, T., Yamaguchi, K., Fujisawa, M., 2009. Cyclin and protamine as prognostic molecular marker for testicular sperm extraction in patients with azoospermia. Fertil. Steril. 91, 1424–1426.

Hashemi, M.S., Mozdarani, H., Ghaedi, K., Nasr-Esfahani, M.H., 2018. Expression of ZMYND15 in Testes of Azoospermic Men and Association With Sperm Retrieval. Urology 114, 99–104.

Herrada, G., Wolgemuth, D.J., 1997. The mouse transcription factor Stat4 is expressed in haploid male germ cells and is present in the perinuclear theca of spermatozoa. J. Cell Sci. 110, 1543–1553.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.

Jedrzejczak, P., Kempisty, B., Bryja, A., Mostowska, M., Depa-Martynow, M., Pawelczyk, L., Jagodzinski, 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.

Jodar, M., Kalko, S., Castillo, J., Ballescà, J.L., Oliva, R., 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Hum. Reprod. 27, 1431–1438.

Jodar, M., Selvaraju, S., Sendler, E., Diamond, M.P., Krawetz, S.A., 2013. The presence, role and clinical use of spermatozoal RNAs. Hum. Reprod. Update 19, 604–624.

Karlic, R., Ganesh, S., Franke, V., Svobodova, E., Urbanova, J., Suzuki, Y., Aoki, F., Vlahovicek, K., Svoboda, P., 2017. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. DNA Res. 24, 129–141.

Kempisty, B., Depa-Martynow, M., Lianeri, M., Jedrzejczak, P., Darul-Wasowicz, A., Jagodzinski, P., 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. Folia Histochem. Cytobiol. 45, 109–113.

Krawetz, S., 2005. Paternal contribution: new insights and future challenges. Nat. Rev. Genet. 6, 633–642.

Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., Krawetz, S.A., 2009. Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. 87, 735–748.

Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A., Carreau, S., 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: Relationship to sperm motility and capacitation. Mol. Hum. Reprod. 10, 535–541.

Lima-Souza, A., Anton, E., Mao, S., Ho, W.J., Krawetz, S.A., 2012. A platform for evaluating sperm RNA biomarkers: Dysplasia of the fibrous sheath -

Testing the concept. Fertil. Steril. 97, 1061-1066.e3.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. PLOS Comput. Biol. 13, e1005457.

Malcher, A., Rozwadowska, N., Stokowy, T., Kolanowski, T., Jedrzejczak, P., Zietkowiak, W., Kurpisz, M., 2013. Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. Fertil. Steril. 100, 1686-1694.e7.

Mao, S., Goodrich, R.J., Hauser, R., Schrader, S.M., Chen, Z., Krawetz, S.A., 2013. Evaluation of the effectiveness of semen storage and sperm purification methods for spermatozoa transcript profiling. Syst. Biol. Reprod. Med. 59, 287–295.

Mao, S., Sendler, E., Goodrich, R.J., Hauser, R., Krawetz, S.A., 2014. A Comparison of Sperm RNA-Seq Methods. Syst Biol Reprod Med 60, 308–315.

Miki, K., Willis, W.D., Brown, P.R., Goulding, E.H., Fulcher, K.D., Eddy, E.M., 2002. Targeted Disruption of the Akap4 Gene Causes Defects in Sperm Flagellum and Motility. Dev. Biol. 248, 331–342.

Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R., Krawetz, S.A., 1999. A complex population of RNAs exists in human ejaculate spermatozoa: Implications for understanding molecular aspects of spermiogenesis. Gene 237, 385–392.

Modi, D., Shah, C., Sachdeva, G., Gadkar, S., Bhartiya, D., Puri, C., 2005. Ontogeny and cellular localization of SRY transcripts in the human testes and its detection in spermatozoa. Reproduction 130, 603–613.

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.

Nguyen, M.T., Delaney, D.P., Kolon, T.F., 2009. Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. Fertil. Steril. 92, 182–187.

Ni, K., Steger, K., Yang, H., Wang, H., Hu, K., Chen, B., 2014. Sperm Protamine mRNA Ratio and DNA Fragmentation Index Represent Reliable Clinical Biomarkers for Men with Varicocele after Microsurgical Varicocele Ligation. J. Urol. 192, 170–176.

Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., Pera, R.A.R., 2012. Human pre-implantation embryo development. Development 139, 829–841.

Ostermeier, G., Goodrich, R., Moldenhauer, J., Diamond, M., Krawetz, S., 2005. A suite of novel human spermatozoal RNAs. J. Androl. 26, 70–74.

Ostermeier, G., Miller, D., Huntriss, J., Diamond, M., Krawetz, S., 2004. Delivering spermatozoan RNA to the oocyte. Nature 429, 154.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., Krawetz, S.A., 2002. Spermatozoal RNA profiles of normal fertile men. Mech. Dis. 360, 772–777.

Ostermeier, G.C., Goodrich, R.J., Diamond, M.P., Dix, D.J., Krawetz, S.A., 2005. Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. Fertil. Steril. 83, 1687–1694.

Pelloni, M., Paoli, D., Majoli, M., Pallotti, F., Carlini, T., Lenzi, A., Lombardo, F., 2018. Molecular study of human sperm RNA: Ropporin and CABYR in asthenozoospermia. J. Endocrinol. Invest. 41, 781–787.

Pessot, C., Brito, M., Figueroa, J., Concha, I., Yanez, A., Burzio, L., 1989. Presence of RNA in the sperm nucleus. Biochem. Biophys. Res. Commun. 158, 272–278.

Platts, A.E., Dix, D.J., Chemes, H.E.,

Thompson, K.E., Goodrich, R., Rockett, J.C., Rawe, V.Y., Quintana, S., Diamond, M.P., Strader, L.F., Krawetz, S.A., 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum. Mol. Genet. 16, 763–773.

Ponting, C.P., Oliver, P.L., Reik, W., 2009. Evolution and Functions of Long Noncoding RNAs. Cell 136, 629–641.

Qiu, J., Ren, Z., Yan, J., 2016. Identification and functional analysis of long non-coding RNAs in human and mouse early embryos based on single-cell transcriptome data. Oncotarget 7.

Quinn, J.J., Chang, H.Y., 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17, 47–62.

R Core Team, 2014. R: A language and environment for statistical computing. [WWW Document]. R Found. Stat. Comput. Vienna, Austria.

RStudio Team, 2015. RStudio: Integrated Development for R.

Salas-Huetos, A., Blanco, J., Vidal, F., Mercader, J.M., Garrido, N., Anton, E., 2014. New insights into the expression profile and function of micro-ribonucleic acid in human spermatozoa. Fertil. Steril. 102, 213–222.

Saunders, C., Swann, K., Lai, F., 2007. PLCzeta, a sperm-specific PLC and its potential role in fertilization. Biochem. Soc. Simp. 74, 23–36.

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

Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., Krawetz, S.A., 2013. Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. 41, 4104–4117.

Steger, K., Wilhelm, J., Konrad, L., Stalf, T., Greb, R., Diemer, T., Kliesch, S., Bergmann, M., Weidner, W., 2008. 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. 23, 11–16.

Supek, F., Bosnjak, M., Skunca, N., Smuc, T., 2011. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One 6.

T.Moon, R., Brown, J.D., Torres, M., 1997. WNTs modulate cell fate and behavior during vertebrate development. Trends Genet. 13, 157–162.

Tomlinson, M.J., 2016. Uncertainty of measurement and clinical value of semen analysis: has standardisation through professional guidelines helped or hindered progress? Andrology 4, 763–770.

Wang, H., Zhou, Z., Xu, M., Li, J., Xiao, J., Xu, Z.Y., Sha, J., 2004. A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. J. Mol. Med. 82, 317–324.

Wang, Liguo, Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J.E., Dasari, S., Vedell, P.T., Barman, P., Wang, Liewei, Weinshiboum, R., Jen, J., Huang, H., Kohli, M., Kocher, J.P.A., 2016. Measure transcript integrity using RNA-seq data. BMC Bioinformatics 17, 1–16.

Wanowska, E., Kubiak, M.R., Rosikiewicz, W., Makałowska, I., Szcześniak, M.W., 2018. Natural antisense transcripts in diseases: From modes of action to targeted therapies. Wiley Interdiscip.

Rev. RNA 9, 1–16.

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W., Zhang, J., Bi, X., Dai, J., Zhang, Q., Lu, Z.J., Gao, G., 2016. Critical roles of long noncoding RNAs in Drosophila Spermatogenesis. Genome Res. 26, 1233–1244.

Wichman, L., Somasundaram, S., Breindel, C., Valerio, D.M., McCarrey, J.R., Hodges, C.A., Khalil, A.M., 2017. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. Biol. Reprod. 97, 313–323.

World Health Organization, 2010. WHO laboratory manual for the examination and processing of human semen, 5th ed. ed. Geneva: World Health Organization.

Xu, B., Hao, Z., Jha, K., Zhang, Z., Urekar, C., Digilio, L., Pulido, S., Strauss, J. 3rd, Flickinger, C., Herr, J., 2008. Targeted Deletion of Tssk1 & 2 Causes Male Infertility Due to Haploinsufficiency Bingfang. Dev. Biol. 319, 211–222.

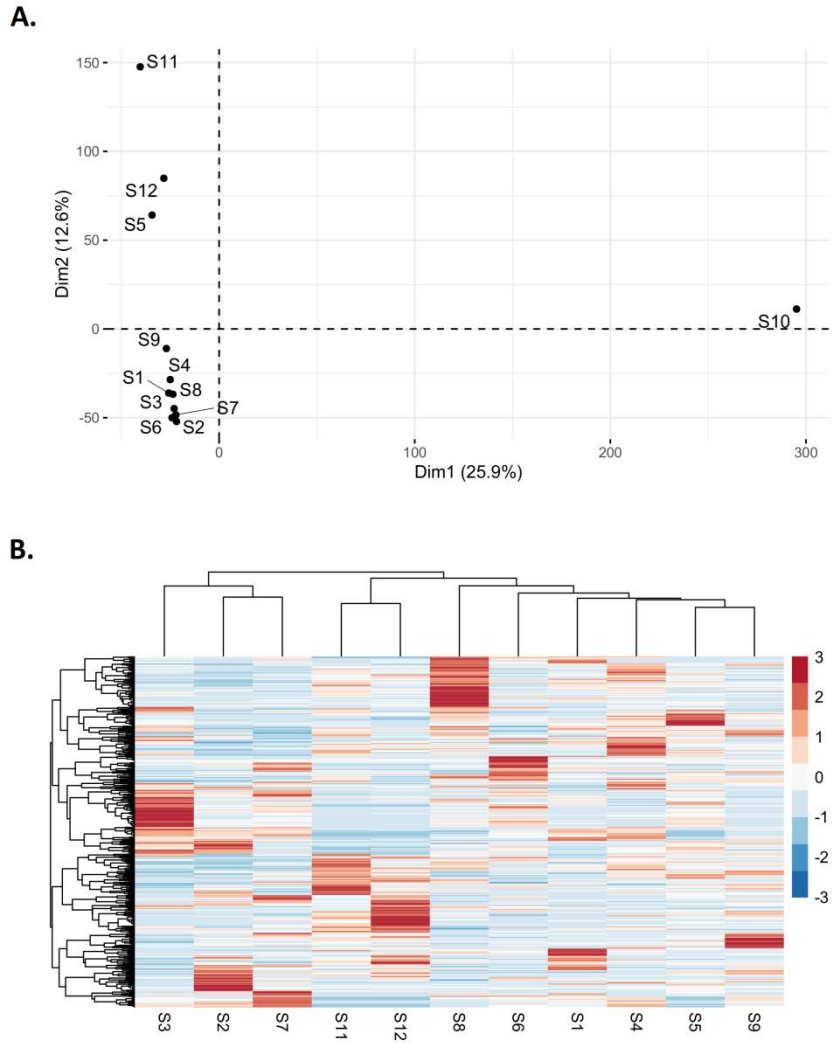
Zhang, C., Li, X.Y., Luo, Z.Z., Wu, T.W., Hu, H., 2019. Upregulation of LINC00982 inhibits cell proliferation and promotes cell apoptosis by regulating the activity of PI3K/AKT signaling pathway in renal cancer. Eur. Rev. Med. Pharmacol. Sci. 23, 1443–1450.

Zhang, L., Liu, Z., Li, X., Zhang, P., Wang, J., Zhu, D., Chen, X., Ye, L., 2015. Low long non-coding RNA HOTAIR expression is associated with down-regulation of Nrf2 in the spermatozoa of patients with asthenozoospermia or oligoasthenozoospermia. Int. J. Clin. Exp. Pathol. 8, 14198–14205.

Zhang, X., Zhang, P., Song, D., Xiong, S., Zhang, H., Fu, J., Gao, F., Chen, H., Zeng, X., 2019. Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm. Biol. Reprod. 100, 982–993.

Zhou, J.-H., Zhou, Q.Z., Lyu, X.-M., Zhu, T., Chen, Z.-J., Chen, M.-K., Xia, H., Wang, C.-Y., Qi, T., Li, X., Liu, C.-D., 2015. The expression of cysteine-rich secretory protein 2 (CRISP2) and its specific regulator miR-27b in the spermatozoa of patients with asthenozoospermia. Biol Reprod 92, 1–9.

Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S., Li, Z., 2016. Dynamics of the Transcriptome during Human Spermatogenesis: Predicting the Potential Key Genes Regulating Male Gametes Generation. Sci. Rep. 6, 19069.



Supplemental Figure 1. Sample homogeneity analyses.

A) Principal Component Analysis (PCA) of the gene expression profile of the 12 analyzed samples (S1–S12).

B) Heatmap clustering of the samples S1-S9 and S11-S12 (S10 has been excluded as an outlier due to PCA results) concerning the 500 genes with higher Fragment Per Kilobase (FPKM) mean value.

Patient ID	Age	Semen parameters				RNA sample parameters			
		Ejaculate volume (mL)	Sperm motility (%PR+NP)	Sperm morphology (%NF)	Sperm concentration (10^6 spz/mL)	Millions of spz	RNA/spz (fg)	Quantity (ng)	Purity (260/280 nm)
S1	24	2.5	60	4	58.0	19	11.79	224	1.88
S2	28	4.2	42	4	63.6	28	17.33	487	1.81
S3	28	5.0	77	10	110.0	18	7.10	128	2.26
S4	19	1.8	59	4	73.3	17	16.57	282	2.19
S5	23	2.8	70	5	76.0	24	2.78	65	1.67
S6	30	4.3	39	6	53.0	15	7.73	116	1.71
S7	31	4.0	61	4	110.0	16	10.56	170	1.65
S8	23	3.0	68	8	112.5	20	6.13	123	1.43
S9	19	3.9	74	4	50.6	24	5.21	123	1.55
S10	24	2.0	62	5	71.0	15	181.24	2643	1.82
S11	20	2.6	60	4	50.3	43	3.44	148	1.75
S12	20	4.2	63	7	52.0	19	4.99	97	1.85

Supplemental Table 1. Demographic data, semen parameters, and sperm RNA measures of the analyzed population.

Gene name	Primer characteristics		Amplicon characteristics		
	Sequence (5' - 3')	Annealing temperature (°C)	cDNA length (bp)	DNA length (bp)	Chromosome location
PRM1	For: CAGAGTTCCACCTGCTCACA	62	331	422	chr16:11,280,901
	Rev: GGATGGTGGCATTTTCAAGA				-11,281,322
GAPDH	For: CGACCACTTTGTCAAGCTCA	64	228	332	chr12:6,537,970
	Rev: AGGGGTCTACATGGCAACTG				6,538,301
CD45	For: CCTTGAACCCGAACATGAGT	60	179	3307	chr1:198,709,727
	Rev: ATCTTTGAGGGGGATTCCAG				198,713,033
KIT	For: CCCAACACAACTTCCTTATGATC	59.8	113	393	chr4:54727484
	Rev: CATAAGCAGTTGCCTCAACAAC				54727876
CDH1	For: GAACCTCTGTGATGGAGGTCAC	61	269	544	chr16:68811684
	Rev: GTCAGTGACTGTGATCACAGCTG				68812227

Supplemental Table 2. Description of the primers used in the RT-PCR quality controls performed in all RNA samples.

Bp = base pairs; For = forward; Rev = reverse

Supplemental Table 3. List of the *Highly Stable* mRNAs and lncRNAs (i.e.: 200 mRNAs and 200 lncRNAs with the highest average Transcript Integrity Number or TIN).

mRNA	TIN	mRNA	TIN	lncRNA	TIN	lncRNA	TIN
PRM2	83.75	AL162231.1	38.11	AC133550.1	6.96	AC012615.6	1.47
TNP1	80.46	CLPB	38.09	AC027796.4	6.31	AC018816.1	1.45
TSSK6	59.57	PLCD3	38.05	AC104964.2	5.93	Z98257.1	1.44
GPX4	57.69	CCNY	37.98	PRDM16-DT	5.86	ARHGAP22-IT1	1.43
RBBP6	56.64	LRCH4	37.94	AP000346.1	5.81	SSTR5-AS1	1.43
ACAP1	54.84	AKAP4	37.92	AL078623.1	5.33	AC007663.2	1.43
ZNF513	54.20	FAM153A	37.88	AL353770.3	5.18	AC012531.1	1.41
CRISP2	52.69	VCX3B	37.81	AC012363.2	4.89	AC011472.2	1.40
SPEM2	52.36	TRIM36	37.81	AC091053.2	4.62	AC026894.1	1.39
PRR30	51.04	PRR7	37.81	AL627171.2	4.34	AC015908.2	1.39
DNAJC4	50.41	FBXW5	37.79	AC090023.2	3.92	LINC01356	1.39
UBAP2	49.56	WDR74	37.78	LINC02425	3.79	AC016738.2	1.38
TMEM191B	49.39	ST6GALNAC2	37.76	AC006064.4	3.64	ADAMTS9-AS2	1.35
CCDC188	49.27	UBXN6	37.74	AC104964.1	3.40	AC015853.1	1.35
ZC3H18	48.94	CKB	37.69	AC124301.1	3.05	ZNF426-DT	1.35
EGR3	48.20	USP25	37.66	UBOX5-AS1	2.96	LINC00298	1.34
MTPAP	48.07	UBAP2L	37.60	AP001160.1	2.92	AC015853.2	1.33
RSRC2	47.89	CCDC136	37.59	AC113189.1	2.84	AC092725.1	1.33
TMEM191C	47.83	CALCOCO2	37.57	AC124301.2	2.83	AC018767.3	1.33
NSUN4	47.31	KIAA1324	37.49	AC068385.1	2.79	AC015908.7	1.32
KRT15	47.25	PTOV1	37.39	LINC01141	2.76	AC011124.2	1.32
TEX54	47.24	HYOU1	37.29	AC013652.1	2.70	HDAC11-AS1	1.31
GOLGA6L2	47.19	FBRS	37.18	AL590989.1	2.70	CACNA1C-IT3	1.29
GIGYF2	46.50	CCDC196	37.17	AC011498.2	2.68	APTR	1.28
ACSBG2	46.15	ACE	37.09	AC139256.2	2.68	FAM230D	1.28
H3F3B	46.11	REXO1	37.08	AC098850.4	2.64	AP001005.3	1.28
JUND	46.03	TCP11	37.05	AC008763.1	2.61	AFAP1-AS1	1.27
TXNDC2	45.74	YWHAZ	36.85	AC009041.1	2.58	AC092650.1	1.26
FKBP8	45.73	ETNK2	36.81	Z97987.1	2.55	AC009158.1	1.26
DGCR6L	45.73	NFIC	36.81	AL136531.3	2.54	AC004702.1	1.26
CITED4	45.04	MAP4	36.80	LINC00905	2.54	LINC00654	1.25
PHF1	44.88	GRINA	36.76	AC004471.2	2.51	FBXO3-DT	1.25
UBC	44.84	MAP7D1	36.59	AC011491.3	2.47	RRN3P2	1.24
PHOSPHO1	44.50	CCDC91	36.55	AC012615.2	2.42	HMGA2-AS1	1.24
GLUL	44.35	VCY1B	36.53	AP000769.2	2.39	LAMTOR5-AS1	1.23
ANKRD9	44.32	ACRBP	36.53	AL034417.2	2.38	AC009623.1	1.23
DDX5	44.21	YTHDC1	36.46	JARID2-AS1	2.37	AL513365.2	1.23
CLK3	44.06	PPDPF	36.34	AC010300.1	2.37	AC022382.1	1.23
LMNB2	43.39	MOSPD3	36.32	AC073349.1	2.31	AC022762.2	1.22
TPPP2	43.36	DNAAF3	36.30	AL158175.1	2.25	AP001172.1	1.22
LMNTD2	43.26	AJM1	36.22	AL022341.2	2.20	SPATA42	1.21
MAPK8IP3	43.22	AC011448.1	36.10	AC025580.2	2.16	AC022762.1	1.21

CARHSP1	43.16	DDX3X	36.10	AL137798.2	2.16	MIR7-3HG	1.21
GOLGA6L10	43.00	PROCA1	35.85	AL590729.1	2.11	AL136301.1	1.20
RERE	42.95	GOLGA8A	35.82	AC026620.1	2.11	AC011444.2	1.19
MEX3D	42.89	TRIM17	35.75	LINC02543	2.10	AC092422.1	1.17
SRRM5	42.85	SPZ1	35.73	AF131216.4	2.07	Z99916.1	1.17
TP53INP2	42.72	HDAC11	35.72	AC004528.2	2.07	LINC01128	1.16
LENG8	42.24	NDUFA13	35.71	AC082651.4	2.06	AL031428.1	1.15
NUMBL	42.19	PCYT2	35.70	AC016292.1	2.05	LINC01956	1.15
PKM	41.99	NOLC1	35.67	AC104581.3	2.01	AC004951.4	1.15
FAM229A	41.79	DBN1	35.64	AC025419.1	2.01	AL049795.2	1.13
PCSK4	41.78	FAM153B	35.61	AL022324.4	1.99	AC004231.1	1.12
SPATA18	41.77	SPAG4	35.59	AC009065.6	1.96	AC017116.1	1.12
SORBS3	41.66	SLFNL1	35.55	AC115284.1	1.96	AC004147.4	1.12
DNAJA4	41.57	UBA52	35.45	AC008742.1	1.95	AC099684.1	1.11
TEX44	41.53	GOLGA6D	35.34	AC078778.1	1.95	LINC01722	1.11
KIF1C	41.49	EEF1D	35.29	FP671120.2	1.95	AP001453.3	1.11
SPATA3	41.46	NFKBIB	35.21	AP001350.1	1.92	AC022816.1	1.11
HMGB4	41.26	ELOA2	35.20	AC004771.1	1.91	SDCBP2-AS1	1.10
SLC22A18	41.19	CLEC16A	35.11	Z98259.1	1.90	LINC01339	1.10
HPCA	41.18	PABPN1	35.09	AL160237.1	1.88	AL591178.1	1.10
FNDC11	41.15	PHF7	35.08	AC074135.1	1.88	AC012531.2	1.09
BAG1	41.09	SIRT2	35.06	AL392086.1	1.87	AC009093.2	1.09
CFL1	41.04	YJEFN3	34.97	AC005253.1	1.85	AL512598.2	1.08
ZSWIM9	41.03	SH3RF2	34.94	AC022098.1	1.80	AC209154.1	1.08
NRBP1	40.98	RAB3IP	34.89	AC017000.1	1.79	AC002472.1	1.07
RARA	40.96	PCBP4	34.88	LINC02079	1.77	FAM230C	1.07
CSNK1G2	40.83	SRRM1	34.87	AL592166.1	1.77	SNHG29	1.07
PTK7	40.78	HNRNPU	34.77	C20orf78	1.76	TMEM220-AS1	1.05
EPB41L3	40.54	GPI	34.75	LINC00115	1.74	AC135983.1	1.05
LUZP1	40.50	GORASP1	34.63	AL669831.5	1.74	AC004448.2	1.05
BRD2	40.47	QRICH1	34.57	LINC01566	1.74	AP002754.1	1.04
GTF2F1	40.33	TESK1	34.52	AC010401.1	1.74	LINC01135	1.04
UBB	40.32	RNF38	34.52	AC008543.1	1.72	LINC02367	1.04
LPIN1	40.28	BRD9	34.40	AL353616.2	1.71	AC006299.1	1.04
AKAP1	40.27	ISG20L2	34.36	MIR193BHG	1.69	AC109492.1	1.04
VRK3	40.26	ESPN	34.33	AP4B1-AS1	1.68	AC068533.3	1.03
PRM3	40.26	ZNF683	34.22	FAM230F	1.68	AC005329.3	1.03
CDIP1	40.10	TRIM28	34.18	AC020663.1	1.66	MIR924HG	1.02
HGS	39.93	MEX3C	34.13	AC006504.5	1.66	MYHAS	1.02
C10orf62	39.77	TPPP3	34.09	AL079307.1	1.65	AC021351.1	1.02
ODF2	39.65	MEX3B	34.07	AC240565.1	1.65	AC009093.10	1.02
C1orf194	39.63	CDHR2	34.05	AC024575.1	1.65	Z98259.3	1.01
TTC7A	39.63	H1FNT	34.05	AL590096.1	1.63	KF456478.1	1.01
STK32C	39.52	STOML2	33.96	TTC39C-AS1	1.62	AC120045.1	1.00
REEP6	39.12	CNN1	33.95	AC068707.1	1.61	AC015687.1	1.00
PDZD8	38.98	EHD1	33.94	AL590556.3	1.60	LINC01357	1.00
CRIP2	38.70	FBXO24	33.92	AC015908.3	1.58	AL591684.2	1.00
CCDC9	38.66	ODF3B	33.90	AC073957.1	1.57	AC051619.4	1.00
ZNF706	38.63	FAM53C	33.90	AP003721.3	1.54	AL022323.4	0.99

RANGAP1	38.63	NUMA1	33.88	LINC01002	1.54	AC010880.1	0.99
PDXDC1	38.61	SOCS1	33.87	AC091806.1	1.51	PRICKLE2-AS1	0.99
TLE4	38.61	TSSK3	33.84	AC020743.2	1.51	LINC02130	0.98
MICOS13	38.56	PRKCZ	33.74	TMEM161B-AS1	1.51	AP001198.1	0.98
SF1	38.51	DYRK1B	33.68	AC138649.1	1.50	LINC01137	0.98
EIF4G1	38.40	IGSF9	33.58	CCDC144NL-AS1	1.50	AC007014.1	0.98
RNF44	38.37	JUP	33.56	AC005899.8	1.49	LINC01005	0.97
HNRNPH1	38.27	ZMIZ2	33.51	AL139424.1	1.49	AC025279.1	0.97
CEMP1	38.12	CHD4	33.50	AL079307.2	1.47	AC005324.1	0.97

Appendix II

Materials and methods

7.1. Study populations

Semen samples from 83 individuals were employed for the consecution of the objectives of this doctoral thesis. Samples were obtained from a single ejaculate collected after 3-7 days of sexual abstinence in sterile containers, and cryopreserved with EggYolk Freezing Medium (IRVINE) at 1:1 proportion. Inclusion criteria of the individuals comprised normal karyotype, no previous exposure to any genotoxic agent, and no history of chemotherapy, radiotherapy or chronic illness.

Semen parameters were classified according to WHO and Kruger criteria (Kruger et al., 1988; World Health Organization, 2010) that establish as normality thresholds: total sperm number of $\geq 39 \cdot 10^6$, percentage of motile sperm of ≥40%, and percentage of normal forms of ≥4% (WHO criteria) or ≥14% (Kruger criteria).

Collection of the semen samples needed for the consecution of Publications 1 and 2 was done as a part of a previous research project aimed to describe the miRNA profile of human spermatozoa (Salas-Huetos et al., 2016, 2015, 2014). They comprised a group of fertile donors with normal semen parameters (n=10, S1-S10), and four groups of infertile patients: asthenozoospermic (n=10, S11-S20), teratozoospermic (n=10, S21-S30), oligozoospermic (n=10, S31-S40), and normozoospermic or UMI (n=8, S41-S48). The demographic data and seminal parameters of these patients are summarized in Table 8.

Table 8. Age and seminal parameters of the samples included in Publications 1 and 2.

Patient	Age	Profile	Sperm concentration (10^6 spz/mL)	Motility (%PR + %NP)	Morphology (%NF)
S1	28	Fertile	110.0	52	7 [(1)]
S2	20	Fertile	59.3	47	7 [(1)]
S3	22	Fertile	87.0	50	>4 [(1)]
S4	25	Fertile	50.0	59	>4 [(1)]
S5	30	Fertile	60.0	31	12 [(1)]
S6	33	Fertile	30.0	65	8 [(1)]
S7	30	Fertile	95.0	53	18 [(1)]
S8	29	Fertile	32.3	56	4 [(1)]
S9	24	Fertile	103.0	44	6 [(1)]

S10	24	Fertile	45.0	55	4 [(1)]
S11	32	Infertile (A)	126.0	35	12 [(1)]
S12	36	Infertile (A)	98.0	35	10 [(1)]
S13	31	Infertile (A)	65.3	40	16 [(2)]
S14	41	Infertile (A)	99.4	35	5 [(1)]
S15	35	Infertile (A)	60.3	35	13 [(1)]
S16	41	Infertile (A)	104.0	40	14 [(2)]
S17	42	Infertile (A)	22.3	24	7 [(1)]
S18	49	Infertile (A)	43.0	22	7 [(1)]
S19	44	Infertile (A)	69.0	31	5 [(1)]
S20	39	Infertile (A)	78.0	29	7 [(1)]
S21	33	Infertile (T)	133.0	73	1 [(1)]
S22	35	Infertile (T)	38.0	49	3 [(1)]
S23	48	Infertile (T)	85.0	41	3 [(1)]
S24	34	Infertile (T)	66.6	54	3 [(1)]
S25	36	Infertile (T)	160.0	46	3 [(1)]
S26	42	Infertile (T)	90.0	56	3 [(1)]
S27	41	Infertile (T)	28.0	51	2 [(1)]
S28	35	Infertile (T)	55.2	45	10 [(2)]
S29	28	Infertile (T)	21.2	45	6 [(2)]
S30	32	Infertile (T)	47.2	60	10 [(2)]
S31	37	Infertile (O)	13.0	43	4 [(1)]
S32	35	Infertile (O)	13.0	64	5 [(1)]
S33	38	Infertile (O)	13.3	36	4 [(1)]
S34	50	Infertile (O)	11.3	54	5 [(1)]
S35	43	Infertile (O)	13.3	43	5 [(1)]
S36	39	Infertile (O)	11.0	57	4 [(1)]
S37	45	Infertile (O)	5.3	40	4 [(1)]
S38	44	Infertile (O)	7.6	48	5 [(1)]
S39	49	Infertile (O)	7.0	53	4 [(1)]
S40	41	Infertile (O)	15.0	48	7 [(1)]
S41	34	Infertile (N)	54.0	60	13 [(1)]
S42	37	Infertile (N)	142.0	50	13 [(1)]
S43	45	Infertile (N)	131.0	60	17 [(1)]
S44	43	Infertile (N)	74.3	60	58 [(1)]
S45	39	Infertile (N)	68.8	50	15 [(1)]
S46	45	Infertile (N)	78.4	50	15 [(1)]
S47	40	Infertile (N)	66.3	50	13 [(1)]
S48	33	Infertile (N)	121.0	50	30 [(1)]

Spz = spermatozoa; PR = Progressive; NP = Non Progressive; NF = Normal Forms; A = Asthenozoospermia; T = Teratozoospermia; O = Oligozoospermia; N = Normozoospermia.

[(1)] Normal morphology: WHO 2010

[(2)] Normal morphology: Kruger 1988

Publication 2 also included a validation phase of the results in an additional cohort of patients. For that purpose, semen samples from nine fertile normozoospermic donors (V1-V9) and 14 infertile patients with different seminal alterations (V10-V23) were analyzed. Demographic data and seminal parameters of these individuals are summarized in Table 9.

Table 9. Age and seminal parameters of the samples included in the validation assay performed in Publication 2.

Patient	Age	Profile	Sperm concentration (10^6 spz/mL)	Motility (%PR + %NP)	Morphology (%NF) [(1)]
V1	27	Fertile	220.0	53	7
V2	32	Fertile	62.0	55	17
V3	22	Fertile	45.0	>40	20
V4	31	Fertile	69.6	66	7
V5	19	Fertile	85.0	51	>4
V6	21	Fertile	63.0	46	>4
V7	22	Fertile	75.0	58	7
V8	24	Fertile	53.0	>40	>4
V9	26	Fertile	74.0	45	>4
V10	N/A	Infertile (T)	73.5	75	2
V11	N/A	Infertile (OA)	7.0	36	>4
V12	37	Infertile (AT)	46.9	30	3
V13	29	Infertile (A)	18.4	25	18
V14	N/A	Infertile (OA)	14.2	39	4
V15	32	Infertile (O)	14.1	45	18
V16	N/A	Infertile (T)	285.7	85	3
V17	37	Infertile (T)	78.4	60	7
V18	N/A	Infertile (T)	15.2	79	3
V19	N/A	Infertile (A)	447.2	38	5
V20	N/A	Infertile (OT)	6.1	44	1
V21	47	Infertile (AT)	156.8	35	6
V22	42	Infertile (T)	36.7	50	4
V23	N/A	Infertile (OT)	10.7	62	3

Spz = spermatozoa; PR = Progressive; NP = Non Progressive; NF = Normal Forms; N/A = Not Available; A = Asthenozoospermia; T = Teratozoospermia; O = Oligozoospermia; OA = Oligoasthenozoospermia; AT = Asthenoteratozoospermia; OT = Oligoteratozoospermia.

[(1)] Morphology criteria: WHO 2010

The study of Appendix I was carried on in semen samples from a population of 12 fertile normozoospermic donors (P1-P12). Demographic data and seminal parameters of these individuals are summarized in Table 10.

Table 10. Age and seminal parameters of the samples included in Appendix I.

Patient	Age	Profile	Sperm concentration (10^6 spz/mL)	Motility (%PR + %NP)	Morphology (%NF) [(1)]
P1	24	Fertile	58.0	60	4
P2	28	Fertile	63.6	42	4
P3	28	Fertile	110.0	77	10
P4	19	Fertile	73.3	59	4
P5	23	Fertile	76.0	70	5
P6	30	Fertile	53.0	39	6
P7	31	Fertile	110.0	61	4
P8	23	Fertile	112.5	68	8
P9	19	Fertile	50.6	74	4
P10	24	Fertile	71.0	62	5
P11	20	Fertile	50.3	60	4
P12	20	Fertile	52.0	63	7

Spz = spermatozoa; PR = Progressive; NP = Non Progressive; NF = Normal Forms.

[(1)] Morphology criteria: WHO 2010

7.2. Experimental protocols

7.2.1. Sperm purification by Somatic Cell Lysis Buffer

As semen samples were initially stored in liquid nitrogen at -256ºC, an initial thawing step was followed to start the processing. Samples were left at room temperature for 5 minutes, followed by an incubation of 5 minutes at 37ºC.

In order to isolate the sperm cells of the samples, the SCLB method (adapted from Goodrich et al., 2007) was employed for achieve the lysis of the non-sperm cell fraction contained in the semen.

Composition of SCLB

- 0.1% of Sodium dodecyl sulfate
- 0.5% Triton X
- Diethylpyrocarbonate water

Protocol

1. Transfer the thawed samples to sterile tubes (15 mL of capacity).
2. Centrifuge at 2500 g for 10 minutes and discard the supernatant.
3. Resuspend the pellet with 1 mL phosphate buffered saline (PBS).
4. Add 1 mL of SCLB and mix by pipetting up and down.
5. Adjust final volume to 13 mL with SCLB and mix by vortex.
6. Incubate on ice for 30 minutes.
7. After 20 minutes of incubation, perform sperm count in a Neubauer chamber to check the absence of somatic cells. If somatic cell ratio over spermatozoa is under 1:10000, proceed to the next step. If the ratio is higher, centrifuge the sample at 1200 g for 10 minutes at 4ºC. Remove supernatant and repeat the protocol from Step 3.
8. Centrifuge the sample at 1200 g for 10 minutes at 4ºC.
9. Remove supernatant.
10. The purified sperm fraction will be concentrated at the bottom of the tube (white pellet).

7.2.2. Sperm RNA extraction

During the present project, three different protocols of sperm RNA extraction have been used:

1) TRIzol® (Thermo Fisher Scientific)

2) mirVana™ PARIS™ RNA and Native Protein Purification Kit (Thermo Fisher Scientific)

3) Maxwell® RSC simplyRNA Cells Kit (Promega)

Table 11 summarizes the parameters of RNA quality and quantity of the obtained samples and the method used in each case.

Table 11. Quantification, purity and extraction method of all sperm RNA samples.

Patient	Work	Profile	Total RNA (ng)	RNA/cell (fg)	RNA purity (260/280)	Extraction method
S1	P1, P2	Fertile	3697	64.00	1.92	TRIzol
S2	P1, P2	Fertile	1229	63.00	1.94	TRIzol
S3	P1, P2	Fertile	828	48.00	1.93	TRIzol
S4	P1, P2	Fertile	696	85.00	1.84	TRIzol
S5	P1, P2	Fertile	737	99.00	1.85	TRIzol
S6	P1, P2	Fertile	712	54.00	1.84	TRIzol
S7	P1, P2	Fertile	402	37.00	1.84	TRIzol
S8	P1, P2	Fertile	458	26.00	1.97	TRIzol
S9	P1, P2	Fertile	428	38.00	1.85	TRIzol
S10	P1, P2	Fertile	373	42.00	1.96	TRIzol
S11	P1, P2	Infertile (A)	1225	47.00	1.83	TRIzol
S12	P1, P2	Infertile (A)	477	11.00	1.79	TRIzol
S13	P1, P2	Infertile (A)	895	11.00	1.77	TRIzol
S14	P1, P2	Infertile (A)	207	10.00	1.62	TRIzol
S15	P1, P2	Infertile (A)	1103	42.00	1.77	TRIzol
S16	P1, P2	Infertile (A)	2066	27.00	1.78	TRIzol
S17	P1, P2	Infertile (A)	2152	324.00	1.80	TRIzol
S18	P1, P2	Infertile (A)	5313	304.00	1.83	TRIzol
S19	P1, P2	Infertile (A)	392	133.00	1.72	TRIzol
S20	P1, P2	Infertile (A)	9532	357.00	1.80	TRIzol
S21	P1, P2	Infertile (T)	497	9.00	1.74	TRIzol
S22	P1, P2	Infertile (T)	765	68.00	1.75	TRIzol
S23	P1, P2	Infertile (T)	3256	52.00	1.84	TRIzol
S24	P1, P2	Infertile (T)	3085	81.00	1.80	TRIzol
S25	P1, P2	Infertile (T)	1924	22.00	1.81	TRIzol
S26	P1, P2	Infertile (T)	1923	35.00	1.75	TRIzol
S27	P1, P2	Infertile (T)	944	73.00	1.75	TRIzol

S28	P1, P2	Infertile (T)	3165	223.00	1.77	TRIzol
S29	P1, P2	Infertile (T)	414	20.00	1.60	TRIzol
S30	P1, P2	Infertile (T)	769	23.00	1.80	TRIzol
S31	P1, P2	Infertile (O)	352	50.00	1.67	TRIzol
S32	P1, P2	Infertile (O)	463	165.00	1.61	TRIzol
S33	P1, P2	Infertile (O)	811	101.00	1.74	TRIzol
S34	P1, P2	Infertile (O)	698	194.00	1.64	TRIzol
S35	P1, P2	Infertile (O)	382	36.00	1.77	TRIzol
S36	P1, P2	Infertile (O)	446	57.00	1.70	TRIzol
S37	P1, P2	Infertile (O)	482	56.00	1.75	TRIzol
S38	P1, P2	Infertile (O)	493	112.00	1.36	TRIzol
S39	P1, P2	Infertile (O)	586	127.00	0.93	TRIzol
S40	P1, P2	Infertile (O)	1260	350.00	1.43	TRIzol
S41	P1, P2	Infertile (N)	937	18.00	1.72	TRIzol
S42	P1, P2	Infertile (N)	2540	55.00	1.74	TRIzol
S43	P1, P2	Infertile (N)	2868	35.00	1.73	TRIzol
S44	P1, P2	Infertile (N)	311	6.00	1.77	TRIzol
S45	P1, P2	Infertile (N)	3632	10.00	1.74	TRIzol
S46	P1, P2	Infertile (N)	660	12.00	1.60	TRIzol
S47	P1, P2	Infertile (N)	2213	73.00	1.73	TRIzol
S48	P1, P2	Infertile (N)	1310	19.00	1.72	TRIzol
V1	P2	Fertile	279.00	50.73	1.72	TRIzol
V2	P2	Fertile	253.00	21.90	1.70	TRIzol
V3	P2	Fertile	233.00	62.97	1.58	TRIzol
V4	P2	Fertile	154.00	22.81	1.58	TRIzol
V5	P2	Fertile	281.00	66.90	1.64	TRIzol
V6	P2	Fertile	327.00	51.90	1.66	TRIzol
V7	P2	Fertile	172.00	19.88	1.73	TRIzol
V8	P2	Fertile	49.00	10.43	1.72	TRIzol
V9	P2	Fertile	252.00	24.23	1.61	TRIzol
V10	P2	Infertile (T)	612.00	27.13	1.94	TRIzol
V11	P2	Infertile (OA)	220.00	114.58	1.41	TRIzol
V12	P2	Infertile (AT)	2382.00	5.91	1.76	TRIzol
V13	P2	Infertile (A)	51.00	0.87	1.87	TRIzol
V14	P2	Infertile (OA)	426.00	55.47	1.89	TRIzol
V15	P2	Infertile (O)	60.00	0.46	1.72	TRIzol
V16	P2	Infertile (T)	556.00	17.55	1.92	TRIzol
V17	P2	Infertile (T)	162.00	0.59	1.66	TRIzol
V18	P2	Infertile (T)	410.00	50.25	2.09	TRIzol
V19	P2	Infertile (A)	724.00	11.17	1.95	TRIzol
V20	P2	Infertile (OT)	322.00	41.93	1.85	TRIzol
V21	P2	Infertile (AT)	2685.00	4.63	1.71	TRIzol
V22	P2	Infertile (T)	58.00	0.51	1.39	TRIzol
V23	P2	Infertile (OT)	330.00	34.38	1.79	TRIzol
P1	AI	Fertile	224.00	11.79	1.88	mirVana
P2	AI	Fertile	486.50	17.33	1.81	mirVana
P3	AI	Fertile	127.75	7.10	2.26	mirVana

P4	AI	Fertile	281.75	16.57	2.19	mirVana
P5	AI	Fertile	65.38	2.78	1.67	mirVana
P6	AI	Fertile	115.50	7.73	1.71	mirVana
P7	AI	Fertile	169.75	10.56	1.65	mirVana
P8	AI	Fertile	122.50	6.13	1.43	mirVana
P9	AI	Fertile	122.50	5.21	1.55	mirVana
P10	AI	Fertile	2642.50	181.24	1.82	mirVana
P11	AI	Fertile	148.05	3.44	1.75	Maxwell
P12	AI	Fertile	97.05	4.99	1.85	Maxwell

Spz = spermatozoa; PR = Progressive; NP = Non Progressive; NF = Normal Forms; A = Asthenozoospermia; T = Teratozoospermia; O = Oligozoospermia; N = normozoospermia; OA = Oligoasthenozoospermia; AT = Asthenoteratozoospermia; OT = Oligoteratozoospermia; P1 = Publication 1; P2 = Publication 2; AI = Appendix I.

7.2.2.1. Sperm RNA extraction by TRIzol® (Thermo Fisher Scientific)

TRIzol method is based on the different solubility of the homogenized cell sample components in organic solvents. Therefore, sperm samples are dissociated and homogeneized by TRIzol reagent, and chloroform is further employed to separate RNA fraction.

Protocol

1. Add TRIzol® reactive to the isolated sperm extract (1 mL per 5-10 · 10^6 cells).
2. Incubate at room temperature for 5 minutes.
3. Add 0.2 mL of chloroform for each milliliter of TRIzol® added in Step 1. Mix vigorously and centrifuge at 12,000 g for 15 minutes at 4ºC.
4. Three separate phases will be visible: the lower organic phase (corresponding to proteins and lipids), the interphase (corresponding to DNA), and the upper aqueous phase (corresponding to RNA). Collect the RNA phase by pipetting carefully. Add this phase to a new sterile tube.
5. Repeat Steps 3 and 4 in order to obtain a higher RNA purity.
6. Add 0.5 mL of isopropanol (2-propanol) at 4ºC per mL of TRIzol® used in Step 1 and mix.
7. Incubate for 4 - 5h at -20ºC, or for 12 - 24h at 4ºC. RNA will precipitate to the bottom of the tube.
8. Centrifuge at 14,000 g for 10 minutes at 4ºC. Remove supernatant.
9. Add 1mL of ethanol 75 % at 4ºC per mL of TRIzol® used in Step 1. Centrifuge at 14,000 g for 4 - 5 minutes at 4ºC. RNA will be visible as a white pellet at the bottom of the tube. Remove supernatant.
10. Repeat Step 9 twice.
11. Let the RNA pellet dry at room temperature for 30 minutes.
12. Dissolve the pellet in 10 µL of RNase free H_2O.
13. Add 0.5 µL of rDNase I and 1 µL of 10X buffer (Life Technologies). Mix and incubate for 20 minutes at 37ºC.
14. Remove rDNase I by repeating Steps 1 - 12 (enzyme will be discarded within the organic phase during the TRIzol® - chloroform separation).
15. Store the final 10 µL of RNA samples at -80ºC.

7.2.2.2. Sperm RNA extraction by mirVana™ PATIS™ RNA and Native Protein Purification Kit (Thermo Fisher Scientific)

MirVana™ kit is based on a mixture between phenol-chloroform phase separation and RNA adsorption to the surface of filtering columns.

Protocol

1. Add 500 µL Cell Disruption Buffer (at 4ºC) and 10 µL DTT solution (0.15g/mL, at 4ºC) to a maximum of 10^6 sperm cells. Mix by vortex for 15 seconds.
2. Incubate at 56ºC for 15 minutes.
3. Add 500 µL Denaturing Solution (at 4ºC) and incubate on ice for 5 minutes.
4. Add 500 µL of Acid Phenol:Chloroform (at 4ºC) for every 500 µL of sample preparation. Mix by vortex for 45 seconds.
5. Centrifuge at 12,000 g for 5 minutes. Collect the upper transparent phase and add it to a new clean tube.
6. Add 1.25 volumes of ethanol 100% (at 4ºC).
7. Add the mixture (up to 700 µL each time) to a column placed in a new clean tube. Centrifuge at 12,000 g for 30 seconds and discard the eluate.
8. Repeat Step 7 until the whole sample volume has been filtered.
9. Add 700 µL of Wash Solution 1 to the column. Centrifuge at 12,000 g for 15 seconds.
10. Add 500 µL of Wash Solution 2/3 to the column. Centrifuge at 12,000 g for 15 seconds.
11. Repeat Step 10.
12. Add 50 µL of Elution Solution (at 95ºC) to the column. Incubate for 5 minutes at room temperature.
13. Centrifuge at 12,000 g for 30 seconds.
14. Store the final volume of 50 µL RNA samples at -80ºC.

7.2.2.3. Sperm RNA extraction by Maxwell® RSC simplyRNA Cells Kit (Promega)

Maxwell® kit is an automatized method performed by the Maxwell® instrument after loading the processed samples and the required reagents into a cartridge provided by the manufacturer. This method is based on the isolation of RNA by their adsorption to a set of paramagnetic beads.

Protocol

1. Add 200 µL of 1-Thioglycerol/Homogenization Solution mixture (20 µL 1-Thioglycerol/1 mL Homogenization Solution) at 4ºC to the pellet of sperm cells.
2. Mix by vortex until the pellet is completely resuspended. Preserve the sample in ice until it is used (it can also be stored at -80ºC at this Step).
3. Prepare the cartridge provided by the kit: place it in the deck tray of the Maxwell® instrument, peel back the seal, place a plunger in Well #8, and place a clean elution tube in the predefined position of the deck tray. Add 50 µL of nuclease-free water to the bottom of the tube.
4. Add 200 µL of Lysis Buffer to the 200 µL of sample. Vortex for 15 seconds. Transfer the 400 µL of lysate to Well #1 of the cartridge.
5. Add 10 µL of DNase I solution (lyophilized DNase I with 275 µL of nuclease-free water and 5 µL of blue dye) to Well #4 of the cartridge.
6. Setup the Maxwell instrument and run the simplyRNA Cell program.
7. Store the final volume of 50 µL RNA samples at -80ºC.

7.2.3. Quality controls

The RNA fractions were subjected to several quality controls in order to assess its quantity, purity, integrity, and the absence of DNA and RNA from non-sperm cells. These controls are summarized in Table 12.

Table 12. Summary of the employed sperm RNA quality controls.

Method		**RNA concentration**	**RNA purity**	**RNA integrity**	**Sperm purification**
Spectrophotometry (Nanodrop 2000)		ng/µL	260/280 nm ratio		
Fluorometry (Quantus)		ng/µL			
RT-PCR	*PRM1, GAPDH*		DNA absence	cDNA amplification	
	CD45				Absence of DNA and RNA from leukocytes
	KIT				Absence of DNA and RNA from germ cells
	CDH1				Absence of DNA and RNA from epithelial cells
Nanoelectrophoresis (Bioanalyzer 2100): Nano6000 RNA chip					rRNA degradation

7.2.3.1. Spectophotometry and fluorometry measurement of RNA quantity and purity

RNA concentration was measured by a full-spectrum UV-Vis spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific), which allows detecting RNA concentrations ranging from 2 ng/μL to 3000 ng/μL. Additionally, RNA purity of the samples was determined by assessing spectrophotometric measures at 260 nm and 280 nm [(2)].

To provide more accurate measures of RNA concentration, a fluorometer system (Quantus™, Promega) was employed to asses RNA quantity of the samples destined to RNA-seq analysis (P1-P12, from Appendix I). Samples were diluted with RNase-free water in 1:100 proportion, and mixed with an equal volume of dye solution (QuantiFluor® dye and Tris-EDTA buffer 1X in 1:1,000 proportion). The dilution was measured by Quantus and RNA concentration (ng/μL) was estimated.

[(2)] A RNA sample with optimal purity should present a 260/280 ratio of 1.8-2.0 (Fleige and Pfaffl, 2006).

7.2.3.2. Nanoelectrophoresis analysis of rRNA degradation

Agilent Nano 6000 RNA nanoelectrophoresis chips were used to evaluate the presence of RNA transcripts from 25 to 6000 nt. By the obtained profiles, the absence of specific size peaks corresponding to 18S and 28S rRNA were verified [(3)].

Briefly, RNA samples were diluted in RNase free water until reaching a concentration of 25-50 ng in 2 µL. Samples and the molecular-weight marker (5 µL) were denatured at 70 ºC for 2 minutes. Afterward, they were loaded in their corresponding well of the chip, as well as the gel-dye (9 µL), and the conditioning solution (9 µL). The chip was centrifuged at 2,400 rpm for 1 minute and loaded into the Agilent 2100 Bioanalyzer (Agilent Technologies).

[(3)] A nanoelectrophoresis signal corresponding to non-degraded 18S and 28S indicates the presence of non-sperm cells in the sample.

7.2.3.3. RT-PCR and PCR of *PRM1, GAPDH, CD45, KIT* and *CDH1*

The integrity of sperm RNA transcripts, the lack of sperm DNA, and the lack of non-sperm cell contamination were analyzed through a set of RT-PCR reactions.

Reverse Transcription (cDNA Synthesis)

Sperm RNA was subjected to reverse transcription using random primers. For every RNA sample, a reaction tube was prepared using the High-Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific) as follow:

Component	Volume (µL) per reaction
RT Buffer (10X)	0.40
dNTPs (with dTTPs)	0.16
MultiScribe Reverse Transcriptase (50 U/µL)	0.20
RT random primers (10X)	0.40
RNase-free H_2O	0.84
RNA sample (5-10 ng/µL)	2.00
Total volume	4.00

dNTP = Deoxynucleotide triphosphate; dTTP = Deoxythymidine triphosphate.

All components were mixed and incubated on ice for 5 minutes. RT-PCR was performed in the thermocycler Mastercycler ep Gradient S 96 (Eppendorf) under the following conditions:

Cycles	Time	Temperature
1	10 minutes	25 ºC
1	120 minutes	37 ºC
1	5 seconds	85 ºC

Polymerase Chain Reaction (PCR)

The synthesized cDNA was subjected to PCR. The primers of the selected genes were designed with the software Primer3 v.3.0.0 (www.primer3.wi.mit.edu), and the Gene Sorter and In-Silico PCR tools of UCSC Genome Browser (https://genome.ucsc.edu/). The characteristics and sequences of the primers used are summarized in Table 13.

Table 13. Characteristics and sequences of the primers used in the sperm RNA quality controls.

Gene	Primer characteristics		Amplicon characteristics		
	Sequence (5' - 3')	**Annealing temperature (°C)**	**cDNA length (bp)**	**DNA length (bp)**	**Chr location**
PRM1	For: CAGAGTTCCACC TGCTCACA Rev: GGATGGTGGCAT TTTCAAGA	62	331	422	chr16:11, 280, 901-11, 281, 322
GAPDH	For: CGACCACTTTGT CAAGCTCA Rev: AGGGGTCTACAT GGCAACTG	64	228	332	chr12:6, 537, 970+6, 538, 301
CD45	For: CCTTGAACCCGA ACATGAGT Rev: ATCTTTGAGGGG GATTCCAG	60	179	3307	chr1:198, 709, 727+198, 713, 033
KIT	For: CCCAACACAACTT CCTTATGATC Rev: CATAAGCAGTTG CCTCAACAAC	59.8	113	393	chr4:54,727, 484+54,727, 876
CDH1	For: GAACCTCTGTGA TGGAGGTCAC Rev: GTCAGTGACTGT GATCACAGCTG	61	269	544	chr16:68,811, 684+68,812,2 27

For = forward, Rev = Reverse, bp = base pairs, chr = chromosome

For each sample, three PCR tubes were prepared for every single gene using the AmpliTaq Gold® DNA Polymerase with GeneAmp® 10X PCR Gold Buffer (Thermo Fisher Scientific) as follows:

Component	Volume (µL) per reaction
Buffer II	2.50
$MgCl_2$	2.00
dNTPs	2.00
Forward primer [(1)]	2.00
Reverse primer [(1)]	2.00
TaqGold enzyme	0.25
RNase-free H_2O	13.25
cDNA sample	1.00
Total volume	25.00

dNTP = Deoxynucleotide triphosphate.

[(1)] Primer set corresponding to the amplified marker gene.

All components were mixed and incubated on ice for 5 minutes. PCR was performed on the Mastercycler ep Gradient S 96 thermal cycler (Eppendorf) under the following conditions:

Cycles	Time	Temperature
1	10 minutes	94 °C
35	45 seconds	94°C
	45 seconds	PRM1: 62 °C
		GAPDH: 64 °C
		CD45: 60 °C
		KIT: 59.8 °C
		CDH1: 61 °C
	1 minute	72 °C

The size of the obtained amplicons was analyzed by agarose gel electrophoresis.

Composition of 2% Agarose gel

- 60 mL of Tris-Borate-EDTA (TBE) buffer
- 1.2 g agarose
- 3 µL SYBR Safe DNA Gel Stain (Thermo Fisher Scientific).

cDNA samples were mixed with a loading buffer and placed into the wells of the gel along with a molecular-weight size marker of 0.1–1 kb (Promega).

Electrophoresis was performed at 90 mV for 45 minutes.

The gels were visualized using the E-Box VX2 Complete Imaging System (Vilber Lourmat).

7.2.4. Single sperm miRNAs quantification by qRT-PCR

7.2.4.1. Reverse Transcription (cDNA Synthesis)

During the development of Publication 2, single qRT-PCR assays of 10 different miRNAs were performed. These miRNAs included:

1) The eight members of the four stable pairs described in Publication 2: 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.

2) The two normalizers for singleplex qPCR studies described in Publication 1: hsa-miR-100-5p and hsa-miR-30a-5p.

For the retrotranscription of the miRNA, gene-specific stem-loop primers were used (TaqMan® microRNA Assays, Thermo Fisher Scientific). The sequences of the miRNAs and the used primers are summarized in Table 14.

Table 14. Stem loop-primers used in the retrotranscription of the miRNAs included in the qRT-PCR single assays.

miRNA	Mature miRNA sequence (5'-3')	Stem-loop sequence (5'-3')
hsa-miR-942-5p	UCUUCUCUGUUUUGGCCAUGUG	AUUAGGAGAGUAUCUUCUCUGUUUUGGCCAUGUGUGUACUCACAGCCCCUCACACAUGGCCGAAACAGAGAAGUUACUUUCCUAAU
hsa-miR-1208	UCACUGUUCAGACAGGCGGA	CACCGGCAGAAUCACUGUUCAGACAGGCGGAGACGGGUCUUUCUCGCCCUCUGAUGAGUCACCACUGUGGUGG
hsa-miR-296-5p	AGGGCCCCCCCUCAAUCCUGU	AGGACCCUUCCAGAGGGCCCCCCCUCAAUCCUGUUGUGCCUAAUUCAGAGGGUUGGGUGGAGGCUCUCCUGAAGGGCUCU
hsa-miR-328-3p	CUGGCCCUCUCUGCCCUUCCGU	UGGAGUGGGGGGGCAGGAGGGGCUCAGGGAGAAAGUGCAUACAGCCCCUGGCCCUCUCUGCCCUUCCGUCCCCUG
hsa-miR-139-5p	UCUACAGUGCACGUGUCUCCAGU	GUGUAUUCUACAGUGCACGUGUCUCCAGUGUGGCUCGGAGGCUGGAGACGCGGCCCUGUUGGAGUAAC
hsa-miR-1260a	AUCCCACCUCUGCCACCA	ACCUUUCCAGCUCAUCCCACCUCUGCCACCAAAACACUCAUCGCGGGGUCAGAGGGAGUGCCAAAAAAGGUAA
hsa-miR-34b-3p	CAAUCACUAACUCCACUGCCAU	GUGCUCGGUUUGUAGGCAGUGUCAUUAGCUGAUUGUACUGUGGUGGUUACAAUCACUAACUCCACUGCCAUCAAAACAAGGCAC
hsa-miR-93-3p	ACUGCUGAGCUAGCACUUCCCG	CUGGGGGCUCCAAAGUGCUGUUCGUGCAGGUAGUGUGAUUACCCAACCUACUGCUGAGCUAGCACUUCCCGAGCCCCCGG
hsa-miR-100-5p	AACCCGUAGAUCCGAACUUGUG	CCUGUUGCCACAAACCCGUAGAUCCGAACUUGUGGUAUUAGUCCGCACAAGCUUGUAUCUAUAGGUAUGUGUCUGUUAGG
hsa-miR-30a-5p	UGUAAACAUCCUCGACUGGAAG	GCGACUGUAAACAUCCUCGACUGGAAGCUGUGAAGCCACAGAUGGGCUUUCAGUCGGAUGUUUGCAGCUGC

A reaction tube was prepared for each miRNA assay and RNA sample using the TaqMan® MicroRNA Reverse Transcription kit (Thermo Fisher Scientific) and the primer pool of the TaqMan® microRNA Assays as follows:

Component	**Volume (µL) per reaction**
dNTPs (100 mM) Human Primer Pool (5X) MultiScribe Reverse Transcriptasa (50 U/µL) Reverse Transcription Buffer (10X) RNase-free H_2O RNA sample (1-10 ng)	0.15 3.00 1.00 1.50 4.35 5.00
Total volume	15.00

dNTP = Deoxynucleotide triphosphate.

All components were mixed and incubated on ice for 5 minutes. Retrotranscription was performed in the Mastercycler ep Gradient S 96 thermal cycler (Eppendorf) under the following conditions:

Cycles	**Time**	**Temperature**
1	30 minutes	16 ºC
1	30 minutes	42 ºC
1	5 minutes	85 ºC

7.2.4.2. Quantitative Polymerase Chain Reaction

A reaction solution was prepared for every cDNA sample using the TaqMan® microRNA Assays and the TaqMan® Universal PCR Master Mix II, No UNG (Thermo Fisher Scientific). Additionally, a negative control (with RNase-free water instead of cDNA) was added for each miRNA assay.

Component	Volume (µL) per reaction
TaqMan® Assay (20X) Universal PCR Master Mix (2X) RNase-free H_2O cDNA sample	3.60 36.00 27.60 4.80
Total volume	72.00

All components were mixed and maintained in ice until their use.

The total volume of each reaction solution was divided in three aliquots of 20 µL and loaded separately into 384-well plates as technical replicates.

qPCR was performed in the thermocycler 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific) under the following conditions:

Cycles	Time	Temperature
1	10 minutes	95 ºC
40	15 seconds	95 ºC
	1 minute	60 ºC

7.2.5. RNA-seq library construction and sequencing

Intending to perform a RNA-seq analysis, the corresponding RNA samples were subjected to a library construction of polyadenilated transcripts. This process was performed by IGATech (Udine, Italy). The followed protocol is briefly summarized below:

1. Allicuote 1-10 ng from the RNA samples.
2. Construct cDNA libraries using the Universal Plus mRNAseq with NuQuant kit (NuGEN), according to the manufacturer's intructions.
3. Perform a selection of poly(A) RNA molecules by Oligo(dT) beads and elute the bound molecules.
4. Submit the selected RNAs to an enzymatic fragmentation.
5. Prepare a RT-PCR with the fragmented RNA, using a mixture of random and oligo(dT) primering.
6. Use a DNA end-repair reaction to generate blunt ends of the cDNA.
7. Perform a strand-specific adaptor ligation and preamplify the cDNA by PCR.

The quality controls performed throughout the library construction process included:

1) The quantification of the obtained libraries by Qubit Fluorometer (Thermo Fisher), to verify that the quantity of cDNA material was enough for the sequencing protocol.

2) The size distribution verification of the libraries using the Bioanalyzer DNA High Sensitivity assay (Agilent).

Finally, paired-end sequencing was performed using the HiSeq 2500 platform (Illumina) to produce fragments of 125 bp. Library sizes were set to 30M reads per sample.

The Bcl2Fastq pipeline (v.2.0.2) was used to perform de-multiplexing of the obtained data and produce the final raw data file (*fastq*) of each analyzed sample.

7.3. Scripts

7.3.1. Mean-centering Restricted (R script)

```
DATA_RAW <- t(DATA)
for (i in 1:ncol(DATA_RAW))
       {
       DATA_RAW[,i][DATA_RAW[,i] > 35] <- NA
       }

DATA_FOR_MEAN_CENTERED <- DATA_RAW

write.table(DATA_FOR_MEAN_CENTERED,"RAW_DATA_ready_FOR_MEAN_CENTERED.
txt",sep="\t")

meanCenter = function(rawData, restricted=TRUE, ...) {
       if (restricted) {
              means = apply(X=na.omit(rawData), MARGIN=2, FUN=mean)
       } else {
              means = apply(X=rawData, MARGIN=2, FUN=mean,
na.rm=TRUE)
       }
       meanCenteredData = rawData - data.frame(lapply(X=means,
FUN=rep, times=nrow(rawData)))
       return(meanCenteredData)
}

MEAN.Centered.norm1 <- meanCenter(DATA_FOR_MEAN_CENTERED[1:384,],
restricted=TRUE)
MEAN.Centered.norm2 <- meanCenter(DATA_FOR_MEAN_CENTERED[385:768,],
restricted=TRUE)

MEAN_all_restricted <- rbind(MEAN.Centered.norm1,MEAN.Centered.norm2)

rownames(MEAN_all_restricted)<-rownames(DATA_FOR_MEAN_CENTERED)

DATA_MEAN<-data.frame(MEAN_all_restricted)

DATA_MEAN<-data.frame(t(DATA_MEAN))

DATA_MEAN$subjects<-rownames(DATA_MEAN)

write.table(DATA_MEAN,"DATA_quantileNorm.txt",sep="\t",row.names=F)
```

7.3.2. Concordance Correlation Restricted (R script)

```
fastCCC = function(x, y) {
	pearson = cor(x, y)
	m1 = mean(x)
	m2 = mean(y)
	s1 = sqrt(mean(x*x) - m1*m1)
	s2 = sqrt(mean(y*y) - m2*m2)
	correction = (2 * s1 * s2) / (s1^2 + s2^2 + (m1-m2)^2)
	ccc = pearson * correction
	return(ccc)
}

selectCCRNormalizers = function(data, ctThreshold=35,
ccRankCutoff=10, normCount=NA) {
	require(epiR)
	data[data>ctThreshold] = NA
	data = na.omit(data)
	if (nrow(data) < ccRankCutoff) {
		stop(paste("Need at least ccRankCutoff =",
ccRankCutoff, "fully observed rows."))
	}
	geneSymbol = rownames(data)
	sampleMeans = apply(X=data, MARGIN=2, FUN=mean)
	allCCCVals = apply(X=data, MARGIN=1, FUN=fastCCC,
y=sampleMeans)
	cccOrder = order(allCCCVals, decreasing=TRUE)[1:ccRankCutoff]
	data = data[cccOrder,]
	geneSymbol = geneSymbol[cccOrder]
	excludedData = data[-1,]
	excludedGeneSymbol = geneSymbol[-1]
	data = as.data.frame(data[1,])
	normalizerSymbol = geneSymbol[1]
	converged = FALSE
	groupMeanCCC = list(epi.ccc(unlist(data), sampleMeans))
	while (!converged) {
		bestCCC = list(rho.c = list(est = -Inf))
		bestCCCIndex = -Inf
		for (addedIndex in 1:nrow(excludedData)) {
			potentialData = rbind(data,
as.data.frame(excludedData[addedIndex,]))
			potentialMean = apply(X=potentialData, MARGIN=2,
FUN=mean)
			cccVal = fastCCC(potentialMean, sampleMeans)
			if (cccVal > bestCCC$rho.c$est) {
```

```
                                bestCCC = epi.ccc(potentialMean,
sampleMeans)
                                bestCCCIndex = addedIndex
                        }
                }
                converged = (bestCCC$rho.c$est <=
groupMeanCCC[[nrow(data)]]$rho.c$upper)
                if (!converged || !is.na(normCount)) {
                        groupMeanCCC[[length(groupMeanCCC)+1]] = bestCCC
                        data = rbind(data,
as.data.frame(excludedData[bestCCCIndex,]))
                        normalizerSymbol = c(normalizerSymbol,
excludedGeneSymbol[bestCCCIndex])
                        excludedData = excludedData[-bestCCCIndex,]
                        excludedGeneSymbol = excludedGeneSymbol[-
bestCCCIndex]
                }
                if (!is.na(normCount)) {
                        converged = (nrow(data) >= normCount)
                }
        }

        return(list(normalizers=normalizerSymbol,
cccTrace=groupMeanCCC))
}
```

7.3.3. STAR genome index generation and reads alignment (Linux command line)

7.3.3.1. Genome index generation from FASTA file of the human genome (GRCh38 assembly)

```
module load bioinfo-tools star/2.5.3a

star --runThreadN 9 --runMode genomeGenerate --genomeDir
genomeIndexhDirectoryPath --genomeFastaFiles
Homo_sapiens.GRCh38.dna.primary_assembly.fa
```

7.3.3.2. Alignment of raw *fastq* files to genome

```
module load bioinfo-tools star/2.5.3a

star --genomeDir genomeIndexhDirectoryPath --outSAMtype BAM
SortedByCoordinate Unsorted --runThreadN 9 --readFilesIn
Sample1.Reads1.fastq.gz Sample1.Reads2.fastq.gz --seedSearchStartLmax
12 --outFilterScoreMinOverLread 0.3 --alignSJoverhangMin 15 --
outFilterMismatchNmax 33 --outFilterMatchNminOverLread 0 --
outFilterType BySJout --outSAMunmapped Within --outSAMattributes NH
HI AS NM MD --outSAMstrandField intronMotif --outWigType bedGraph --
readFilesCommand zcat --outFileNamePrefix sample1Alignment
```

7.3.4. Gene and lncRNA count obtention with FeatureCounts (Linux command line)

7.3.4.1. Gene count from *BAM* files using GTF annotation file (GRCh38 assembly)

```
module load bioinfo-tools subread/1.5.2

featureCounts -a annotationFile.gtf -p -s 1 -t exon -g gene_id -T 9 -
o geneCountOutputName.txt sample1Alignment.bam
```

7.3.4.2. lncRNA count from *BAM* files using GTF annotation file of lncRNA sequences (GENCODE)

```
module load bioinfo-tools subread/1.5.2

featureCounts -a lncRNA.annotationFile.gtf -p -s 1 -M -O -Q 0 -o
lncRNAcountOutputName.txt -t exon -g gene_id -T 9
sample1Alignment.bam
```

7.3.5. TIN calculation (Linux command line)

This analysis requires building a table in 12-column BED format containing information about the transcripts from the hg38 assembly. The file containing the aligned reads (bam file) of each sample require to be sorted by coordinate and indexed which can be done with samtools:

```
module load bioinfo-tools samtools/1.9
samtools sort sample1Alignment.bam -o sample1AlignmentSorted.bam -@ 9
samtools index -b sample1AlignmentSorted.bam
```

From all the needed files, the tin.py function of the RSeq package can be used to calculate TIN score of each transcript in every sample:

```
module load bioinfo-tools rseqc/2.6.4
tin.py -i sample1AlignmentSorted.bam –r hg38Annotation.bed
```

7.3.6. lncRNA *cis* target genes prediction with BEDTools (Linux command line)

A table in BED format containing chromosome coordinates of the lncRNAs of interest must be constructed as an initial step. This table is compared to a BED table of human genome sequences to get those genes comprised in a 10 kb window around each lncRNA.

```
module load bioinfo-tools BEDTools/2.27.1

bedtools window -a lncRNA_annotation.bed -b hg38.bed -w 10000 >
outputCisTargets.txt
```

8 Abbreviations and genes

Abbreviations

µg	Microgram
AUC	Area Under the Curve
Bp	Base Pair
CAGE	Cap Analysis of Gene Expression
CAR	Chromatin-associated RNA
CCR	Concordance Correlation Restricted
cDNA	Complementary DNA
circRNA	Circular RNA
CNV	Copy Number Variations
Ct	Threshold Cycle
DAVID	Database for Annotation, Visualization and Integrated Discovery
DE-miRNA	Differentially Expressed microRNA
DFS	Dysplasia of the Fibrous Sheath
DGCR8	DiGeorge Syndrome Critical Region 8
dL	Deciliter
DRM	Detergent Resistant Membrane
endo-siRNA	Endogenous small interfering RNA
ES	Elongating Spermatid
EST	Expressed Sequence Tags
FDR	False Discovery Rate
fg	Femtograms
FN	False Negative
For	Forward
FP	False Positive
FPKM	Fragments Per kilobase Per Million
FSH	Follicle-Stimulating Hormone
GO	Gene Ontology
hCG1	Human Chorionic Gonadotrophin
ICMART	International Committee for Monitoring Assisted Reproductive Technology
ICSI	Intracytoplasmic Sperm Injection
IMI	Idiopathic Male Infertility
ISH	In Situ Hybridization
IUO	Infertility of Unknown Origin
LH	Luteinizing Hormone
LINE1	Long Interspersed Nuclear Elements 1
lncRNA	Long non-coding RNA

LOESS	Locally Estimated Scatterplot Smoothing
LOWESS	Locally Weighted Scatterplot Smoothing
MC	Mean-Centering
MCR	Mean-Centering Restricted
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
miRISC	miRNA-induced Silencing Complex
miRNA	microRNA
mIU	Million International Units
mL	Mililiter
MRHL	Meiotic Recombination Hot Spot Locus
mRNA	Messenger RNA
mse-tsRNAs	Mature-sperm-enriched tRNA-derived small RNAs
ncRNA	Non-coding RNA
NF	Normal Forms
ng	Nanogram
NGS	Next Generation Sequencing
NOA	Non-obstructive Azoospermia
NP	Non progressive motility
nt	Nucleotides
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PGC	Primordial Germ Cells
piRNA	PIWI-interacting RNA
PR	Progressive motility
pre-miRNA	Precursor microRNA
pri-miRMA	Primary microRNA
ProSpg	Prospermatogonia
qRT-PCR	Real-time Reverse Transcription PCR
Rev	Reverse
REViGO	Reduce and Visualize Gene Ontology
RIN	RNA Integrity Number
RISC	RNA-induced Silencing Complex
RNA-seq	RNA sequencing
ROC	Receiver Operating Characteristic
ROS	Reactive Oxygen Species
RPKM	Reads Per Kilobase Per Million
rRNA	Ribosomal RNA
RS	Round Spermatid
RT-PCR	Reverse Transcription PCR

SAGE	Serial Analysis of Gene Expression
SCLB	Somatic Cell Lysis Buffer
SCOS	Sertoli-cell-only Syndrome
snaR	Small ILF3/NF90-associated RNA
sncRNA	Small non-coding RNA
snoRNA	Small nucleolar RNA
SNP	Single Nucleotide Polymorphism
snRNA	Small Nuclear RNA
Spc	Spermatocyte
Spg	Spermatogonia
Spz	Spermatozoa
sRNA-seq	small RNA Sequencing
SSC	Spermatogonial Stem Cell
TE	Transposable Elements
TEC	To be Experimentally Confirmed
TIN	Transcript Integrity Number
TN	True Negative
TP	True Positive
TPM	Transcripts Per Million
TSS	Transcription Start Site
UMI	Unexplained Male Infertility
UTR	Untranslated Region
WHO	World Health Organization
Wnt	Wingless int-1
ZGA	Zygote Genome Activation

Genes

A2M	Alpha-2-macroglobulin
ACRV1	Acrosomal Vesicle Protein 1
ACSBG2	Long-chain-fatty-acid—CoA ligase
ACTB	β-actin
AGO	Argonaute
AKAP4	A-kinase Anchoring Protein 4
ANXA2	Annexin A2
AURKC	Aurora Kinase C
BCL2	B Cell Lymphoma 2
BRD2	Bromodomain-containing Protein 2
CAB39L	Calcium Binding Protein 39 Like Gene
CATSPER	Cation Channel Sperm Associated 1
CBLN1	Cerebellin 1 Precursor
CCN	Cyclin
CD45	Protein Tyrosine Phosphatase Receptor Type C
CDH1	Cadherin-1
CDY1	Chromodomain Y1
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHD2	Chromodomain Helicase DNA Binding Protein 2
CRISP2	Cysteine-rich Secretory Protein 2
DAZ	Deleted In Azoospermia
DAZL	Deleted In Azoospermia Like
DICER	Ribonuclease type III
DMR	DMRT1 related gene
DNAH1	Dynein Axonemal Heavy Chain 1
DPY19L2	Dpy-19 Like 2
DROSHA	Class 2 ribonuclease III
eNOS	Nitric Oxide Synthase 3
ERBB3	Erb-b2 Receptor Tyrosine Kinase 3
ESX1	ESX Homeobox 1
FOXG1B	Forkhead Box G1
GA17	Eukaryotic Translation Initiation Factor 3 Bubunit M
GAPDH	Glyceraldehyde 3-phosphate Dehydrogenase
GGT1	Gamma-Glutamyltransferase 1
GPCR	G Protein-coupled Receptor

HBEGF	Heparin Binding EGF Like Growth Factor
HILS1	Spermatid-specific Linker Histone H1-like Protein
HLA-E	Human Leukocyte Antigen-E
HOX2A	Homeobox Protein Hox-2a
HSBP-1	Heat Shock Factor Binding Protein 1
HSFY	Heat Shock Factor Y Chromosome
HSPA2	Heat-Shock Protein A2
IAP	Intracisternal A-particle
ICSBP	Interferon Consensus Sequence-binding Protein
JNK2	C-Jun Kinase 2
JUN	Ju-nana
LDHC	Lactate Dehydrogenase C
MCM8	Minichromosome Maintenance 8 Homologous Recombination Repair Factor
MT-CO	Mitochondrially Encoded Cytochrome C Oxidase
NF1	Neurofibromin 1
NFAT	Nuclear Factor of Activated T-cells Protein Family
NFjB	Nuclear Factor Kappa B
NIPBL	Delangin
nNOS	Nitric Oxide Synthase 1
NPC2	Epididymal Protein 1
Nrf2	Nuclear Factor Erythroid-Derived 2-Like 2
NUMA1	Nuclear Mitotic Apparatus Protein 1
NUP155	Nucleoporin 155
OAZ3	Ornithine Decarboxylase Antizyme 3
ODF	Outer Dense Fiber of Sperm Tails
PARK7	Parkinsonism Associated Deglycase
PKM	Pyruvate Kinase M1/2
PLC-Z	Phospholipase C-Z
PMS1	DNA Mismatch Repair Protein
PPGB	Cathepsin A
PPP2R2B	Protein Phosphatase 2 Regulatory Subunit Beta
PRM	Protamine
PSG1	Pregnancy-Specific B-1-glycoprotein 1
RANBP9	RAN Binding Protein 9
RBMY1	RNA Binding Motif Protein, Y-linked, Family 1
RMND5B	Required for Meiotic Nuclear Division 5 Homolog B
RNF7	Ring Finger Protein 7
RPL7	Ribosomal Protein L7
RPL9	Ribosomal Protein L9

RPS10	Ribosomal Protein S10
RPS3A	Ribosomal Protein S3A
RXRB	Retinoid X Receptor Beta
S100A6	S100 Calcium Binding Protein A6
SCAMP2	Secretory Carrier Membrane Protein 2
SEPT12	Septin 12
SGP2	Clusterin
SKP1	S-phase Kinase-associated Protein 1A
SLC26A8	Solute Carrier Family 26 Member 8
SP1	Specificity Protein 1
SPACA4	Sperm Acrosome Associated 4
SPAM1	Sperm Adhesion Molecule 1
SPATA	Spermatogenesis Associated
SPATS1	Spermatogenesis Associated Serine Rich 1
SPG2	Proteolipid Protein 1
SPP1	Secreted Phosphoprotein 1
SPZ1	Spermatogenic Leucine Zipper 1
SRP	Signal Recognition Particle
SRY	Sex Determining Region Y
STAT4	Signal Transducer and Activator of Transcription 4
SUI1	Eukaryotic Translation Initiation Factor 1
SYCE1	Synaptonemal Complex Central Element Protein 1
TAF	TATA-Box Binding Protein Associated Factor 4b
TCF3	Transcription Factor 3
TCP11	T-complex 11
TEX	Testis Expressed
TGFB1	Transforming Growth Factor Beta 1
TGIFLX	TGFB Induced Factor Homeobox 2 Like X-linked
THRA	Thyroid Hormone Receptor Alpha
TLK2	Tousled Like Kinase 2
TMEM225	Transmembrane Protein 225
TMSL3	TMSB4X Pseudogene 8
TNP	Transition Protein
TNPAIP3	TNF-alpha Induced Protein 3
TNRC6	Trinucleotide Repeat–containing 6
TPX1	Testis-specific Protein
TRBP	Nuclear Receptor Coactivator 6
TRKA	Tyrosine Kinase Receptor A
TRY1	Trypsin 1
TSSK6	Testis Specific Serine Kinase 6

TSX	Testis-specific X-linked
TTC7A	Tetratricopeptide Repeat Domain 7A
TYROBP	TYRO Protein Tyrosine Kinase Binding Protein
U6	RNU6B
UBAC1.2	Ubiquitin-associated Domain-containing Protein 1
UBE2B	Ubiquitin Conjugating Enzyme E2 B
UBE2D2	Ubiquitin Conjugating Enzyme E2 D2
UBN1	Ubinuclein 1
UBQLN3	Ubiquilin 3
USP25	Ubiquitin Specific Peptidase 25
VDAC3	Voltage Dependent Anion Channel 3
VIM	Vimentin
VSIG4	V-set and Immunoglobulin Domain Containing 4
WBSCR28	Williams-Beuren Syndrome Chromosomal Region 28 Protein
WNT5A	Wnt Family Member 5A
ZMYND15	Zinc Finger Mynd-containing Protein 15