Reproductive Biology and Endocrinology

RESEARCH

Open Access

The composition of human sperm sncRNAome: a cross-country small RNA profiling

Poonam Mehta[1,2] and Rajender Singh[1,2*]

Abstract

Background Over the last decade, numerous studies have implicated sperm-borne small non-coding RNAs (sncRNAs) in fertility and transgenerational inheritance. Spermatozoa contain a variety of small RNAs; however, inter-individual and inter-population variations in the human sperm sncRNA content (sncRNAome) have not yet been ascertained.

Methods We performed sncRNA sequencing in 54 normozoospermic proven fertile Indian donors. We also obtained a second semen sample from 13 donors and a third semen sample from eight donors and repeated sncRNA sequencing. To better understand sperm sncRNAome similarities and variations, sncRNA sequencing data for eligible Chinese ($n=87$), US ($n=14$), and Spanish ($n=2$) normozoospermic (fertile or presumptive fertile) samples were downloaded and analyzed in a uniform manner. sncRNA data were compared within and across populations to identify similarities and differences.

Results In Indian samples, rsRNAs (13.71–78.76%), YsRNAs (0.64-76.53%) and tsRNAs (5.63–35.16%) constituted the major fraction and miRNAs, piRNAs, mt-tsRNAs, and other sncRNAs constituted the minor fraction. Across three other populations, rsRNAs (11–80%) and tsRNAs (10–60%) constituted the major fraction, and YsRNAs (0.62–4.28%), miRNAs (0.41–7.37%), piRNAs (1.37–4.36%), mt-tsRNAs (0.14–4.33%), and other sncRNAs constituted the minor fraction. Only 47 miRNAs were consistent across the Indian samples, and only 17 miRNAs were consistent across the four populations. Interestingly, all piRNAs detected in Indian samples were derived from the chromosome 15 piRNA cluster, which were also predominantly present in other populations. tRNA-Gly-GCC contributed approximately 50% of the tsRNA pool across all populations. The mt-tsRNAs also originated majorly from one mt-tRNA that differed across populations. Among the rsRNAs, the maximum number of reads belonged to 28S, followed by 18S, 5S, 5.8S, and 45S in decreasing order. Y4sRNAs were the most abundant YsRNAs, while the second most common contributor differed across populations.

Conclusions The human sperm sncRNAome has a 'core component' that shows small variations and a 'peripheral component' that shows significant variations across individuals and populations. The availability of the normal human sperm sncRNAome would help delineate biologically meaningful variations from sample-to-sample natural/random variations.

*Correspondence:
Rajender Singh
rajender_singh@cdri.res.in

Full list of author information is available at the end of the article

Keywords Sperm RNAs, Sperm small RNAs, Human sperm RNAs, Transgenerational inheritance, Sperm sncRNAome

Introduction

Spermatozoa were discovered to possess minor quantities of RNA in 1962 [1]; however, their significance remained unknown for more than five decades. Early studies on sperm RNAs suggested that they are remnants of past activity without any significance in mature sperm. Later, direct evidence of the massive makeover of the sperm RNA repertoire during epididymal maturation of spermatozoa [2–4] and their actual delivery to the oocyte [5, 6] suggested their biological significance. Some studies provided direct evidence for the role of certain small RNAs in sperm fertility and early embryonic development. For example, Liu et al., (2012) showed that sperm-borne miR-34c was required for the first cleavage division in mice following fertilization [7]. Considering the presence of small and regular RNAs in sperm and their relationship with sperm fertility and early embryonic development, some investigators proposed exploiting them as fertility biomarkers. A number of studies in the last decade have tested the suitability of sperm RNAs, particularly small RNAs, as sperm fertility biomarkers [8–11]. We performed a cross-comparison across all these studies and found that close to 400 sperm miRNAs have been considered as fertility biomarkers [12].

On a different note, Lamarck proposed the theory of inheritance in 1809, stating that the characters acquired during an organism's life could be inherited by way of use and disuse in his lifetime. He supported his theory by citing examples, the most famous of which was the stretching and elongation of the giraffe neck, among others. However, neither Lamarck nor others could explain such inheritance, which was applicable to certain characters, but not to all. A notable study in 2005 exposed pregnant rats to endocrine disruptors and observed a transgenerational phenotype of reduced spermatogenic potential that could be linked with DNA methylation changes [13]. This study set the stage for future studies that would discover an equivalent role for sperm RNAs in transgenerational inheritance. In the first study implicating sperm RNAs in transgenerational inheritance, homozygous wild-type progeny of the null mutant *Kit* gene showed a phenotype that was driven by changes in sperm RNAs [14]. In an early study on environmental exposure, Gapp et al., (2014) subjected mice to trauma early in life, leading to behavioural issues in a transgenerational manner that were linked to alterations in sperm and brain miRNAs [15].

The above studies sparked investigations into the role of sperm RNAs in transgenerational inheritance. Later, several other animal studies provided strong evidence of how paternal environment/experiences, such as stress [16–18], diet [19, 20], and exercise [21, 22] can influence the sperm RNA profile and affect the upcoming generations in an intergenerational or transgenerational manner. For example, a high fat diet in mice resulted in increased levels of 5′tsRNA-Gly-GCC in mature sperm, resulting in enhanced gluconeogenesis in offspring [20]. A recent study in mice showed that paternal exposure to a high fat diet resulted in upregulation of the 5' fragmentation of mt-tRNAs in spermatozoa, which were transferred to the oocyte and altered gene expression [23]. Following animal studies, human studies have corroborated the effects of diverse environmental factors, such as smoking, obesity, stress, and diet on the sperm small non-coding RNAs (sncRNAs) profile [24, 25]. A study on human sperm from lean and obese men found that the expressions of specific microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), tRNA derived small RNAs (tsRNAs), and small nuclear RNAs (snRNAs) were significantly altered in the obese group [26]. An analysis of sncRNAs in young, healthy Finnish men showed a major change in mitochondrial tRNA derived small RNAs (mt-tsRNAs) with continuous variation in BMI [23]. Similarly, studies on human sperm also showed that dietary changes result in significant changes in mt-sRNAs in human sperm [27].

These findings suggest that the biological significance of sperm sncRNAs is now well understood and is gaining deeper insights rapidly with their growing linkage with food habbits, diet, exercise, stress, well-being, nutrition, ageing, fertility, infertility, IVF outcomes, and their potential applications in RNA therapeutics. A number of human studies have compared sperm sncRNAs between various comparison groups and identified and emphasized differences in their levels. Notwithstanding the significance of sperm-borne sncRNAs, the composition of the small non-coding RNA repertoire in normal human sperm (snRNAome) and the degree of natural variation across individuals and populations have not yet been elucidated. Some studies have attempted to determine the normal human sperm sncRNAome [28]; however, concrete efforts in this direction are lacking [29]. Considering the significance of sperm sncRNAs in transgenerational inheritance, their potential as sperm fertility/quality biomarkers, and enormous research interest in sperm sncRNAs, it is critical to study their composition, diversity, and variability in human sperm across individuals and populations. We hypothesize that sperm sncRNAome has 'core transcripts' with least variations and 'peripheral transcripts' with high variations. To address this, we analyzed the sperm sncRNA profiles in 54 normozoospermic fertile Indian men, including

repeated sampling from a few donors to address inter- and intra-individual variability. In addition, we also downloaded sncRNA data for fertile/presumptive fertile men from other populations and performed comparisons to understand population-wise similarities and differences in an attempt to define the normal human sperm sncRNAome.

Materials and methods

Ethical approval

This study was approved by the Institutional Human Ethics Committee of the Central Drug Research Institute, Lucknow (CDRI/IEC/2015/A1). The study was conducted in accordance with the declaration of Helsinki guidelines. Written informed consent was obtained from all the participants.

Subject recruitment

Semen samples were collected from healthy, normozoospermic fertile volunteers. The inclusion criteria were as follows: (1) Proven fertility with at least one child, (2) age group 20–55 years, and (3) normal seminogram as per the WHO 2021 criteria. The exclusion criteria consisted of the following: (1) history of a major medical illness, (2) individuals under any kind of medication (antipsychotics, antidepressants etc.), and (3) alcohol/cigarette/tobacco addiction. The subjects were requested to observe abstinence for three days before sample donation. Using this set of inclusion and exclusion criteria, we recruited 54 normozoospermic individuals in the study. All these individuals had a BMI in the 20–25 range and none of them was diabetic. We did not record their dietary habits, which are likely to be significantly different among the participants. To address intra-individual variations in sperm RNAs, we collected repeated samples from some donors after a gap of one month. For the first repeat, only 13 individuals turned up, and for the second repeat only 8 individuals turned up. This way, we collected one sample each from 54 individuals, of whom 13 donated a second semen sample and eight donated a third semen sample as well.

Sample preparation, RNA isolation, and quality analysis

Semen samples were allowed to liquify at 37 °C for 15 min. After liquefaction, the samples were analyzed under a microscope according to the 6th edition of the WHO manual 2021 (https://www.who.int/publications/i/item/9789240030787). The samples were washed with 1X PBS and the cells were collected by centrifugation at 200 g for 15 min in an Eppendorf centrifuge (5804, Benchtop Centrifuge with A-4-44 Rotor). One ml of the somatic cell lysis buffer (SCLB) was added to the pellet, which was resuspended by proper mixing. The SCLB volume was raised to 13 ml, and the tube was incubated on ice for 30 min. The samples were checked under a microscope to ensure the removal of somatic cells, and a final centrifugation at 200 g for 15 min was performed. The MasterPure Complete DNA and RNA Purification Kit (Lucigen, USA) was used for RNA isolation. We followed the instructions given in the kit manual for processing semen samples. Briefly, the sperm pellet was resuspended in 150 µl of the tissue and cell lysis buffer with one µl of proteinase K (50 µg/ml) and incubated in a dry bath at 65 °C for 15 min with repeated vortexing every 5 min. The samples were placed on ice for three minutes, and MPC solution was added for protein precipitation, followed by vortexing for 10 s and centrifugation at 14,000 g for 10 min. The supernatant was aspirated in another Eppendorf tube, and total nucleic acids were precipitated. Total nucleic acids were treated with DNase I enzyme for 15–20 min at 37 °C, followed by the reprecipitation of proteins and RNA precipitation from the supernatant by adding chilled isopropanol. We ran a 2% agarose gel to ensure the complete removal of DNA and performed a bioanalyzer check for RNA quality (Agilent, USA). Since sperm do not have intact ribosomal RNA, a formal RIN calculation was not performed.

Analysis for purity of sperm RNAs

RNA preparation was quantified using a nanodrop spectrophotometer and cDNA was prepared using the high-capacity reverse transcription kit (Applied Biosystems, USA). We checked for the presence of three cell-specific marker transcripts: PRM1 for sperm, CD45 for leukocytes, and CDH1 for immature germ cells. Primer sequences used for these reactions are listed in Supplementary Table 1.

Small RNA library preparation and sequencing

The NEBNext Multiplex small RNA library prep kit (New England Biolabs, USA) was used for library preparation with the total RNA as input (~50 ng per reaction). We followed the protocol provided in the manual. Briefly, the first step was to ligate the 3' adaptor to the RNAs. The adapters were preheated at 70 °C along with 50 ng of total RNA, and a 3' ligation mix was added to the reaction. The reaction was incubated at 16 °C for 18 h to increase the ligation efficiency of modified RNAs (e.g. piRNAs). The next day, RT primers were added to carry out 5' adaptor ligation perfectly to the 5' end. They hybridize with the ligated and free 3' adaptors. dsDNAs are not substrates for the next step ligation enzyme, i.e., T4 RNA ligase 1. The 5' adaptor ligation reaction was incubated for one hour at 25 °C, followed by a reverse transcription reaction. The libraries were amplified over 15 PCR cycles, using index-specific primers for each sample. For size selection of specific small RNA libraries, we ran the libraries on a 6% polyacrylamide gel, and the specific

band corresponding to 140–150 bp size was excised from the gel. The excised band was crushed with RNase-free disposable pestles, and 250 µl of DNA gel elution buffer was added. The tube was rotated overnight at room temperature. The gel debris fluid was passed through cellulose acetate filters (Millipore Sigma, USA) and centrifuged at 16,000 rpm for two minutes. The flow-through was collected, and one µl of linear acrylamide, 60 µl of sodium acetate (pH-5.5), and 750 µl of chilled ethanol were added, followed by overnight incubation at -80 °C for library precipitation. The next day, the tube was spun at 14,000 g for 30 min at 4 °C, and the pellet was air dried and resuspended in TE buffer. The prepared libraries were checked on a 2100 bioanalyzer using a high sensitivity DNA chip (Agilent, USA). Small RNA sequencing was performed using Novaseq 6000 (Illumina, USA) in single end mode with a read length of 60 bp and a depth of 15–30 million reads.

Screening and SRA data download for various populations

We searched PubMed using a combination of keywords such as "human sperm" AND "small RNA", "fertile samples" AND "sperm RNA", "human spermatozoa RNA" to identify studies on normal human sperm sncRNA data from other populations. After screening all outcomes individually, we identified only four small RNA sequencing studies for which raw data were available; two from the USA [28, 30]) and one each from Spain [31] and China [32]. We collected the methodology and demographic details from all studies and summarized them in Supplementary Fig. 1 and Supplementary Table 2, respectively. Raw files were downloaded from the GEO Omnibus database (Krawetz et al., 2011-GSE21191; Pantano et al., 2015, SRP029517; Hua et al., 2019-GSE110190; Morgan et al., 2020-GSE159155), and the data were re-analyzed in the same manner as our samples. We could not get adapter information for Illumina's Small RNA DGE v1.0 kit used in Krawetz et al., (2011) study; therefore, we excluded this study from further analysis.

Small RNA data analysis

The data generated in our lab for Indian samples, as well as the downloaded datasets, were quality checked using the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimgalore was used for quality filtering (q > 28), adaptor trimming (3' adaptor sequence given), and read length selection (19-35nt) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The trimmed files were used for small RNA analysis using the SPORTS1.1 package [33]. The SPORTS1.1 pipeline performs sequential mapping to identify each class of small RNA for both the mapped and unmapped reads in the human reference genome. We considered only hg38 mapped reads (with zero mismatch) and further mapped them sequentially onto database-specific files for each small RNA (downloaded from https://github.com/junchaoshi/sports1.1?tab=readme-ov-file#pre-compile). The sRNA mapping to various biotypes was performed in a sequential manner, as shown in Fig. 1A. The piRNAdb reference file instead of the piRBase file was used for the identification of piRNAs.

The number of unique sequences mapped to each sncRNA biotype is presented in the supplementary data (Supplementary data 1). We collected the read count data generated for all samples, corresponding to the small RNA annotated and unannotated sequences. We removed all unannotated sequences and generated a read count matrix file for all the annotated sequences in each population. We used annotated reads as the total count for each sample and calculated the count per million (CPM) values. Furthermore, the CPM matrix files were used for in-depth analysis of each small RNA type. Principal component analysis was carried out using 'prcomp' function in R, which showed almost perfect clustering of samples in Indian and Chinese studies, while samples in the US and Spain studies were scattered (Supplementary Fig. 2). In addition to SPORTS1.1, we also used the tRNA analysis of Expression (tRAX) tool for the analysis of tRNA derived fragments [34, 35]. The tRAX pipeline was specifically designed to cover the special characteristics of tRNAs, such as the 3'CCA tail and RNA modifications (https://trna.ucsc.edu/tRAX/).

sncRNA characterization in sperm

In small RNA sequencing studies, a ladder pattern is generally observed, where sequences usually differ by 1-5nt from either the 5' or the 3' end. As a result, there is more than one sequence for a particular small RNA. We considered such reads only in the case of miRNAs and summed up all those reads that had the same annotation and their 3'end differed by 1–3 nucleotides in length. We did not perform this exercise for other small RNAs because in the case of piRNAs, different annotations are available for reads differing in length even by one nucleotide, and the nomenclature of the remaining small RNAs as per their fragmentation pattern is not very well defined.

Software and packages used

The software used for small RNA analysis have been described in detail in their respective sections. In addition to this, we used graph pad prism and 'R' packages (version 4.4.0) [36] for data processing and plotting. We used ggplot2 [37], dplyr [38], GenomicAlignments [39], GenomicFeatures [39], Rsamtools [40], and BSgenome.Hsapiens.UCSC.hg38 [41].

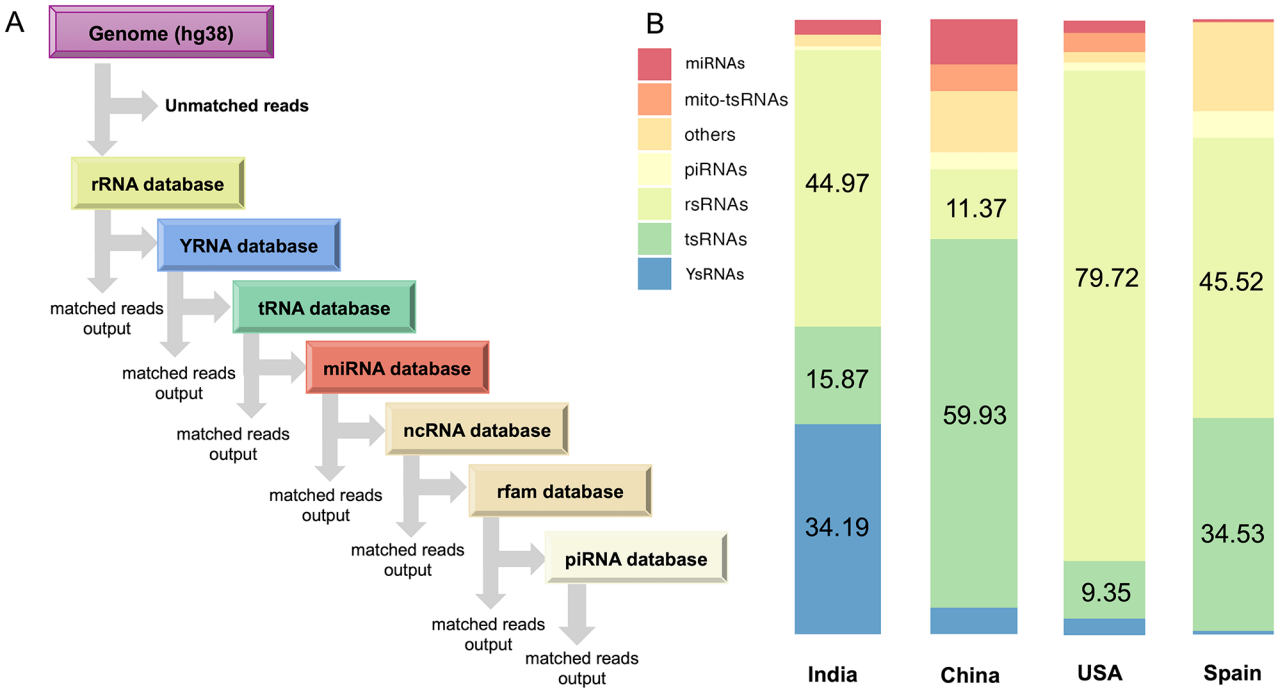

Fig. 1 Sequential mapping of small RNAs to various databases (**A**). Stacked bar chart showing the proportion of various sncRNAs in sperm samples from India, China, USA and Spain (**B**)

Statistical analysis

Descriptive statistics were used to plot the data. The mean values were calculated using Microsoft Excel. Box plots were generated for each miRNA and piRNA to compare their expression levels. The box plots were arranged according to their inter-quartile range for ease of comparison. For comparison of expression, CPM were plotted for various sncRNAs. Log2CPM was occasionally used to plot data over a wide range for ease of presentation. Statistical analysis to compute the level of significance was not performed for any of the sncRNAs analyzed in this study.

Results

Demographic details of the participants

We recruited 54 normozoospermic fertile individuals from Uttar Pradesh, India. The participants represented Indo-European, Dravidian, and Indo Asiatic ethnicities. Each individual had fathered at least one child (time length not recorded), with an average of two children per subject in the pool. The mean age and semen parameters of the participants are shown in Supplementary Table 2. The samples varied in age (21–55 years), and the sperm concentration was between 15 and 150 million/ml. We also collected a second semen sample from 13 subjects (others did not come for repeated donation) and a third semen sample from eight individuals (others did not turn up for the third time), after an interval of approximately one month from the first and second time points.

High purity of sperm RNAs

Somatic cell lysis buffer (SCLB) treated samples were examined under a microscope to ensure the complete removal of somatic cells. PCR amplification from the sperm cDNA preparation showed the presence of *PRM1* and the absence of *CD45* and *CDH1*, suggesting high purity of the sperm preparation (Supplementary Fig. 3).

Ribosomal RNA-derived small RNAs (rsRNAs) are most abundant in spermatozoa

All samples were screened for the composition of sperm sncRNAs. Sequencing data were mapped to all known possible classes of small RNAs, including rsRNA, YsRNA, tsRNA, miRNA, and piRNA in a sequential manner (Fig. 1A). rsRNA were the most frequent biotype in the Indian, US and Spanish samples, but tsRNA constituted the highest proportion in Chinese samples (Fig. 1B). The proportion of other biotypes varied significantly from one population to the other with tsRNA taking the second spot in the US and Spanish samples and YsRNAs taking the second spot in Indian samples (Fig. 1B).

Consistent miRNAs differ across populations

The miRNA reads comprised 2.42% in Indian, 7.37% in Chinese, 1.99% in US, and 0.41% in Spanish samples (Fig. 1B). miRNA analysis showed peaks between 19 and 23nt with 20-22nt being the most prominent peaks across the populations. There were slight differences across populations with the predominant peak at 21nt in

Indian, 22nt in Chinese and US, and 19-23nt in Spanish samples (Fig. 2A). The miRNAs detected in at least 80% of the samples were 47, 128, 38, and 59 in Indian, Chinese, US, and Spanish samples, respectively (Fig. 2B-E). The most and least abundant miRNAs varied across populations (Fig. 2B-E).

Next, we investigated whether repeated sampling from the same individual showed the same set of miRNAs. To address the intra-individual variability in sperm RNAs, we collected a second ($n = 13$) and a third ($n = 8$) semen sample from some of the donors. We found that out of 47, 20 miRNAs replicated the second and third times perfectly with some variations in their levels (perfect category), five miRNAs replicated in the second and third semen samples almost perfectly with an occasional absence (almost perfect category), and the remaining 22 miRNAs failed to show up quite frequently (poor replication category) upon repeated sampling (Fig. 3A and B). The reasons behind the occasional or frequent absence of some of these miRNAs is an interesting question; however, the perfect replication of a good number of them suggests their high potential as sperm fertility biomarkers.

Although approximately 500 miRNAs have been detected in human sperm across various studies, the number of consistent miRNAs in each population is much smaller (47, 128, 38, and 59 miRNAs in Indian, Chinese, US, and Spanish samples, respectively). The comparison of sperm miRNAs across the four populations identified the consistent presence of only 17 miRNAs (detected in at least 80% of individuals), which, in addition to suggesting similarities, also suggests large differences in miRNAs that are present at significant concentrations (Fig. 3C). We checked these 17 miRNAs for their classification into the replication categories (perfect, almost perfect, or poor) as defined above for Indian samples. Fourteen out of the 17 miRNAs belonged to the perfect or almost perfect replication category (Fig. 3C). The replication of these miRNAs across populations and upon repeated sampling (from Indian donors) suggests their strong potential as biomarkers of sperm fertility. Out of these 17 miRNAs, 15 have been shown to be altered in oligozoospermic infertility across one or more studies [12, 42–48].

Sperm piRNAs are chromosome 15 cluster derived

The piRNA reads comprised 0.65% in Indian, 2.73% in Chinese, 1.37% in US, and 4.36% in Spanish samples. For further analysis, we focused on piRNAs that were consistently present in 80% of the samples. Sperm piRNAs across all four populations were in the 19-35nt range, with major peaks at 28-30nt in Indian, 28-31nt in Chinese, 29-31nt in US, and 29-30nt in Spanish samples (Fig. 4A). The expression of various piRNAs differed within and across populations (Fig. 4B-D), affecting the position of piRNAs in the expression plots. Because piRNAs exist in the form of clusters, we looked for clusters of sperm piRNAs. Interestingly, in Indian samples, almost all piRNAs (except one sequence) belonged to one cluster on chromosome 15 (chromosome15:101752465–101774324) (Fig. 5A). For repeated samples from Indian donors, we aligned the piRNA sequences with the human genome and plotted the chromosome 15 cluster. This analysis showed the same pattern of piRNAs, demonstrating reproducibility and mapping to the chromosome 15 cluster (Fig. 5B).

In the case of the Chinese samples, because of the very high number of piRNAs, we used the proTRAC cluster prediction tool [49], which suggested four clusters, of which the major cluster was on chromosome 15 (101752465–101774324), followed by chromosome X (9405386–9414248), chromosome 11 (45683725–45704710), and chromosome 22 (37358565–37369113). In the US samples, we observed only 15 piRNA sequences, of which 8 mapped to the chromosome 15 cluster; the rest of the sequences mapped to diverse regions. In Spanish samples, we observed a very high number of piRNAs, due to which we again used proTRAC for cluster prediction, observing 15 piRNA clusters, including chromosome 15 (101752476–101776323), chromosome X (9405387–9414248), chromosome 22 (37356389–37369214), chromosome 21 (44483822–44490808), chromosome 19 (16037007–16039258), chromosome 18 (11657816–11670763), chromosome 15 (92573335–92609735), chromosome 15 (62219903–62262234), chromosome 15 (51288043–51302890), chromosome 14 (88142904–88155115), chromosome 12 (9327135–9335865), chromosome 12 (3467798–3476185), chromosome 11 (45676547–45719877), chromosome 9 (79865203–79870210), and chromosome 6 (33871622–33914736). Comparison across the four populations showed the consistent presence of piRNAs originating from the chromosome 15 cluster, making it the most important cluster to investigate further with regard to its importance in spermatogenesis, fertility, and embryonic development.

Sperm tsRNAs are majorly contributed by tRNA-Gly-GCC

The tsRNA reads comprised 15.87% in Indian, 59.93% in Chinese, 9.35% in US, and 34.53% in Spanish samples. The majority of sperm tsRNAs-5' were in the range of 30-32nt in Indian, 32-33nt in Chinese, 32-33nt in US, and 29-30nt in Spanish samples (Fig. 6A). The tsRNA-3' were, in general, much lower in frequency (< 2%) in comparison to 5' tsRNAs, which could be due to inadequate coverage of these fragments owing to the 3' modifications in these molecules as suggested previously [50] and/or due to their natural occurrence at relatively low frequencies. The

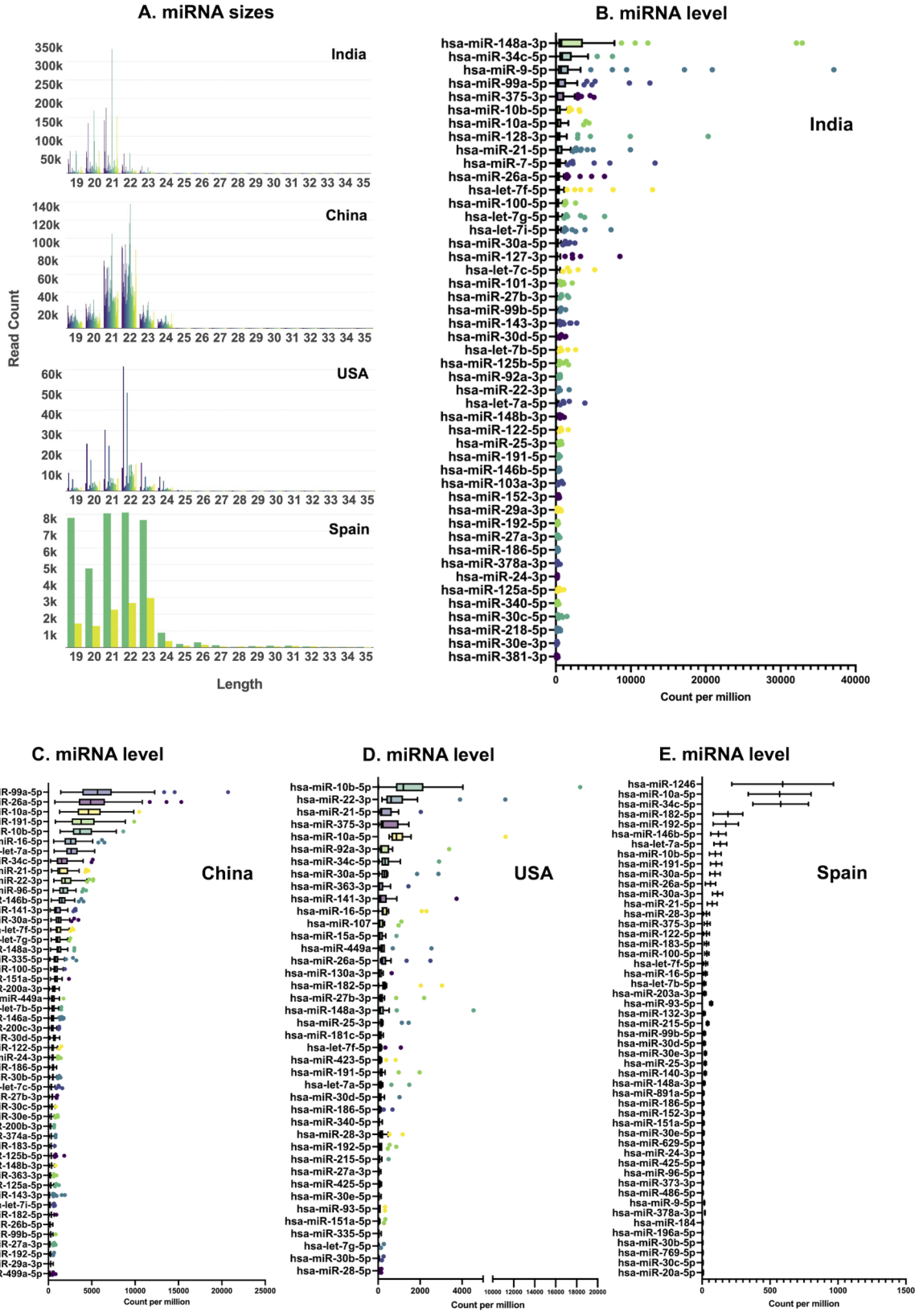

Fig. 2 miRNA sizes across the four tested populations (**A**). The levels of various miRNAs observed in each population (consistent in at least 80% samples) are plotted using box plots and arranged according to the inter-quartile range (**B-E**). A maximum of 50 entities are plotted (remaining miRNAs are listed in supplementary data 1)

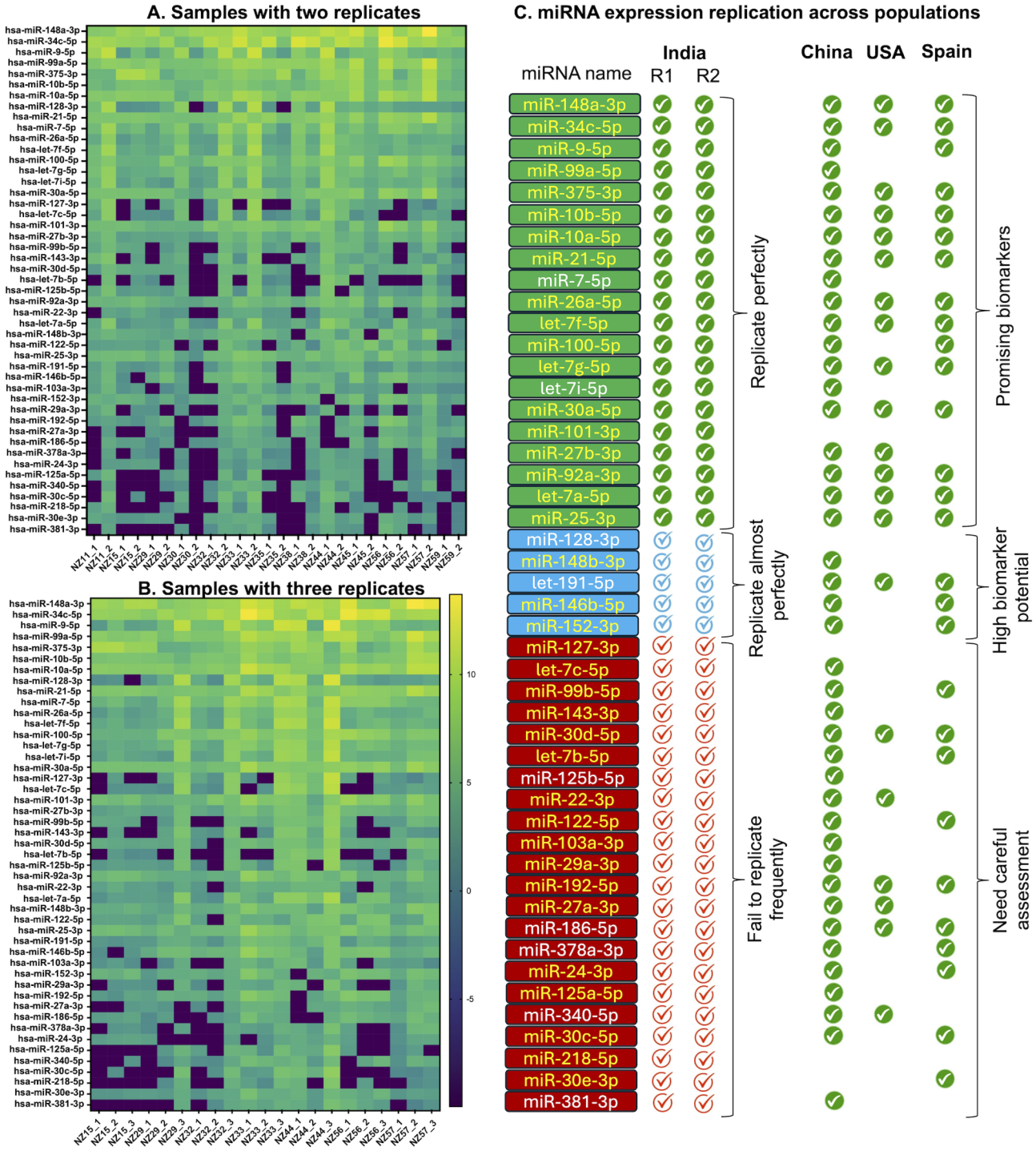

Fig. 3 The level of various miRNAs (log2CPM) across Indian samples with two replicates (**A**) and three replicates (**B**), and the replication of each of the 47 miRNAs observed in Indian samples upon repeated sampling (R1 and R2) and their presence in other populations (**C**). miRNAs in the yellow font have been reported as biomarkers of sperm fertility by previous studies

major peaks for tsRNA-3' were in the range 34-35nt in India and China; however, these peaks were not observed in the USA and Spain. Thus, the major tsRNA peak in the case of tsRNA-5' was slightly shorter than that of tsRNA-3', which is supported by previous observations [51]. We might have missed longer lengths, as we had applied a cut-off of 35 during data analysis. In contrast to tsRNA-5', different tsRNA-3' sizes were more homogeneous in frequency.

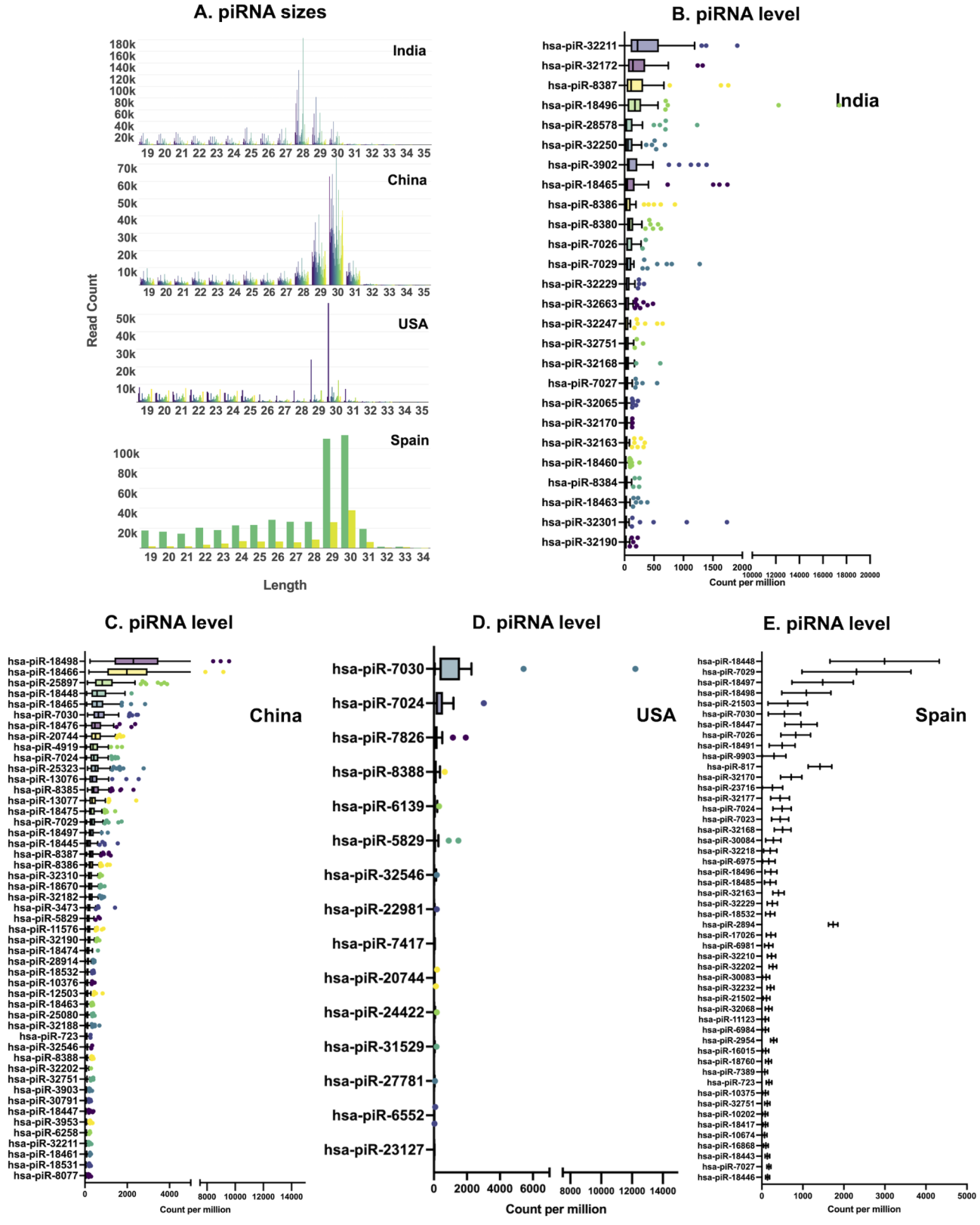

Fig. 4 piRNA sizes across the four tested populations (**A**). The levels of various piRNAs observed in each population (consistent in at least 80% of samples) are plotted using box plots and arranged according to the interquartile range (**B-E**). A maximum of 50 entities are plotted (remaining piRNAs are listed in supplementary data 1)

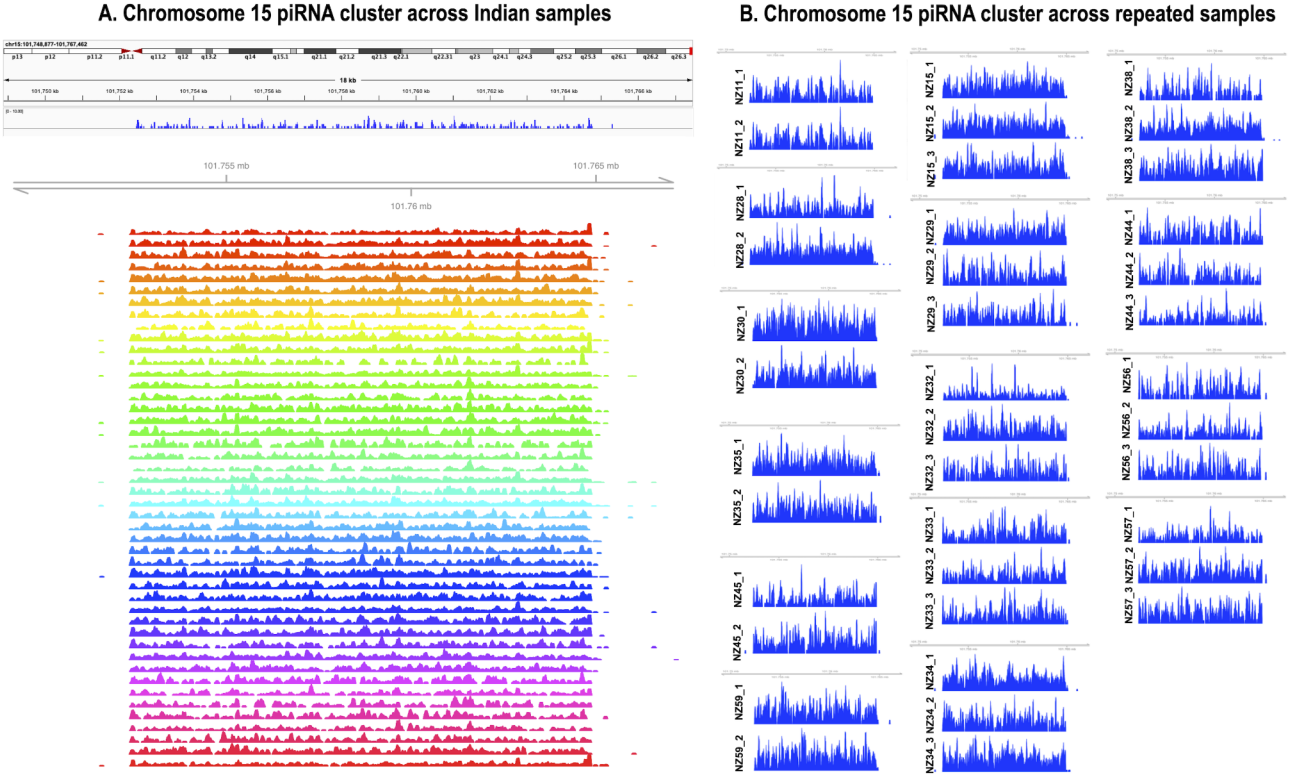

Fig. 5 Almost all piRNAs observed in Indian samples mapped to the chromosome 15 piRNA cluster (**A**) even during repeated sampling for the second and the third time (**B**)

Next, the relative contribution of each tRNA to the total tsRNA pool was examined. We observed that the majority of tsRNAs were contributed by tRNA-Gly-GCC, which consistently accounted for approximately 50% of the total tsRNA pool across all the populations (Fig. 6B). However, the second and the third most significant contributors differed significantly across populations. For example, tRNA-Glu-CTC was the second most common contributor in Indian and Spanish samples, whereas tRNA-Glu-TTC and tRNA-Val-CAC were the second most common contributors in Chinese and US samples, respectively. The next step was to assess the sample-to-sample variations in the tsRNA pool. We observed that tRNA-Gly-GCC remained the top contributor across most of the samples; however, occasionally tRNA-Gly-GCC was taken over by other tRNAs in a few samples (Fig. 6C). These sample to sample variations were observed across all the four populations tested in this study.

Another question was whether all tRNAs were fragmented in all possible ways, i.e., 5' half, 3' half, i-tRF, and 5' and 3' fragments. We observed three major fragmentation characteristics; 1) not all tRNAs fragmented in all possible ways, some fragmented preferentially into 5' half, 5' fragments, and i-tRFs, while others fragmented preferentially into 3' half or 3' fragments, 2) some tRNAs showed specific fragmentation patterns irrespective of the population, 3) some tRNAs fragmented differently across populations. For example, tRNA-Asn-GTT fragmented into specific fragments irrespective of the population, whereas tRNA-Arg-ACG and tRNA-Arg-TCG differed in their fragmentation patterns across populations (Supplementary Fig. 4). tRNA-Gly-GCC and tRNA-Glu-CTC gave rise mostly to 5' halves, 5' fragments, and i-tRFs (Fig. 7). While, tRNA-Glu-CTC did not give rise to 3' halves, tRNA-Glu-TTC gave rise to 3' halves in addition to 5' halves and 5' fragments but did not produce i-tRFs (Fig. 7). This showed a consistent pattern of fragmentation for a particular tRNA, which differed from that of other tRNAs. This consistency in the fragmentation patterns was maintained across populations. For example, tRNA-Glu-TTC gave rise to 3' and 5' halves, but not to i-tRFs, irrespective of the population tested (Fig. 7).

Mito-tsRNAs in sperm are significantly different from nuclear tsRNAs

Mitochondria have their own tRNAs, and similar to nuclear tRNAs, they produce tsRNAs (mt-tsRNAs). Mt-tsRNA reads comprised 0% (consistent in 80% of samples) in Indian, 4.33% in Chinese, 3.13% in the US, and 0.14% in Spanish samples (Fig. 8A). Mt-tsRNAs showed

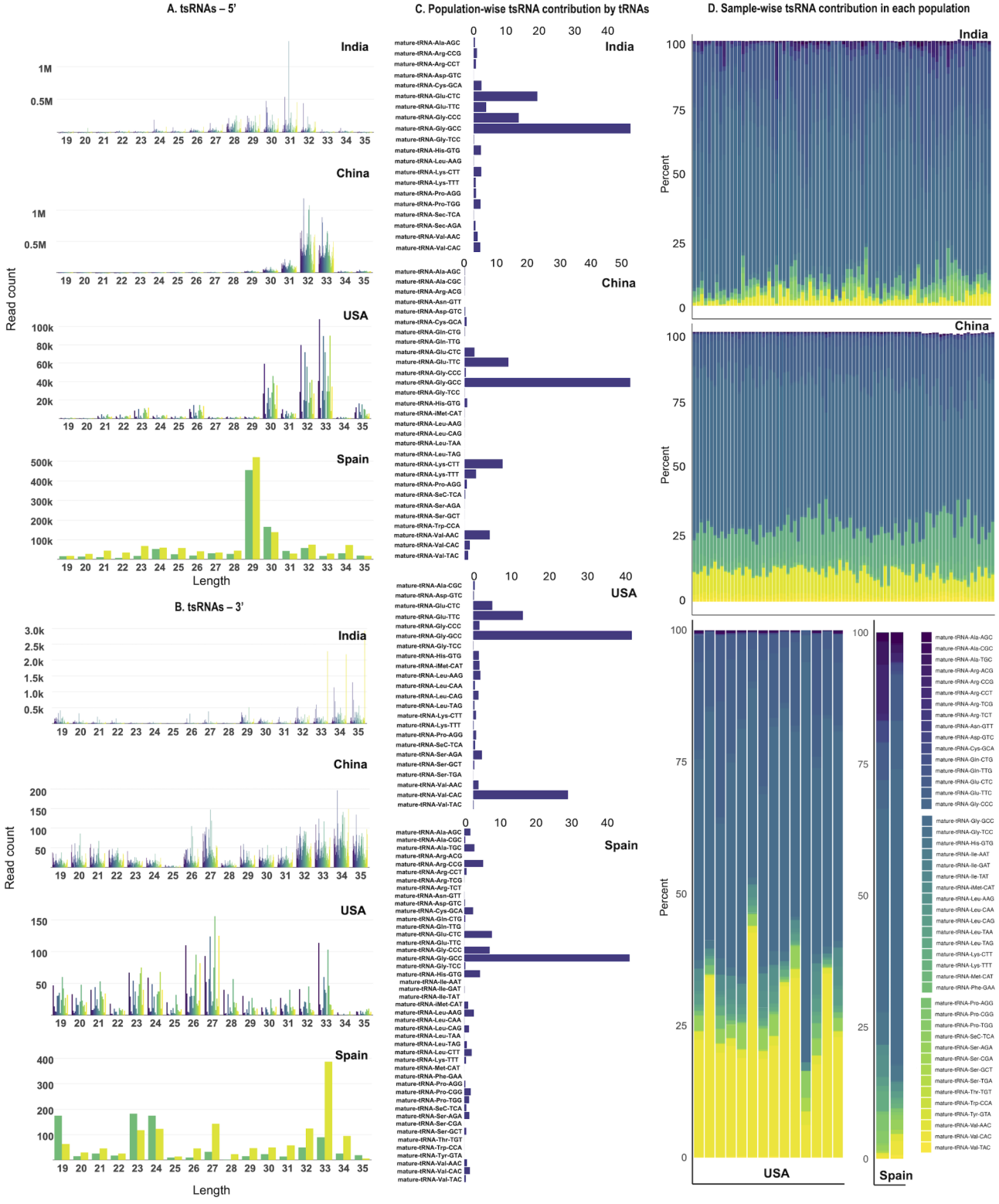

Fig. 6 tRNAs-5′(**A**) and tRNAs-3′(**B**) fragments observed across the four populations. The average percent contribution of various tRNAs to the tsRNA pool in each population (**C**), and the contribution of each tRNA to the tsRNA pool in each sample across the four populations (**D**)

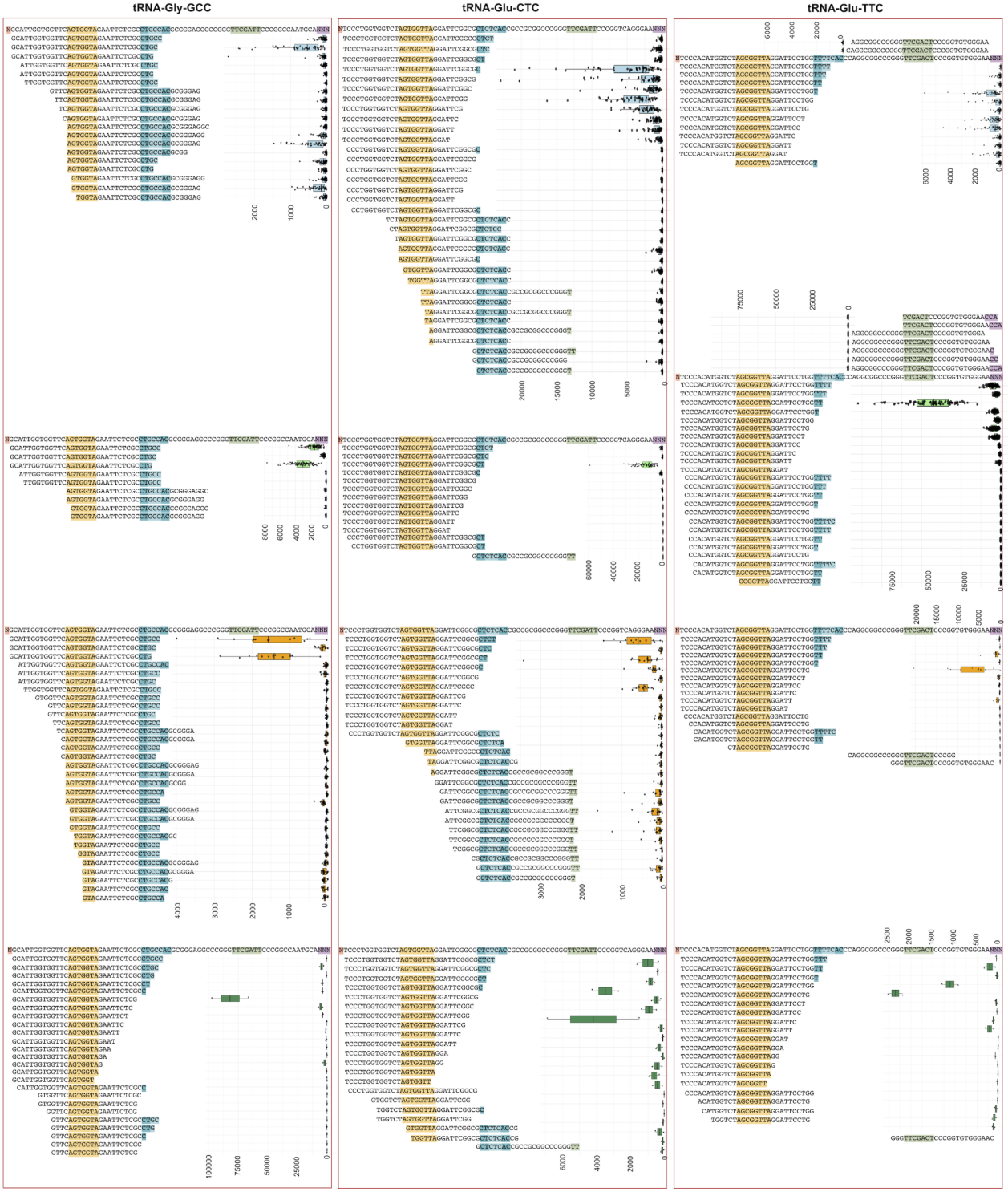

Fig. 7 Representative fragmentation pattern of a few tRNAs showing the generation of various fragments along with their frequencies. This highlights the similarities and differences in the fragmentation pattern of tRNAs. Yellow highlight represents the D loop, turquoise highlight represents the anticodon loop, light green highlight represents the T loop, and purple highlight represents the CCA tail

fragments in the 19-35nt range, but the major peaks were more widely distributed than those of nuclear tsRNAs. In Indian samples, no mt-tsRNA was shared across 80% of the samples; therefore, a percentage cut-off was not applied. mt-tsRNA-5' showed major peaks at 19-25nt in Indian, at 28-33nt in Chinese, at 29-33nt in the US, and were wide-spread (23-35nt) in Spanish samples. However, these data did not show a consistent pattern across

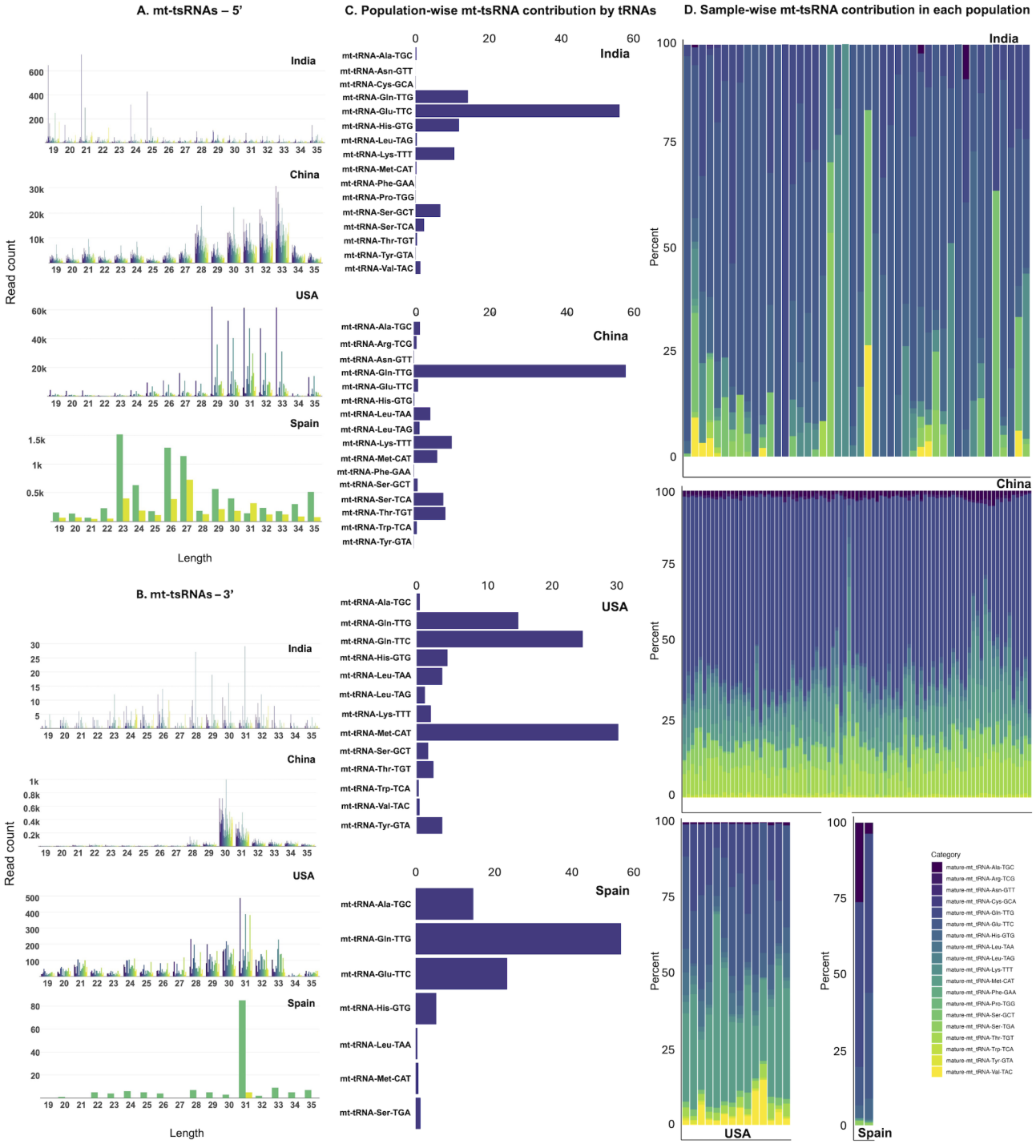

Fig. 8 mt-tRNAs-5′ (**A**) and mt-tRNAs-3′ (**B**) fragments observed across the four populations. The average percent contribution of various mt-tRNAs to the mt-tsRNA pool in each population (**C**), and the contribution of each mt-tRNA to the mt-tsRNA pool in each sample across the four populations (**D**)

populations, which may be due to the inadequate coverage of mt-tsRNAs as a result of technical differences. In the case of mt-tsRNA-3′, the range was 19-31nt, with major peaks at 28-31nt in Indian, at 30-31nt in Chinese, at 28-33nt in US, and at 31nt in Spanish samples.

In contrast to nuclear tsRNAs, which predominantly originate from tRNA-Gly-GCC across all populations, the top contributor to mt-tsRNAs differed across populations (Fig. 8B). For example, the top contributors to mt-tsRNA were tRNA-Glu-TTC in Indian, tRNA-Gln-TTG in Chinese, tRNA-Met-CAT in US, and tRNA-Gln-TTG

in Spanish samples. Similarly, the second most common contributor also differed among the four populations. The top contributors to mt-tsRNAs were completely different from those of the nuclear tsRNAs. Regarding similarity, both nuclear and mitochondrial tsRNAs had a significant contribution from tRNA-Glu-TTC, whereas all other tsRNAs differed significantly between the nucleus and the mitochondria. In addition, the mt-tsRNA pool composition differed across individuals, with the top contributor making up to 90% of the mt-tsRNAs in certain samples, which were occasionally replaced by other mt-tsRNAs (Fig. 8C). For example, tRNA-Glu-TTC contributed more than 90% of the tsRNA pool in most Indian samples and less than 10% in a few samples, whereas tRNA-Gln-TTG contributed up to 60% in certain Chinese samples and less than 20% in a few samples (Fig. 8C). Similar sample to sample variations were also observed in the US and Spanish samples as well.

Further investigation on the origin of mt-tsRNAs suggested that not all mt-tRNAs gave rise to 5'-tRF, 3'-tRFs, or i-tRFs. Instead, different mt-tRNAs preferentially produce certain types of mt-tsRNAs (Fig. 9). Only rarely, did a particular mt-tRNA give rise to all three major types of tsRNAs, including 5'-tRF, 3'-tRF, and i-tRF. With regard to differences, mt-tRNA-Gln-TTG-1-1 gave rise to 5'-halves, 5' fragments, i-tRFs, and 3' fragments in Indian samples, but in Chinese samples, it gave rise to 5' halves and 5' fragments only. Regarding similarities, mt-tRNA-Glu-TTC-1-1 produces only i-tRFs across all populations tested.

rsRNAs in sperm are largely derived from 28S and 18S rRNAs

The rsRNA reads comprised 44.97% in Indian, 11.37% in Chinese, 79.72% in US, and 45.52% in Spanish samples (Fig. 1). Depending on their origin, rsRNAs can be classified into five types; 5S, 5.8S, 18S, 28S, and 45S [33]. Since we extracted the library band of 140-160nt for our samples, we observed rsRNAs 19-35nt long that originated from all five types of rRNAs (5S, 5.8S, 18S, 28S, and 45S). Across all the populations, the most abundant rsRNAs were generated from 28S, followed by 18S, 5S, 5.8S, and 45S-derived rsRNAs (Fig. 10A). Previous studies have also shown that rsRNAs derived from 18S to 28S are most abundant in human sperm [24]. We observed different fragmentation patterns across all rRNA types, with 18S and 28S rRNAs fragmented into all sizes (19-35nt) across all populations with a decreasing frequency of longer fragments (Fig. 10B). In contrast, 5S rRNA, 5.8S rRNA, 45S rRNA, and occasionally 28S rRNA showed a fragmentation pattern in the form of certain peaks (Fig. 10B). Interestingly, the fragmentation peaks originating from these rRNAs differed across populations. For example, 28S-derived rsRNAs showed peaks at 19nt and 34nt in Indian, at 21nt, 31nt and 35nt in Chinese, no specific peak in the USA, and at 19-21nt in Spanish samples (Fig. 10B). Similarly, 5.8S-derived rsRNAs showed peaks at 27nt in Indian, at 22nt, 26nt, 27nt, 31nt in Chinese, at 19nt, 25nt, 27nt, 35nt in the USA, and at 25-27nt in Spanish samples. The 5S-derived rsRNAs showed peaks at 32-35nt in India, at 31nt in China, at 23-24nt in the USA, and at 31nt in Spain. 45S-dervied rsRNAs showed peaks at 22nt, 23nt, 35nt in Indian, at 19nt, 20nt, 23nt, 32nt in Chinese, at 29-31nt and 35nt in the USA, and at 19nt, 21nt, 29-31nt in Spanish samples (Fig. 10B). The contribution of various rRNAs to rsRNAs varied from sample to sample in each population (Fig. 10C). Mitochondrial origin rsRNAs (12S and 16S) represented a minor proportion. The minor representation of 12S and 16S rsRNAs in human sperm has been reported previously [24].

YsRNA in sperm are largely Y4RNA derived

The YsRNA reads comprised 34.19% in Indian, 4.28% in Chinese, 2.72% in US, and 0.62% in Spanish samples (Fig. 1). Among all the YsRNAs, Ys4RNAs were the most frequent, followed by other YsRNAs, the percentage of which differed significantly across populations (Fig. 11A). For example, Ys1RNAs were the second most frequent in India and the USA, but Ys3RNAs were the second most frequent in China, and Ys5RNAs were the second most frequent in Spain (Fig. 11A), although the fraction of each type of YsRNA differed significantly across the populations. Regarding the order of the YsRNA frequencies, India and USA showed the same pattern (Fig. 11A). Regarding the size of YsRNAs, Ys1RNAs showed peaks at 31nt in Indian, at 32nt in Chinese, at 30-33nt in the US, and at 31nt in Spanish samples (Fig. 11B). In the case of Ys3RNAs, peaks were observed at 29-30nt in Indian, at 30-32nt in Chinese, at 30nt in Spanish samples, but the US samples did not show a specific peak. For Ys4RNAs, Indians showed peaks at 30-31nt, Chinese and US showed peaks at 31-32nt, and Spanish showed a peak at 31nt. Ys5RNAs showed a wide fragmentation pattern with peaks spread in the 19-34nt range across all populations, with two major groups of peaks, one around 20-23nt and the second around 29-33nt (Fig. 11B). The contribution of various YRNAs to YsRNAs varied from sample to sample in each population (Fig. 11C).

Discussion

sncRNAome composition in normal human sperm

Almost all sncRNA biotypes were observed across all populations. rsRNAs were the most frequent biotype in Indian, US and Spanish samples, whereas tsRNAs were the major biotype in Chinese samples. In an independent study, Nätt et al., (2019) reported rsRNA as the major biotype in human sperm. It has been emphasized

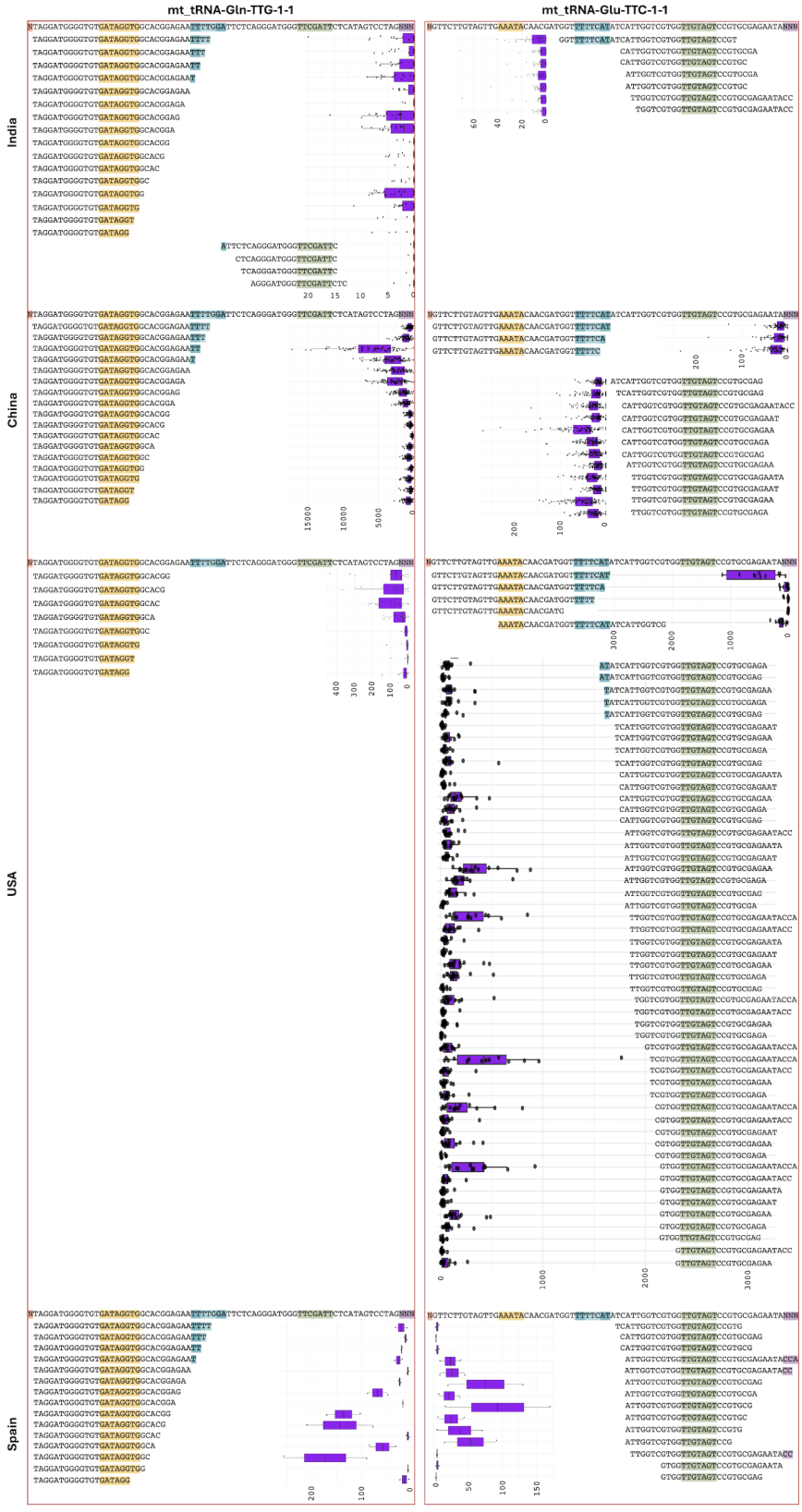

Fig. 9 Representative fragmentation pattern of a few mt-tRNAs showing the generation of various fragments along with their frequencies. This highlights the similarities and differences in the fragmentation pattern of mt-tRNAs. Yellow highlight represents the D loop, turquoise highlight represents the anti-codon loop, light green highlight represents the T loop, and purple highlight represents the CCA tail

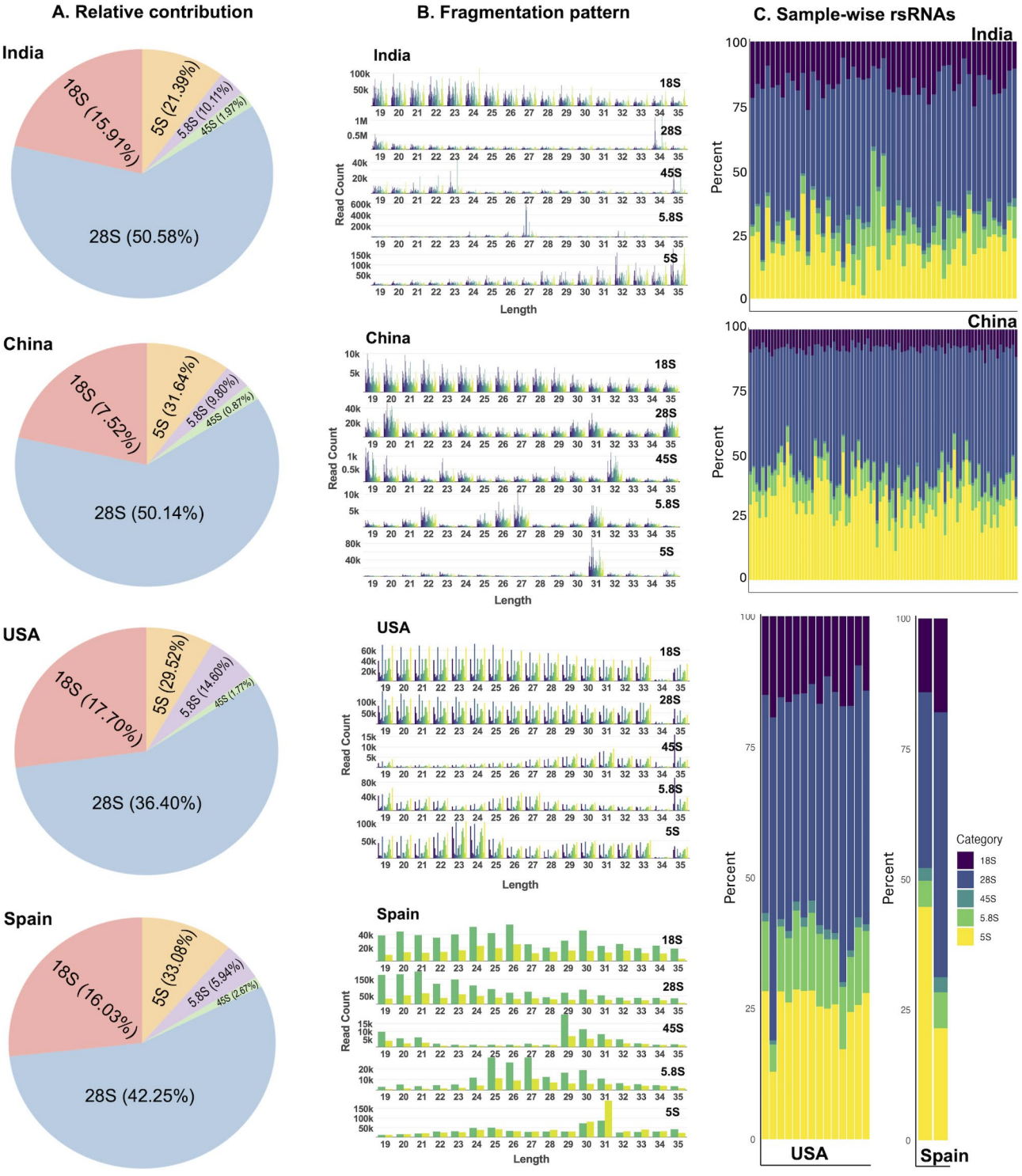

Fig. 10 The average contribution of various rRNAs to rsRNAs in the four populations tested (**A**), fragmentation pattern of various rRNAs into rsRNAs (**B**), the relative contribution of each rRNA to rsRNAs in each sample across the four populations (**C**)

that various modifications to their parental RNAs affect the fragmentation pattern and the generation and functions of sncRNAs [52]. The proportion of each sncRNA biotype in the sperm sncRNAome differed significantly across individuals and populations, which could be due to technical, environmental or biological factors. Technical issues include sperm preparation, RNA isolation, library preparation, and other experimental variables. Recent studies have shown rapid and significant changes in sperm sncRNAs owing to dietary changes [12], making

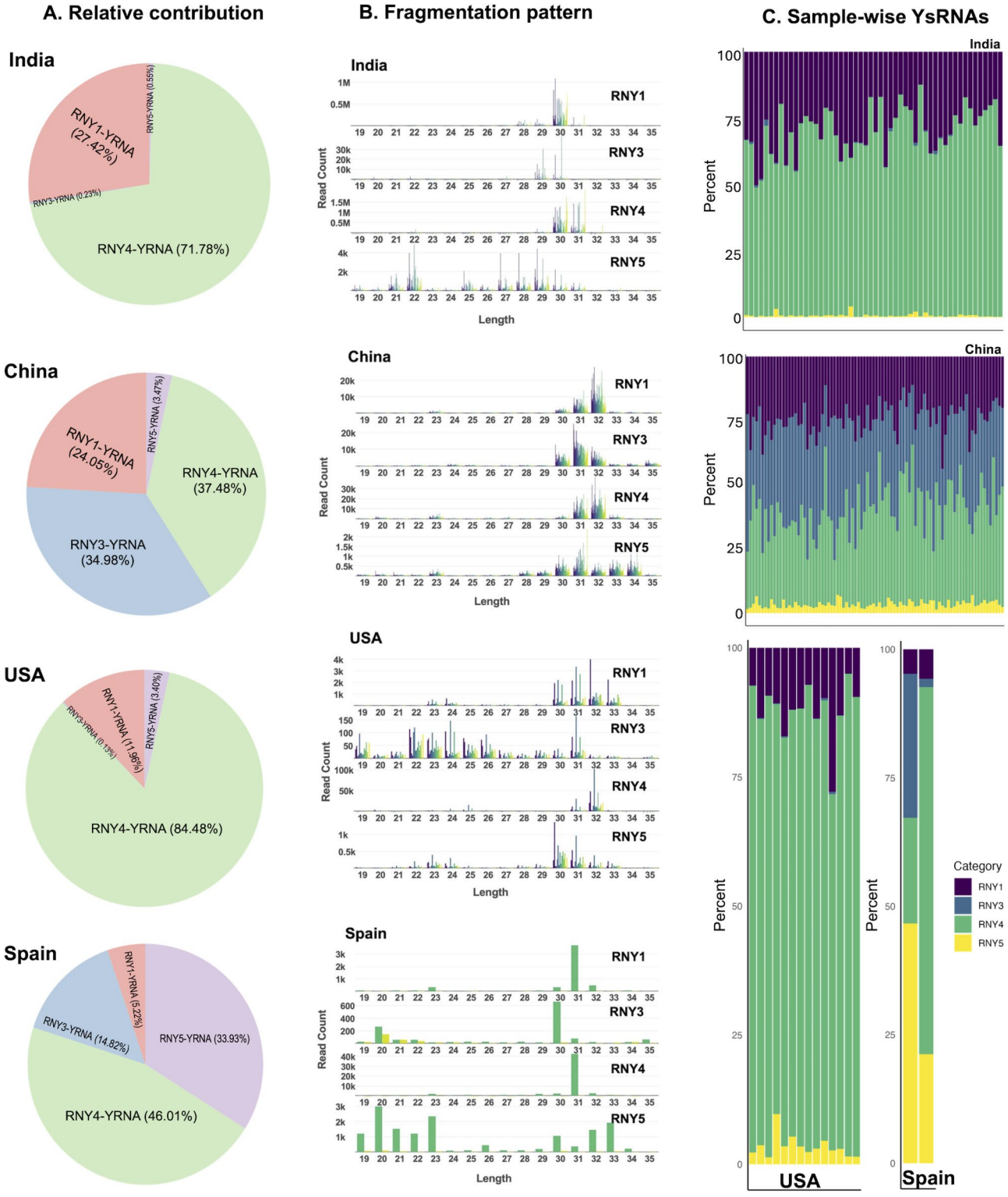

Fig. 11 The average contribution of various YRNAs to YsRNAs in each population (**A**), fragmentation pattern of various YRNAs into YsRNAs (**B**), the relative contribution of each YRNA to YsRNAs in each sample across the four populations (**C**)

environmental factors significant contributors to these variations. Among biological factors, natural biological variations as seen in most biological variables could account for certain variations. With changes in technology and sequencing methods, sncRNA biotypes may see further changes in their relative abundance in the normal human sperm sncRNAome in the future.

Consistent miRNAs in sperm are few

miRNAs constitute a minor fraction (0.41–7.37%, average: 3.04%) of the sncRNAome, but they have been investigated far more in-depth than other sncRNAs. miRNAs showed major peaks at 20-22nt with slight variations across populations; however, the biological significance of these size variations in miRNAs (1-3nt) has not yet been investigated. The different numbers of miRNAs detected across the four populations suggest significant variations in their levels across samples (Fig. 2). Out of more than 500 miRNAs reported in human sperm till date, less than 100 were consistent across samples in each of the four populations and only 17 were consistent across populations, demonstrating high variability and a low level of consistency (Fig. 3). Sperm miRNAs have long been sought as biomarkers of fertility. The foremost requirement for this is their consistent presence in the sperm across individuals and populations. We checked sperm miRNA consistency by repeated sampling from Indian donors and comparison across populations. We found that out of 47 miRNAs, 25 showed perfect or almost perfect replication in repeated sampling. Further comparisons across populations showed that only 17 miRNAs were consistently present in human sperm. A review of the literature showed that close to 400 miRNAs have been considered to have male fertility biomarker potential [8, 12, 53, 54]. Of the 47 miRNAs detected in Indian samples, almost all have been reported as sperm fertility biomarkers in various studies [7, 12, 42–48, 55, 56]; however, we found that half of them are poorly reproducible (Fig. 3). Further comparisons across populations showed that only 17 were consistently present in sperm and only 14 of these were reproducible upon repeated sampling from the same individual. We believe that miRNAs belonging to the reproducible categories have a higher potential as sperm fertility/quality biomarkers than those belonging to the less reproducible categories. Other miRNAs are present either at low levels or show high variability in response to environmental cues, making them ideal for investigating in transgenerational inheritance.

Sperm miRNAs are important for spermatogenesis or embryonic development

Seventeen miRNAs shared by sperm across populations may be critical for spermatogenesis or sperm fertility. A literature search showed that 11 of the 17 miRNAs are important for testis development, spermatogenesis, post-fertilization development or to have value as biomarkers of sperm fertility. miR-375 participates in testicular development [57], miR-10b is important for Sertoli cell maturation [58], and miR-34c-5p [59] and miR-92a-3p [60] play roles in spermatogenesis. Let-7 family members are expressed in type B spermatogonia and primary spermatocytes [61], and miR-30a-5p is known to promote spermatogonial stem cell differentiation [62]. miR-21 is critical for the spermatogonial stem cell population [63], plays a role in sperm capacitation [64], and regulates spermatogonial to spermatocyte differentiation [61]. Let-7 g-5p is testosterone responsive [65] and regulates apoptosis during heat shock [66]. miR-191 is expressed in spermatocytes and round spermatids [59]. miR-186-5p is present in most germ cells and is most abundant in spermatozoa [67]. Although their participation in spermatogenesis is not surprising, their consistent presence in human sperm may be linked with their roles in sperm fertility or embryonic development, which is of significant interest. Only two of these, miR-191-5p [68] and miR-34c-5p [7], have been shown to affect embryo quality and development, respectively.

Sperm piRNAs are low but highly specific and mysterious

piRNAs constitute a minor fraction (0.65–4.36%; average: 2.28%) of the human sperm sncRNAome. piRNAs showed major peaks at 28-31nt across all populations with minor differences in size and frequency. piRNAs were found to be the most abundant biotype in human sperm in only one study [31]. Considering their percentage share, piRNAs cannot be the top contributors to the human sperm sncRNAome. Many of these pseudogene-derived piRNAs have been suggested to target their parental genes in the testis, but their significance in human sperm remains unknown [31]. piRNAs are mysterious with regard to their reduction in mouse caput sperm and reappearance in caudal sperm [4, 69]; however, this definitely suggests their biological significance in sperm. It has been postulated that sperm can produce piRNAs in situ [69], the significance of which remains unclear. We observed that almost all piRNAs in Indian samples originated from the chromosome 15 cluster, while sperm samples from other populations predominantly contained chromosome 15 cluster piRNAs. This is a unistrand (+) pachytene spermatocyte cluster with 15-q26-7499.1 gene annotation having two exons [70] and genomic annotation corresponding to dynamin 1 pseudogene 47 (*DNM1P47*). Previous studies have also suggested that testicular piRNAs are pseudogene-derived [31]. The origin of all sperm piRNAs from the chromosome 15 cluster suggests their specificity (non-random nature) and biological significance that requires further investigation. The first study investigating the functional

importance of sperm piRNAs was recently conducted in Drosophila, identifying distinct sperm piRNA cargoes capable of altering gene expression in the offspring to mediate intergenerational metabolic reprogramming (IGMR), leading to obesity in their offspring [71]. Nevertheless, it remains to be determined whether piRNAs have similar effects in humans. The chromosome 15 piRNA cluster awaits further investigation for its potential significance in spermatogenesis, embryonic development, or transgenerational inheritance.

tsRNAs correlate with sperm quality and mediate transgenerational inheritance

tsRNAs constituted 10–60% (average: 30%) of the sncRNAome and occupied the top spot in Chinese, the second spot in USA and Spanish, and the third spot in Indian samples. tsRNAs-5' showed certain peaks, but tsRNA-3' showed even spread across the 19-35nt range. It would be interesting to investigate whether variations of 1-3nt in tsRNAs exert significant biological effects. The relatively lower frequency of tsRNA-3' in comparison to tsRNA-5' could be due to their lower coverage owing to their 3' modification [50], but it may be attributed to their natural occurrence at a relatively lower frequency. Interestingly, tsRNA derived from tRNA-Gly-GCC accounted for up to 50% of the tsRNA pool across all populations, although the second major contributor was different. While tRNA-Gly-GCC remained the major tRNA contributor in most of the samples, its contribution varied from sample to sample, with another tRNA becoming the major contributor in a few samples. Studies on transgenerational inheritance have largely focussed on the total level of tsRNAs in the sperm sncRNAome. Considering the significant variations in the contribution of individual tRNAs to the tsRNA pool, future studies should also investigate the changes in the composition of the total tsRNA pool. tsRNA size differences across individuals and populations could be due to various biological or technical reasons; however, their significant presence in human sperm suggests their biological significance. tRFs are known to act as mediators of various significant biological processes, including translation, apoptosis, cell survival, antiviral defence, response to starvation, nutrient deficiency, and oxidative or other stresses [72]. The significance of tsRNAs in human sperm has been studied by a few [32, 73]. Hua et al., (2019) reported the impact of ten sperm tsRNAs on embryo quality in IVF [32]. Another study on human sperm reported that tRNA-Gln-TTG derived tsRNAs were significantly associated with sperm quality [73].

In addition to their significance in embryonic development, tsRNAs have also been found to mediate transgenerational inheritance. A study on human sperm from lean and obese men found that the expressions of specific miRNAs, piRNAs, tsRNAs, and snRNAs were significantly altered in the obese group [74]. These findings have been corroborated by experiments in mice where a high fat diet modified the levels of sperm tsRNAs, which showed an intergenerational effect [75]. In a similar study, a high fat diet in mice resulted in an increased level of 5′tsRNA-Gly-GCC in mature sperm, resulting in enhanced gluconeogenesis in the offspring [20]. Studies on human sperm have shown significant changes in sperm tsRNAs with age [76]. Similarly, studies in mice have shown that advanced paternal age results in significant changes in sperm tsRNAs, which contribute to neuropsychiatric disorders that are intergenerational [77]. High inter-individual variability in the human sperm tsRNA profile will facilitate accurate interpretation of tsRNA data in studies on transgenerational and intergenerational inheritance.

Sperm mt-tsRNAs correlate with motility and are highly sensitive to diet

The mt-tsRNAs constitute a minor fraction (0-4.33%, average: 1.90%) of the human sperm sncRNAome. Earlier studies did not pay much attention to sperm mt-tsRNAs. The comparison of mt-tsRNA-5' across other populations suggested a different pattern of mt-tsRNAs in their samples, with 28-35nt representing the major peaks in the US and Chinese samples and 23-27nt representing the major peaks in Spanish samples. In Indian samples, the coverage of mt-tsRNA-5' was poor in comparison to other populations; therefore, we consider others data more reliable in this case. The abundance of 5'-derived mt-tsRNAs was significantly higher than that of 3'-derived mt-tsRNAs; however, the abundance of all mt-tsRNAs in general was much lesser than that of the nuclear tsRNAs. Mt-tsRNAs followed the rule of top contribution by one tRNA; however, unlike nuclear tsRNAs, the top contributor differed across populations. Similar to the nuclear tRNAs, different mt-tRNAs differ in their fragmentation pattern. Mitochondria are specifically enriched in spermatozoa, making them important contributors to the overall repertoire of sncRNAs in the sperm. Since mitochondria regulate critical processes such as apoptosis and energy production, the role of mt-tsRNAs may be far more important than understood. Furthermore, there is the possibility of the leakage of mt-tsRNAs into the cytoplasm, which suggests their impact on cellular functions at a much broader level [77]. mt-tRNAs are slightly different from nuclear tRNAs, and the mitochondrial environment is more prone to oxidative stress and mutations, which can affect mt-tRNA modifications and ultimately their fragmentation. This suggests that mt-tsRNAs are even more dynamic than nuclear tsRNAs.

Mt-tsRNAs appear to be highly sensitive to metabolism and diet. A recent study showed that paternal exposure

to a high-fat diet resulted in the upregulation of 5' fragmentation of mt-tRNAs in mouse spermatozoa, which were transferred to the oocyte and altered gene expression [23]. Previous studies on Drosophila [27] and human [24] sperm have also reported upregulation of mt-tsRNAs in response to a short high-sugar diet. An analysis of sncRNAs in young, healthy Finnish men showed little variation in n-tsRNAs with BMI, but a major change in mt-tsRNAs with continuous variation in BMI [23]. Similarly, studies on human sperm have also shown that dietary changes in humans result in significant alterations in mitosRNAs [27], which positively correlates with total sperm number and motility [27]. Re-analysis of sncRNA data from Hua et al., (2019) [32] also showed a significantly higher level of mt-tsRNAs in the high-quality embryo group [27]. A previous study on Drosophila sperm found that the 5' halves of nuclear tRNA were the dominant biotype, whereas in the case of mt-tsRNA, i-tsRNAs were the most common [27]. In response to dietary changes, i-tsRNAs change most significantly [27]. Specific fragmentation patterns of various mt-tRNAs and differences in their relative contribution to mt-tsRNAs suggest that such studies should also consider changes in the fragmentation pattern of each mt-tRNA and changes in the relative contribution of various tRNAs. The above studies suggest that mt-tsRNAs can be used as sperm quality biomarkers, in addition to their value as transgenerational carriers. Notably, n-tsRNAs are acquired by spermatozoa from epididymosomes or cytoplasmic droplets carried by spermatozoa [4, 78, 79], whereas mt-tsRNAs are generated inside the spermatozoa.

Sperm rsRNAs are abundant but least explored

rsRNAs constituted 11–80% (average: 45.39%) of the human sperm sncRNAome. rsRNAs constituted the top fraction of the sncRNAome in Indian, US and Spanish samples and occupied the second position in Chinese samples. Depending on their origin, rsRNAs can be classified into five types: 5S, 5.8S, 18S, 28S, and 45S [33]. The functional significance of miRNAs, piRNAs, and tsRNAs is somewhat known; however, the same remains to be understood for rsRNAs. The rsRNAs derived from all rRNA types were distributed in the 19-35nt range. Among the rsRNAs, we observed that the maximum reads belonged to 28S, followed by 5S, 5.8S, 18S, and 45S in decreasing order. The maximum contribution of 28S RNA-derived rsRNA reads to human sperm was previously established by the Chinese study included in our population-wise comparison [32]. rsRNAs from 18S to 28S have been shown to be the most abundant in human sperm [24]. A recent study on mouse sperm also observed differential regulation of mt-rsRNAs in response to a high-fat diet [23]. In addition to the studies included in the population-wise analysis presented herein, studies specifically on rsRNAs in human sperm are scarce. In the only other study on human sperm rsRNAs, Chu et al., (2017) compared rsRNA-28S between leukocytospermia and normal human sperm samples and identified that rsRNAs derived from 28S rRNA were diminished or absent in the leukocytospermic infertile group [80]. This study suggests that rsRNAs may have a functional significance in infertility. We observed a minor proportion of mitochondrial rsRNAs (12S and 16S derived), which is supported by previous observations [24]. Although rsRNAs in human sperm occupy the top spot, they have been the least explored with regard to their functional significance.

Sperm YsRNAs are unique and least explored

YsRNAs constitute a significant fraction (0.62–34.19%, average: 10.45%) of the human sperm sncRNAome but have been least explored. 'Y' RNAs were originally observed in ribonucleoprotein complex formation with Ro and La proteins. Due to their presence in the cytoplasm, they were given the name 'Y' RNAs [81]. Y-RNAs are typically 83-112nt in length. In humans, there are four YRNA loci, hY1, hY3, hY4, and hY5, located in a cluster on chromosome 7 that encodes YRNAs. YRNAs are thought to be involved in chromosomal DNA replication in response to UV damage [82, 83], but their exact significance remains unknown. Although the functional significance of YRNAs themselves remains unknown, high-throughput sequencing studies have identified YRNA fragments to be highly abundant in cells, tissues, and body fluids and tumors in humans and mammals [84–86]. A comparison across populations showed Ys4RNAs to be the most abundant, whereas the second most common contributor differed. The highest frequency of Ys4RNAs and Ys1RNAs was also reported in a previous study on German samples [87]. According to this study, the length distribution of Ys1RNAs peaks at 30nt and that of Ys3RNAs peaks at 31nt [87]. 5'-YsRNAs were observed in a significant proportion in sperm heads, whereas 3'-YsRNAs were almost absent in sperm heads. Interestingly, Ys1RNAs were almost completely absent in the oocytes, suggesting their major contribution to the zygote by sperm [87].

YsRNAs have been least explored, and we found only one study on YsRNAs in the human testes and sperm [87]. This study included the analysis of both sperm and all germ cell developmental stages in the testis and observed that YsRNAs were absent in testicular germ cells and made a sudden appearance in sperm heads. This study further investigated and observed that sperm YsRNAs were largely derived from the epididymis in the form of semen exosomes. Therefore, YsRNAs are a part of the soma-to-germline delivered cargo. They have been proposed to play a role in early embryonic development [87];

however, the functional significance of these sncRNAs remains unknown. Considering the similarities in size range and 3' modifications, the study concluded that YsRNAs are functional piRNAs that are not encoded by piRNA clusters [87]. A previous study on human seminal exosomes identified the presence of a significant proportion of YRNAs and YsRNAs, which were suggested to be delivered to sperm [86]. All four YRNAs were found to give birth to YsRNAs in a distinctive pattern, with >95% of the Ys1, Ys3, and Ys5 mapped to the 5' ends, and in the case of Ys4RNA, approximately 80% mapped to the 5' end and 20% mapped to the 3' end. This suggests a selective generation of YsRNAs [86], implying their biological roles. Our study also showed a distinctive pattern in Ys1, Ys3 and Ys4 compared to that in Ys5. It appears that YsRNAs make up a significant proportion of the sperm sncRNAome; however, their significance remains almost entirely unknown.

Sperm sncRNAome has a pattern and is biologically meaningful

For all sncRNA biotypes, the consistent fragmentation pattern suggests them to be non-random. The detected numbers of small RNAs are much fewer than the theoretically possible number, suggesting that they are specifically generated that way. Further, sncRNAs are not generated randomly from all the regions of their parental RNAs. For example, some tsRNAs are specifically generated from the 5' half, whereas others are generated from the 3' half. The selective retention of one strand to generate miRNAs and the selective retention of one half to generate tsRNA has been identified, which may apply to other sncRNAs as well, further endorsing specific mechanisms of their generation and linkage with their biological significance. Thus, there is evidence coming up to suggest that these sncRNAs are not randomly generated functionless entities on their way to elimination, but have biological roles to serve depending on the cell/tissue type. Interestingly, all functions of sncRNAs are not sequence complementarity-dependent, but scaffolding and aptamer-like functions have also been suggested and established [52]. The breakage of larger RNA pieces into smaller units unveils new possibilities of function due to their size, three dimensional structure, loss of self-complementarity of certain sequences, and breakage of scaffolding domains, expanding the possibility of their functions tremendously. Although immense possibilities exist, their functional significance has just begun to be uncovered. The amenability of these small RNAs to various structures and their potential to bind to various proteins provide them with immense avenues for exerting their biological functions way beyond RNAi-dependent mechanisms.

Limitations and cautions

The primary aim of this study was to identify similarities across samples and populations in the sperm sncRNAome. The observed similarities that withstood significant variations in population types and experimental variables suggest the robustness of certain sncRNAome features; nevertheless, we suggest interpreting the differences with caution. sncRNAs, such as rsRNAs and tRFs carry a 2′, 3′ cyclic phosphate, or a 3′ phosphate, which hinders capturing these sncRNAs during library preparation [50]. Apart from the above restriction in the representation of sncRNAs in library preparation, other technical issues that can affect their representation include variations in sperm collection and RNA extraction methods, which vary significantly across studies [4, 88, 89]. For example, treating sperm with detergents for RNA extraction can affect the yield of RNA by 2–3 folds due to the loss of outer membrane and extra-nuclear components [89]. Further, since sperm head and tail both carry small RNAs, differences in the sperm lysis methods could differ in their ability to break sperm heads and tails, further introducing variations. The use of modern methods in comparison to the old techniques of library preparation can ensure capturing of more tRFs; for example, with the use of additional T4 polynucleotide kinase [90, 91]. Even a simple factor, such as the method of sperm preparation, may affect the observed sncRNAs. Due to the well-defined nomenclature in the case of miRNAs and the sufficient literature available, we could figure out consistent and variable miRNAs; however, similar analyses for other sncRNAs await further advancements in their nomenclature and specific identification. We also observed sample to sample and population-wise differences among all sncRNAs. Both technical and environmental factors might contribute to these variations. We tried to reduce the contribution of technical variations by re-analyzing the raw data for all included studies in a uniform way; however, other technical variations introduced by steps involved in data generation could still contribute to the population-wise differences reported in this comparison.

Conclusions and future perspective

The primary focus of this study was to work out the sperm sncRNAome composition by identifying similarities across individuals and populations. We conclude that in Indian samples, rsRNAs (13.71–78.76%), YsRNAs (0.64-76.53%) and tsRNAs (5.63–35.16%) constitute the major fraction, and miRNAs, piRNAs, mt-tsRNAs and other sncRNAs constitute the minor fraction of the sperm sncRNAome. Similarly, in three other populations, rsRNAs (11–80%) and tsRNAs (10–60%) constitute the major fraction and YsRNAs (0.62–4.28%), miRNAs (0.41–7.37%), piRNAs (1.37–4.36%), mt-tsRNAs

(0.14–4.33%), and other sncRNAs constitute the minor fraction of the sperm sncRNAome. Up to 50% of tsRNAs across all populations are contributed by tRNA-Gly-GCC, but the second prominent contributor differs across populations. mt-tsRNAs show a significantly different pattern than nuclear tsRNAs. Though the majority of mt-tsRNAs are also derived from one mt-tRNA, the dominant contributor differs across populations. Further, the top contributing tRNAs to the nuclear and mitochondrial tsRNAs are also different. The majority of rsRNAs are derived from 28S rRNA across all populations. Ys4RNAs constitute the top fraction in the YsRNA pool. In addition to individual to individual variations within a population, sperm sncRNA biotypes also show highly significant population-wise variations. Among sperm miRNAs, less than 100 were consistent in each population and only 17 were consistent across populations, demonstrating a very high level of inconsistency. The most interesting finding is with regard to sperm piRNAs, which consist entirely (in Indian samples) or chiefly (in other populations) of the chromosome 15 cluster derived piRNAs. We believe that the core (consistent) sncRNAs may be critical to spermatogenesis or sperm fertility, and the peripheral (highly variable) sncRNAs may be responsive to environmental cues, which would differ individual to individual even within the same population. Animal studies can be conducted in a more controlled environment, but the human sperm sncRNAome is influenced by a number of environmental factors, rendering its interpretation far more complicated. Significant individual to individual or population-wise variations in the human sperm sncRNAome profile heightens the possibility of misinterpretation of natural variations as biologically meaningful. We believe that the data presented in this study will serve to guide the analysis on human sperm sncRNA data in the future.

Sperm RNAs, particularly, sncRNAs, once thought to be a useless cargo, are in vogue today. From functions in spermatogenesis and embryonic development to acting as the carriers of transgenerational information, these tiny molecules cater big functions, much of which still remains to be comprehended. Defining the composition of the sncRNAome for normozoospermic fertile human sperm would serve six major purposes. First, this will help us define the core sncRNAome composition for normal human sperm, which can identify fertility biomarkers for assisted reproduction. Second, sncRNAs consistent across populations can guide investigations into their roles in spermatogenesis (past activity) and early embryonic development (future activity). Interestingly, all three major sncRNAs; rsRNAs [32], tsRNAs [92, 93], and miRNAs [10] have been shown to affect embryo quality. Third, sncRNAs linked to poor diet, poor lifestyle, obesity, high BMI, and other lifestyle disorders can be used as sperm quality biomarkers for the evaluation of donor sperm in assisted reproduction. Fourth, the extent of natural variations in the sncRNA levels and composition can be used to better interpret the intergenerational or transgenerational inheritance of certain acquired human lifestyle disorders. Fifth, sncRNAome in-depth studies will identify the sncRNAs with potential applications in RNA therapeutics for male infertility treatment in the future. Sixth and the most important, the present effort of defining the human sperm sncRNAome would help in delineating biologically meaningful variations from sample to sample or technical variations, particularly while comparing across studies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12958-025-01358-3.

Supplementary Material 1: Supplementary data 1. Excel sheet with sample-wise data corresponding to each class of sncRNAs.

Supplementary Material 2: Supplementary Fig. 1. Methodology followed by the studies included in this analysis.

Supplementary Material 3: Supplementary Fig. 2. Principal component analysis showing population-wise clustering of samples.

Supplementary Material 4: Supplementary Fig. 3. Sperm RNA purity check by PCR using cell-specific markers.

Supplementary Material 5: Supplementary Fig. 4. tsRNA fragmentation pattern across all tRNAs and the four populations as generated by tRAX pipeline. Purple colour (transcript specific) represents reads that are specific to a transcript, blue colour (isodecoder specific) represents reads that map uniquely to the transcripts with the corresponding anticodon, green colour (isotype specific) represents the reads that map only to transcripts with the corresponding tRNA isotype, and red colour (not amino specific) represent the reads that map more than one isotype.

Supplementary Material 6

Acknowledgements

The authors are thankful to the Council of Scientific and Industrial Research for financial support (BSC0101 and MLP2026). Poonam Mehta would like to thank the University Grants Commission (460/CSIR-UGC NET DEC.2017) for graduate fellowship. This manuscript bears CDRI communication number 10931.

Author contributions

PM, RS conceived the study, PM executed all experiments and generated data. RS secured funding and supervised the study. PM downloaded the data for other populations and performed full analysis. PM and RS participated in writing and reviewing the article. PM and RS edited the final version of the article.

Funding

The study was funded by the Council of Scientific and Industrial Research (CSIR), Govt. of India under project codes BSC0101 and MLP2026.

Data availability

The data underlying this article are available in NCBI repository and can be accessed with BioProject no. PRJNA1141922.

Declarations

Ethics approval

This study was approved by the Institutional Human Ethics Committee of the Central Drug Research Institute, Lucknow (CDRI/IEC/2015/A1). The study was

conducted in accordance with the guidelines of the declaration of Helsinki. Written informed consent was obtained from all participants.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]CSIR-Central Drug Research Institute, Lucknow, India
[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

Received: 7 October 2024 / Accepted: 7 February 2025
Published online: 06 March 2025

References

1. Abraham K, Bhargava P. Nucleic acid metabolism of mammalian spermatozoa. Biochem J. 1963;86:298.
2. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The microRNA signature of mouse spermatozoa is substantially modified during epididymal maturation. Biol Reprod. 2015;93:91–1.
3. Conine CC, Sun F, Song L, Rivera-Pérez JA, Rando OJ. Small RNAs gained during epididymal transit of sperm are essential for embryonic development in mice. Dev Cell. 2018;46:470–80.
4. Sharma U, Sun F, Conine CC, Reichholf B, Kukreja S, Herzog VA, et al. Small RNAs are trafficked from the epididymis to developing mammalian sperm. Dev Cell. 2018;46:481–94.
5. Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. Delivering spermatozoan RNA to the oocyte. Nature. 2004;429:154–154.
6. Sendler E, Johnson GD, Mao S, Goodrich RJ, Diamond MP, Hauser R, et al. Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. 2013;41:4104–17.
7. Liu W-M, Pang RT, Chiu PC, Wong BP, Lao K, Lee K-F et al. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. Proceedings of the National Academy of Sciences. 2012;109:490–4.
8. Abu-Halima M, Hammadeh M, Backes C, Fischer U, Leidinger P, Lubbad AM, et al. Panel of five microRNAs as potential biomarkers for the diagnosis and assessment of male infertility. Fertil Steril. 2014;102:989–97.
9. Corral-Vazquez C, Salas-Huetos A, Blanco J, Vidal F, Sarrate Z, Anton E. Sperm microRNA pairs: new perspectives in the search for male fertility biomarkers. Fertil Steril. 2019;112:831–41.
10. Shi S, Shi Q, Sun Y. The effect of sperm miR-34c on human embryonic development kinetics and clinical outcomes. Life Sci. 2020;256:117895.
11. Abu-Halima M, Khaizaran ZA, Ayesh BM, Fischer U, Khaizaran SA, Al-Battah F, et al. MicroRNAs in combined spent culture media and sperm are associated with embryo quality and pregnancy outcome. Fertil Steril. 2020;113:970–80.
12. Joshi M, Andrabi SW, Yadav RK, Sankhwar SN, Gupta G, Rajender S. Qualitative and quantitative assessment of sperm miRNAs identifies hsa-miR-9-3p, hsa-miR-30b-5p and hsa-mir-122-5p as potential biomarkers of male infertility and sperm quality. Reproductive Biology Endocrinol. 2022;20:122.
13. Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science. 2005;308:1466–9.
14. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature. 2006;441:469–74.
15. Gapp K, Jawaid A, Sarkies P, Bohacek J, Pelczar P, Prados J, et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci. 2014;17:667–9.
16. Rodgers AB, Morgan CP, Leu NA, Bale TL. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. Proceedings of the National Academy of Sciences. 2015;112:13699–704.
17. Gapp K, van Steenwyk G, Germain P-L, Matsushima W, Rudolph KL, Manuella F, et al. Alterations in sperm long RNA contribute to the epigenetic inheritance of the effects of postnatal trauma. Mol Psychiatry. 2020;25:2162–74.
18. Cunningham AM, Walker DM, Ramakrishnan A, Doyle MA, Bagot RC, Cates HM, et al. Sperm transcriptional state associated with paternal transmission of stress phenotypes. J Neurosci. 2021;41:6202–16.
19. Watkins AJ, Dias I, Tsuro H, Allen D, Emes RD, Moreton J, et al. Paternal diet programs offspring health through sperm-and seminal plasma-specific pathways in mice. Proc Natl Acad Sci. 2018;115:10064–9.
20. Wang B, Xia L, Zhu D, Zeng H, Wei B, Lu L, et al. Paternal high-fat diet altered sperm 5'tsRNA-Gly-GCC is associated with enhanced gluconeogenesis in the offspring. Front Mol Biosci. 2022;9:857875.
21. Short A, Yeshurun S, Powell R, Perreau V, Fox A, Kim JH, et al. Exercise alters mouse sperm small noncoding RNAs and induces a transgenerational modification of male offspring conditioned fear and anxiety. Translational Psychiatry. 2017;7:e1114–1114.
22. Ingerslev LR, Donkin I, Fabre O, Versteyhe S, Mechta M, Pattamaprapanont P, et al. Endurance training remodels sperm-borne small RNA expression and methylation at neurological gene hotspots. Clin Epigenetics. 2018;10:1–11.
23. Tomar A, Gomez-Velazquez M, Gerlini R, Comas-Armangué G, Makharadze L, Kolbe T et al. Epigenetic inheritance of diet-induced and sperm-borne mitochondrial RNAs. Nature. 2024; 630: 720-727.
24. Nätt D, Kugelberg U, Casas E, Nedstrand E, Zalavary S, Henriksson P, et al. Human sperm displays rapid responses to diet. PLoS Biol. 2019;17:e3000559.
25. Zhang Y, Chen Q. Human sperm RNA code senses dietary sugar. Nat Reviews Endocrinol. 2020;16:200–1.
26. Chen Q, Yan W, Duan E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. Nat Rev Genet. 2016;17:733–43.
27. Ramesh R, Skog S, Örkenby L, Kugelberg U, Nätt D, Öst A. Dietary sugar shifts mitochondrial metabolism and small RNA biogenesis in sperm. Antioxid Redox Signal. 2023;38:1167–83.
28. Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, et al. A survey of small RNAs in human sperm. Hum Reprod. 2011;26:3401–12.
29. Mehta P, Singh R. Small RNAs: an ideal choice as sperm quality biomarkers. Front reprod health. 2024;6:1329760.
30. Morgan CP, Shetty AC, Chan JC, Berger DS, Ament SA, Epperson CN, et al. Repeated sampling facilitates within-and between-subject modeling of the human sperm transcriptome to identify dynamic and stress-responsive sncRNAs. Sci Rep. 2020;10:17498.
31. Pantano L, Jodar M, Bak M, Ballescà JL, Tommerup N, Oliva R, et al. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. RNA. 2015;21:1085–95.
32. Hua M, Liu W, Chen Y, Zhang F, Xu B, Liu S, et al. Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. Cell Discovery. 2019;5:20.
33. Shi J, Ko E-A, Sanders KM, Chen Q, Zhou T. SPORTS1. 0: a tool for annotating and profiling non-coding RNAs optimized for rRNA-and tRNA-derived small RNAs. Genomics Proteom Bioinf. 2018;16:144–51.
34. Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. Nat Methods. 2015;12:879–84.
35. Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, et al. Efficient and quantitative high-throughput tRNA sequencing. Nat Methods. 2015;12:835–7.
36. Team RC. RA language and environment for statistical computing. Computing: R Foundation for Statistical; 2020.
37. Wickham H. Data Analysis. In: Wickham H, editor. ggplot2: Elegant Graphics for Data Analysis [Internet]. Cham: Springer International Publishing; 2016. pp. 189–201. Available from: https://doi.org/10.1007/978-3-319-24277-4_9
38. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: a grammar of data manipulation. R package version 1.1. 2. Computer software. 2023.
39. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9:e1003118.
40. Morgan M, Pages H, Obenchain V, Hayden N, Rsamtools. Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R Package Version. 2016;1:677–89.
41. Team T. BSgenome. Hsapiens. UCSC. hg38: full genomic sequences for Homo sapiens (UCSC genome hg38). 2023.
42. Abu-Halima M, Hammadeh M, Schmitt J, Leidinger P, Keller A, Meese E, et al. Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. Fertil Steril. 2013;99:1249–e125516.
43. 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.
44. Salas-Huetos A, Blanco J, Vidal F, Godo A, Grossmann M, Pons MC, et al. Spermatozoa from patients with seminal alterations exhibit a differential micro-ribonucleic acid profile. Fertil Steril. 2015;104:591–601.

45. Salas-Huetos A, Blanco J, Vidal F, Grossmann M, Pons MC, Garrido N, et al. Spermatozoa from normozoospermic fertile and infertile individuals convey a distinct Mi RNA cargo. Andrology. 2016;4:1028–36.
46. Tian H, Lv M, Li Z, Peng D, Tan Y, Wang H, et al. Semen-specific miRNAs: suitable for the distinction of infertile semen in the body fluid identification? Forensic Sci International: Genet. 2018;33:161–7.
47. Heidary Z, Zaki-Dizaji M, Saliminejad K, Khorram Khorshid HR. MicroRNA profiling in spermatozoa of men with unexplained asthenozoospermia. Andrologia. 2019;51:e13284.
48. Mokánszki A, Molnár Z, Varga Tóthné E, Bodnár B, Jakab A, Bálint BL, et al. Altered microRNAs expression levels of sperm and seminal plasma in patients with infertile ejaculates compared with normozoospermic males. Hum Fertility. 2020;23:246–55.
49. Rosenkranz D, Zischler H. proTRAC-a software for probabilistic piRNA cluster detection, visualization and analysis. BMC Bioinformatics. 2012;13:1–10.
50. Shigematsu M, Kawamura T, Kirino Y. Generation of 2′, 3′-cyclic phosphate-containing RNAs as a hidden layer of the transcriptome. Front Genet. 2018;9:562.
51. Zhang Y, Ren L, Sun X, Zhang Z, Liu J, Xin Y, et al. Angiogenin mediates paternal inflammation-induced metabolic disorders in offspring through sperm tsRNAs. Nat Commun. 2021;12:6673.
52. Chen Q, Zhou T. Emerging functional principles of tRNA-derived small RNAs and other regulatory small RNAs. J Biol Chem. 2023;299:105225.
53. Zhang H-T, Zhang Z, Hong K, Tang W-H, Liu D-F, Mao J-M, et al. Altered microRNA profiles of testicular biopsies from patients with nonobstructive azoospermia. Asian J Androl. 2020;22:100–5.
54. Tomic M, Bolha L, Pizem J, Ban-Frangez H, Vrtacnik-Bokal E, Stimpfel M. Association between sperm morphology and altered sperm microRNA expression. Biology. 2022;11:1671.
55. Abhari A, Zarghami N, Farzadi L, Nouri M, Shahnazi V. Altered of microRNA expression level in oligospermic patients. Iran J Reprod Med. 2014;12:681–6.
56. Liang G, Wang Q, Zhang G, Li Z, Wang Q. Differentially expressed miRNAs and potential therapeutic targets for asthenospermia. Andrologia [Internet]. 2022 [cited 2024 Dec 26];54. Available from: https://onlinelibrary.wiley.com/doi/https://doi.org/10.1111/and.14265
57. Lian C, Sun B, Niu S, Yang R, Liu B, Lu C, et al. A comparative profile of the microRNA transcriptome in immature and mature porcine testes using Solexa deep sequencing. FEBS J. 2012;279:964–75.
58. Bo W, RAN M, Rong C, PENG F, Hui L, Hu G, et al. miR-10b promotes porcine immature sertoli cell proliferation by targeting the DAZAP1 gene. J Integr Agric. 2019;18:1924–35.
59. Bouhallier F, Allioli N, Lavial F, Chalmel F, Perrard M-H, Durand P, et al. Role of miR-34c microRNA in the late steps of spermatogenesis. RNA. 2010;16:720–31.
60. Gupta A, Vats A, Ghosal A, Mandal K, Sarkar R, Bhattacharya I, et al. Follicle-stimulating hormone-mediated decline in miR-92a-3p expression in pubertal mice sertoli cells is crucial for germ cell differentiation and fertility. Cell Mol Life Sci. 2022;79:136.
61. Luo M, Hao L, Hu F, Dong Y, Gou L, Zhang W, et al. MicroRNA profiles and potential regulatory pattern during the early stage of spermatogenesis in mice. Sci China Life Sci. 2015;58:442–50.
62. Khanehzad M, Nourashrafeddin SM, Abolhassani F, Kazemzadeh S, Madadi S, Shiri E, et al. MicroRNA-30a-5p promotes differentiation in neonatal mouse spermatogonial stem cells (SSCs). Reproductive Biology Endocrinol. 2021;19:85.
63. Niu Z, Goodyear SM, Rao S, Wu X, Tobias JW, Avarbock MR et al. MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. Proceedings of the National Academy of Sciences. 2011;108:12740–5.
64. Xie Y, Xu Z, Wu C, Zhou C, Zhang X, Gu T, et al. Extracellular vesicle-encapsulated mir-21-5p in seminal plasma prevents sperm capacitation via Vinculin inhibition. Theriogenology. 2022;193:103–13.
65. Ma J, Fan Y, Zhang J, Feng S, Hu Z, Qiu W, et al. Testosterone-dependent miR-26a-5p and let-7 g-5p act as signaling mediators to regulate sperm apoptosis via targeting PTEN and PMAIP1. Int J Mol Sci. 2018;19:1233.
66. Janus P, Toma-Jonik A, Vydra N, Mrowiec K, Korfanty J, Chadalski M, et al. Pro-death signaling of cytoprotective heat shock factor 1: upregulation of NOXA leading to apoptosis in heat-sensitive cells. Cell Death Differ. 2020;27:2280–92.
67. Chen X, Che D, Zhang P, Li X, Yuan Q, Liu T, et al. Profiling of miRNAs in porcine germ cells during spermatogenesis. Reproduction. 2017;154:789–98.
68. Xu H, Wang X, Wang Z, Li J, Xu Z, Miao M, et al. MicroRNA expression profile analysis in sperm reveals hsa-mir-191 as an auspicious omen of in vitro fertilization. BMC Genomics. 2020;21:1–9.
69. Hutcheon K, McLaughlin EA, Stanger SJ, Bernstein IR, Dun MD, Eamens AL, et al. Analysis of the small non-protein-coding RNA profile of mouse spermatozoa reveals specific enrichment of piRNAs within mature spermatozoa. RNA Biol. 2017;14:1776–90.
70. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet. 2019;20:89–108.
71. Lempradl A, Kugelberg U, Iconomou M, Beddows I, Nätt D, Casas E et al. Intergenerational metabolic priming by sperm piRNAs [Internet]. 2021 [cited 2024 Jul 21]. Available from: https://doi.org/10.1101/2021.03.29.436592
72. Shaukat A-N, Kaliatsi EG, Stamatopoulou V, Stathopoulos C. Mitochondrial tRNA-derived fragments and their contribution to gene expression regulation. Front Physiol. 2021;12:729452.
73. Chen X, Sun Q, Zheng Y, Liu Z, Meng X, Zeng W, et al. Human sperm tsRNA as potential biomarker and therapy target for male fertility. Reproduction. 2021;161:111–22.
74. Donkin I, Versteyhe S, Ingerslev LR, Qian K, Mechta M, Nordkap L, et al. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. Cell Metabol. 2016;23:369–78.
75. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. Science. 2016;351:397–400.
76. Guo Y, Bai D, Liu W, Liu Y, Zhang Y, Kou X, et al. Altered sperm tsRNAs in aged male contribute to anxiety-like behavior in offspring. Aging Cell. 2021;20:e13466.
77. López-Polo V, Maus M, Zacharioudakis E, Lafarga M, Attolini CS-O, Marques FDM, et al. Release of mitochondrial dsRNA into the cytosol is a key driver of the inflammatory phenotype of senescent cells. Nat Commun. 2024;15:7378.
78. Trigg NA, Eamens AL, Nixon B. The contribution of epididymosomes to the sperm small RNA profile. Reproduction. 2019;157:R209–23.
79. Wang H, Wang Z, Zhou T, Morris D, Chen S, Li M, et al. Small RNA shuffling between murine sperm and their cytoplasmic droplets during epididymal maturation. Dev Cell. 2023;58:779–90.
80. Chu C, Yu L, Wu B, Ma L, Gou L-T, He M, et al. A sequence of 28S rRNA-derived small RNAs is enriched in mature sperm and various somatic tissues and possibly associates with inflammation. J Mol Cell Biol. 2017;9:256–9.
81. Lerner MR, Boyle JA, Hardin JA, Steitz JA. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science. 1981;211:400–2.
82. Rutjes SA, Van Der Heijden A, Van Venrooij WJ, Pruijn GJM, Utz PJ. Rapid Nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. J Biol Chem. 1999;274:24799–807.
83. Krude T, Christov CP, Hyrien O, Marheineke K. Y RNA functions at the initiation step of mammalian chromosomal DNA replication. J Cell Sci. 2009;122:2836–45.
84. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DIK. Deep sequencing of serum small RNAs identifies patterns of 5′ tRNA half and YRNA fragment expression Associated with breast Cancer. Biomark Cancer. 2014;6:BIC.S20764.
85. Meiri E, Levy A, Benjamin H, Ben-David M, Cohen L, Dov A, et al. Discovery of microRNAs and other small RNAs in solid tumors. Nucleic Acids Res. 2010;38:6234–46.
86. Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, et al. Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. Nucleic Acids Res. 2014;42:7290–304.
87. Elzer D, Bremser M, Zischler H. Human sperm heads harbor modified YsRNA as transgenerationally inherited non-coding RNAs. Front Genet. 2023;14:1294389.
88. Mao S, Goodrich RJ, Hauser R, Schrader SM, Chen Z, Krawetz SA. Evaluation of the effectiveness of semen storage and sperm purification methods for spermatozoa transcript profiling. Syst Biology Reproductive Med. 2013;59:287–95.
89. Johnson GD, Mackie P, Jodar M, Moskovtsev S, Krawetz SA. Chromatin and extracellular vesicle associated sperm RNAs. Nucleic Acids Res. 2015;43:6847–59.
90. Liu S, Sharma U, Sperm RNA. Payload: implications for intergenerational epigenetic inheritance. IJMS. 2023;24:5889.
91. Perillo G, Shibata K, Wu P-H. piRNAs in sperm function and embryo viability. Reproduction. 2023;165:R91–102.
92. Grosso JB, Zoff L, Calvo KL, Maraval MB, Perez M, Carbonaro M, et al. Levels of seminal tRNA-derived fragments from normozoospermic men correlate with the success rate of ART. Mol Hum Reprod. 2021;27:gaab017.

93. Nätt D, Öst A. Male reproductive health and intergenerational metabolic responses from a small RNA perspective. J Intern Med. 2020;288:305–20.

Publisher's note