www.sciencetranslationalmedicine.org/cgi/content/full/7/295/295re6/DC1

Supplementary Materials for

Absence of sperm RNA elements correlates with idiopathic male infertility

Meritxell Jodar, Edward Sendler, Sergey I. Moskovtsev, Clifford L. Librach, Robert Goodrich, Sonja Swanson, Russ Hauser, Michael P. Diamond, Stephen A. Krawetz*

*Corresponding author. E-mail: steve@compbio.med.wayne.edu

Published 8 July 2015, *Sci. Transl. Med.* **7**, 295re6 (2015)
DOI: 10.1126/scitranslmed.aab1287

The PDF file includes:

Supplemental Materials and Methods

Assessment of semen parameters

Semen parameters were evaluated in all sperm samples, and the time of this evaluation was designated as the start of the first 90 day cycle. Semen parameters that were evaluated included the volume of the ejaculate, sperm concentration, sperm motility, sperm morphology, the presence of round cells [white blood cells (WBC) and immature germ cells], and sperm DNA damage. Computer-assisted semen analysis (CASA) was performed to assess sperm kinetics using an IVOS analyzer (Hamilton Thorne Bioscience). Sperm morphology and the presence of round cells were assessed from smears (two slides for each sample) after fixation and staining as recommended by WHO in 1999 (32). Semen samples were excluded from the study if the WBC count was ≥ two million leukocytes per ml (33), higher than the acceptable level described by WHO in 2010 (34). Morphology was assessed by strict Kruger's criteria (35). The teratozoospermia index (TZI) was calculated and defined as the number of abnormalities present per abnormal spermatozoon. Four different sperm abnormalities were assessed: the presence of a cytoplasmic droplet, or head, neck/midpiece, or tail abnormalities. Semen samples were also analyzed for sperm DNA damage using the Sperm Chromatin Structure Assay (SCSA) (36, 37) and expressed as the DNA fragmentation index (DFI). Bacterial cultures were established using a separate aliquot of the semen sample. Bacterial colony counts were performed on all morphotypes of possible semen tract pathogens. Pathogen positive semen samples were not included in the study.

Isolation and Sequencing of Sperm RNA

Semen samples frozen at -80°C were thawed, then immediately subjected to purification through a 50% gradient of Puresperm (NidaCon International AB) (38). RNA sequencing required a minimum of two million spermatozoa. Sperm RNA was isolated and purified using a modified RNeasy (Qiagen) protocol, and the RNA quality was assessed as previously described (39,40). Somatic cell contamination was assessed as a function of the leukocyte-specific marker Protein Tyrosine Phosphatase, Receptor

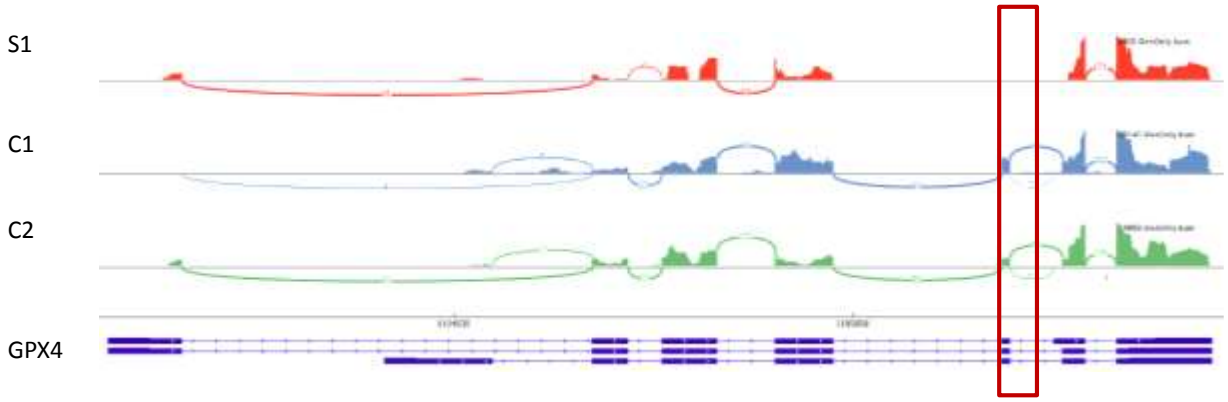
Type, C (*PTPRC*) RNA (13), epithelial cells marker Cadherin 1, Type 1, E-Cadherin (*CDH1*) RNA (41), and immature testicular germ cells marker mast/stem cell growth factor receptor Kit (*KIT*) RNA (41); these were found to be negative in all samples. A total of 2 ng of unfractionated input RNA along with the manufacturer recommended concentration of ERCC RNA spike-in control (Life Technologies) was reverse transcribed and the resulting cDNA amplified using SeqPlex RNA Amplification (Sigma-Aldrich Co.), allowing for the use of samples with a low quantity of spermatozoa (40). Amplified primer dimers were removed prior to library construction using Agencourt AMPure XP (Beckman Coulter), and libraries were constructed with 50 ng of amplified cDNA using NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). Barcoding permitted multiplexed sequencing of samples. Paired-end sequencing was performed for 50 cycles using the Illumina HiSeq-2500 sequencer.

Computational analysis

After demultiplexing (Casava 1.8.2, Illumina), sequences were aligned using Bowtie (42) (v. 2.2.3.0) and splice junction mapper Tophat (43) (v.2.0.12) to human genome build hg19, with supplementary alignment to ribosomal 18S and 28S sequence and ERCC control sequences. To standardize samples for differential comparisons, we developed criteria indicative of reduced initial quality or quantity of sperm RNA. Accordingly, two samples with ≥ 30% of the reads aligned to mtRNA, 19 samples with ≥ 30% of the reads aligned to ERCC (≥ 6x input of 5%), and three samples with ≤ 10% of the reads aligned to rRNA were excluded from analysis. From the 72 remaining samples, transcript abundance by RPKM (reads per kilobase per million) was initially obtained based on total number of reads aligning over all exons of that transcript, normalized to complete transcript length (Genomatix RNA-seq Expression, v.3.10410, Genomatix Software GmbH). Although this measure provides an indication of the importance and overall abundance of the approximately 25,000 annotated human transcripts, it fails to consider commonly found (but unannotated) alternative isoforms typically observed in sperm. Additionally, this broad method cannot independently resolve the irregularities of a single exon of an otherwise abundant transcript. This is illustrated in Supplemental Figure 1, in the case of

Glutathione Peroxidase 4 (*GPX4*), encoding an antioxidant essential for spermatogenesis, fertilization, and early embryogenesis. The absence of RNA-seq reads of the exon and the absence of the splice junction reads joining consecutive exons is highlighted with a red box, indicating its suppression in a specific sample. This level of detail is only possible at the RNA element level. To take full advantage of this degree of resolving power, reads aligning to all exons as annotated by UCSC Genes (~80,000 transcripts, comprising ~280,000 single exons and noncoding elements) were independently surveyed and tallied. Additionally, a number of previously unannotated sperm elements in intergenic and intronic regions (22) were included. The high levels and pervasive appearance of these elements in all sperm samples surveyed suggests that they are required, although their function has yet to be determined (22). A total of 848 distinct intronic and intergenic sperm elements were identified and added to the list of previously annotated sperm elements (22) to form the complete list of elements that were assessed in this study.

There are large differences in the proportion of ribosomal and mitochondrial RNAs within sperm, and considerable differences in the degree of transcript stability (22), even between normal individuals. This yields a marked differential in the absolute abundance of even retained functional RNAs. Differential ranking (Spearman rank) of RNA elements using percentile rank was shown to be a more reliable sample invariant measure of the global transcript population, as compared to absolute expression (RPKM). The latter showed substantial inter-sample variation even after several normalization approaches were used. These included evaluating the representation of ERCC transcript standards and excluding the largely fragmented ribosomal and mitochondrial RNAs.



Supplemental Figure 1: Distribution and junctions of RNA-seq reads of a selected required SRE, *GPX4*. The exon highlighted in a red box is one of the 648 required SREs found on *GPX4* gene. This element has a stable expression across control samples in Group I, such as C1 and C2. A sample from the test set (S1) reveals the absence of this exon. This difference in a single exon can only be detected at the sperm element level.

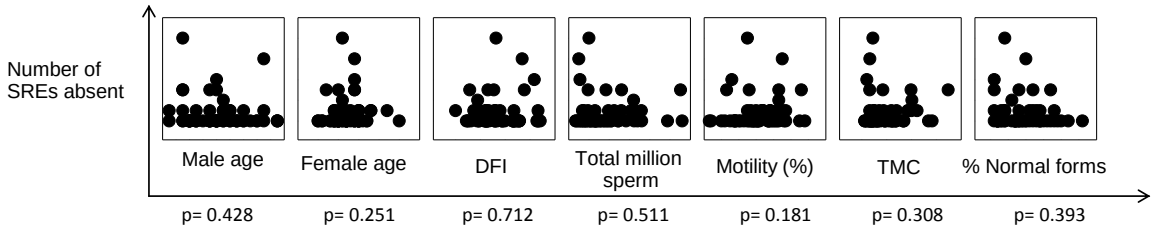


Figure S2. No correlation between the number of absent SREs and sperm parameters or partner age. Spearman correlation does not show any association between number of SREs absent and the partner's age or any semen parameters tested.