INFERTILITY

Comment on "Absence of sperm RNA elements correlates with idiopathic male infertility"

Heike Cappallo-Obermann and Andrej-Nikolai Spiess*

The spermatozoal population used for a recent RNA sequencing study may have been contaminated by somatic cell types because of an inefficient sperm purification procedure.

In a recent article (1), Jodar and co-workers describe a next-generation sequencing–based analysis of spermatozoal RNAs and their correlation to the outcome of natural conception and assisted reproduction techniques. The same group (S. A. Krawetz) has previously published several gene expression studies (2–5) on human spermatozoa that are based on RNA sequencing of purified sperm populations. Sperm purification is a mandatory procedure because ejaculates are routinely contaminated with so-called "round cells," mostly somatic cells such as leukocytes, macrophages, and epithelial cells of the genital tract but also nonsomatic and less differentiated germ cells (spermatocytes and round spermatids). In contrast to sperm concentration, morphology, and motility, no reference values for round cells exist, but the World Health Organization states a consensus value of 1×10^6/ml for peroxidase-positive leukocytes (6). To obtain an overview on the proportion of round cells in ejaculates of men enrolled in an assisted reproduction program at our clinic (Department of Andrology, University Hospital Hamburg-Eppendorf), we have compiled round cell concentrations from 100 patients seen in the year 2014 (fig. S1). More than 30% of these men exhibited round cell concentrations that were well in excess of 5×10^6/ml (red bars), which should render an efficient removal problematic.

One might argue that a low proportion of somatic cell contamination (1 to 2%) is negligible in differential gene expression studies of human spermatozoa. Unfortunately, somatic cells contain a roughly 100- to 200-fold excess of RNA compared to the 20 to 50 fg of a single spermatozoon (2, 7, 8) and therefore create substantial bias by severely shifting the proportion of sperm-specific RNAs in round cell–contaminated ejaculates. This is largely a dilution effect, which we describe in more detail in the Supplementary Materials, and which makes an efficient elimination of somatic cells imperative before RNA isolation. In this respect, there are several methods for obtaining a highly purified sperm population, such as swim-up, density gradient centrifugation, and somatic cell lysis. Although swim-up enriches for the most motile subpopulation and delivers rather low sperm numbers, the latter two have been used in gene expression studies of S. A. Krawetz and his group. Some initial microarray studies used somatic cell lysis (9), whereas recent RNA sequencing studies (2–5), including that of Jodar et al. (1), are based on the 50% density gradient centrifugation protocol described in (7).

Preliminary results in our lab indicated that a 50% density gradient centrifugation is insufficient for somatic cell elimination. When comparing the raw ejaculate (fig. S2A) with 50% density gradients (fig. S2B), only a negligible elimination of round cells takes place. In contrast, 90% density gradients sufficiently remove round cells (fig. S2, C and D), resulting in considerably smaller but purer fractions (fig. S2E).

These findings suggested that the RNA sequencing data in (1) are likely to be contaminated with transcripts originating from somatic cells. To validate this assumption, first, we downloaded the RNA sequencing data from the Gene Expression Omnibus (GEO) database (GSE65683), which are supplied as a summary file with the expression magnitude given in percentile ranks (GSE65683_ClusterData_PercRank_GEOsubmission.xlsx). The University of California, Santa Cruz (UCSC) gene annotations in this file ("Element name") were matched with an Hg19 annotation file downloaded from the Table Browser on the UCSC Web site (https://genome.ucsc.edu/cgi-bin/hgTables) using the R programming language (www.r-project.org). We then extracted all exons for a set of various somatic transcripts expressed in the genital tract as well as for germ cells. More specifically, these were PRM1/PRM2/ODF1/TNP1/LELP1/SMCP/OAZ3/PFN4/ACRV1/KIT for differentiated and undifferentiated germ cells, the epithelial marker CDH1, the prostate-specific transcripts TGM4 and MSMB, the seminal vesicle–specific transcripts SEMG1 and SEMG2, and the leukocyte markers IL8 and CD45/PTPRC. The expression of all exons for all 72 samples was displayed as a heatmap with expression values between 0 (background expression) and 1 (maximum expression), as defined by the percentile ranks (Fig. 1). We discovered that all samples were contaminated to an extremely high degree with prostate and seminal vesicle markers, and almost all exons for TGM4, MSMB, SEMG1, and SEMG2 displayed expression values in the upper 30% (>0.7) for all samples (marked at the bottom of the heatmap), nearly matching the saturated expression of the differentiated germ cell markers (marked at the top of the heatmap). Second, we ordered the samples in the heatmap by hierarchical clustering (Manhattan distance and complete linkage) based on all 278,605 features of the GSE65683 file. This approach identified a clear separation of the samples into two distinct clusters (Fig. 1). A closer inspection of the two clusters revealed that the first cluster (marked yellow) contains somatic transcripts for CDH1 and IL8, indicating a contamination with epithelial and leukocyte cell populations. Confusingly, the authors mention in their Supplementary Materials (1) that CDH1 is absent in all samples (as deduced from polymerase chain reaction results). Moreover, the samples' clustering does not tally with the treatment regime as defined in Table 1 of (1) ("Group," right side of heatmap), similar to what is mentioned in their Table 1 legend. In this respect, it would have been informative to correlate the somatic cell contamination with the per sample fertility treatment outcome. However, these data are not available, so the impact of somatic cell contamination on this parameter cannot be adequately assessed.

To validate these findings on an independent data set, we interrogated the microarray-based gene expression data (GSE26881) from another research group, consisting of 18 ejaculates that were purified by the same 50% density gradient centrifugation protocol (10). For the seven somatic marker transcripts used in Fig. 1, we found high, but variable, expression throughout the samples, with IL8, SEMG1, and TGM4

Department of Andrology, University Hospital Hamburg-Eppendorf, 20246 Hamburg, Germany.
*Corresponding author. Email: a.spiess@uke.de

Cappallo-Obermann and Spiess, *Sci. Transl. Med.* 8, 353tc1 (2016)   24 August 2016

1 of 3
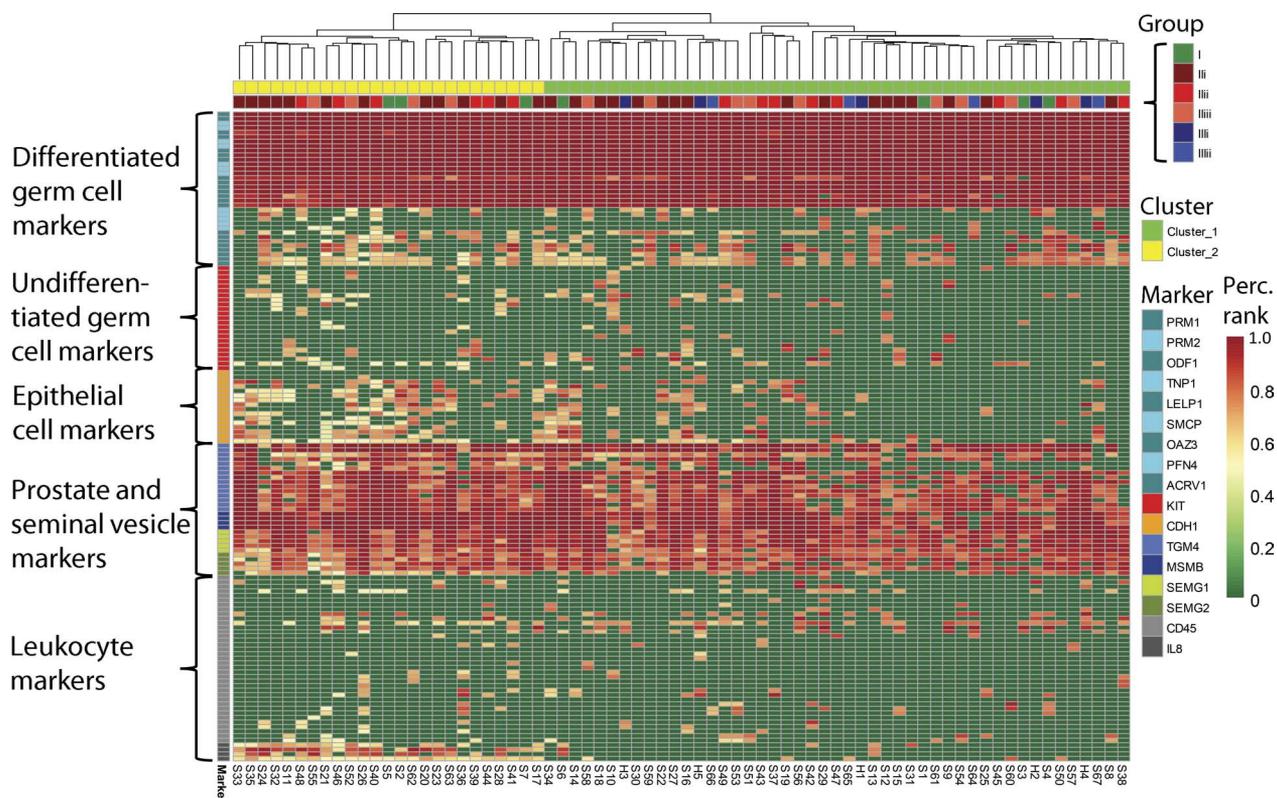
Fig. 1. Heatmap display for all exons of 17 different germ cell and somatic markers. The expression of all exons in the 72 samples from (*1*) is displayed with heatmap colors in the interval [0, 1], as given by the percentile ranks ("Perc. rank") in data set GSE65683 (GEO). A strong expression of prostate and seminal vesicle markers *TGM4/MSMB/* *SEMG1/SEMG2* is observable in all samples, and Cluster 2 (yellow) also contains the epithelial and leukocyte markers *CDH1* and *IL8*, respectively. Samples were clustered on the basis of all 278,605 features, which resulted in two clearly separable groups ("Cluster") that do not correlate with the original grouping ("Group") defined in (*1*).

exhibiting values in the upper half of the expression range (fig. S3). This corroborates an inefficient removal of somatic cells by this procedure resulting in somatic transcript contamination in gene expression studies.

Last, we interrogated whether the putative dilution effect arising from somatic transcripts described in the Supplementary Materials is manifested in the RNA sequencing data. We calculated Spearman rank correlations for 12 somatic and germ cell transcripts included in Fig. 1 and summarized them in a correlation matrix (fig. S4). The red cells in this table demonstrate strong negative correlations between differentiated germ cell markers and epithelial/leukocyte/prostate markers, most likely from the dilution of sperm-specific transcripts such as *PRM1/PRM2* in a somatic cell proportion–dependent manner. Strong positive correlations (green cells) can largely be explained by the co-expression of germ cell and somatic markers within the same cell population, such as *PRM1/PRM2* or *TGM4/MSMB*.

All these observations allowed us conclude that the problem of somatic transcript contamination in gene expression studies of human spermatozoa has not been solved adequately to date and is a topic that needs further investigation.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/8/353/353tc1/DC1
Theoretical considerations.
Fig. S1. Round cell concentrations for 100 ejaculates.
Fig. S2. Analysis of round cell elimination by density gradient centrifugation.

Fig. S3. Somatic transcript contamination in published gene expression data.
Fig. S4. Spearman rank correlation matrix for 12 different markers.
References (*11–13*)

REFERENCES AND NOTES

1. M. Jodar, E. Sendler, S. I. Moskovtsev, C. L. Librach, R. Goodrich, S. Swanson, R. Hauser, M. P. Diamond, S. A. Krawetz, Absence of sperm RNA elements correlates with idiopathic male infertility. *Sci. Transl. Med.* **7**, 295re6 (2015).
2. M. Jodar, S. Selvaraju, E. Sendler, M. P. Diamond, S. A. Krawetz; Reproductive Medicine Network, The presence, role and clinical use of spermatozoal RNAs. *Hum. Reprod. Update* **19**, 604–624 (2013).
3. S. A. Krawetz, A. Kruger, C. Lalancette, R. Tagett, E. Anton, S. Draghici, M. P. Diamond, A survey of small RNAs in human sperm. *Hum. Reprod.* **26**, 3401–3412 (2011).
4. E. Sendler, G. D. Johnson, S. Mao, R. J. Goodrich, M. P. Diamond, R. Hauser, S. A. Krawetz, Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res.* **41**, 4104–4117 (2013).
5. G. D. Johnson, E. Sendler, C. Lalancette, R. Hauser, M. P. Diamond, S. A. Krawetz, Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Mol. Hum. Reprod.* **17**, 721–726 (2011).
6. WHO, *Laboratory Manual For The Examination And Processing Of Human Semen* (ed. 3, 2010), p. 107.
7. R. J. Goodrich, E. Anton, S. A. Krawetz, Isolating mRNA and small noncoding RNAs from human sperm. *Methods Mol. Biol.* **927**, 385–396 (2013).
8. H. Cappallo-Obermann, W. Schulze, H. Jastrow, V. Baukloh, A.-N. Spiess, Highly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly. *Mol. Hum. Reprod.* **17**, 669–678 (2011).
9. C. Lalancette, A. E. Platts, G. D. Johnson, B. R. Emery, D. T. Carrell, S. A. Krawetz, Identification of human sperm transcripts as candidate markers of male fertility, *J. Mol. Med.* **87**, 735–748 (2011).

Cappallo-Obermann and Spiess, *Sci. Transl. Med.* **8**, 353tc1 (2016)   24 August 2016

2 of 3

10. S. E. Pacheco, E. A. Houseman, B. C. Christensen, C. J. Marsit, K. T. Kelsey, M. Sigman, K. Boekelheide, Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLOS One* **6**, e20280 (2011).

11. S. S. Sommer, J. E. Cohen, The size distributions of proteins, mRNA, and nuclear RNA. *J. Mol. Evol.* **15**, 37–57 (1980).

12. E. Shapiro, T. Biezuner, S. Linnarsson, Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).

13. H. Cappallo-Obermann, C. Feig, W. Schulze, A.-N. Spiess, Fold-change correction values for testicular somatic transcripts in gene expression studies of human spermatogenesis. *Hum. Reprod.* **28**, 590–598 (2013).

Citation: H. Cappallo-Obermann, A.-N. Spiess, Comment on "Absence of sperm RNA elements correlates with idiopathic male infertility." *Sci. Transl. Med.* **8**, 353tc1 (2016).

Science Translational Medicine

Comment on "Absence of sperm RNA elements correlates with idiopathic male infertility"

Heike Cappallo-Obermann and Andrej-Nikolai Spiess

ARTICLE TOOLS	http://stm.sciencemag.org/content/8/353/353tc1
SUPPLEMENTARY MATERIALS	http://stm.sciencemag.org/content/suppl/2016/08/22/8.353.353tc1.DC1
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/7/295/295re6.full http://stm.sciencemag.org/content/scitransmed/6/249/249ra108.full http://stm.sciencemag.org/content/scitransmed/5/212/212ra160.full http://stm.sciencemag.org/content/scitransmed/3/92/92ps31.full http://stm.sciencemag.org/content/scitransmed/3/92/92ra65.full http://stm.sciencemag.org/content/scitransmed/8/353/353tr1.full
REFERENCES	This article cites 12 articles, 1 of which you can access for free http://stm.sciencemag.org/content/8/353/353tc1#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service