A platform for evaluating sperm RNA biomarkers: dysplasia of the fibrous sheath—testing the concept

Aletheia Lima-Souza, D.V.M.,[a,b] Ester Anton, Ph.D.,[a,c] Shihong Mao, Ph.D.,[a] Won Jin Ho, M.S.,[a] and Stephen A. Krawetz, Ph.D.[a]

[a] Department of Obstetrics and Gynecology, Center for Molecular Medicine and Genetics, C. S. Mott Center, Wayne State University School of Medicine, Detroit, Michigan; [b] Department of Animal Science, Federal University of Ceara, Campus do Pici, Fortaleza, Ceara, Brazil; and [c] Unitat de Biologia Cel·lular Facultat de Biociències, Universitat Autònoma de Barcelona, Bellaterra, Spain

Objective: To investigate the potential use of correlative microarray-based transcript pairs as candidate markers for male fertility using dysplasia of the fibrous sheath (DFS) as an affected model. It is widely recognized that microarray technology may be limited by cost and that the quality of the transcript remains relatively unknown. To address these issues, we analyzed the stable transcript pairs by qPCR with a systematic primer design process.
Design: Experimental study.
Setting: University.
Patient(s): Men with proven fertility and men with a diagnosis of DFS.
Intervention(s): None.
Main Outcome Measure(s): Primer sequences for six genes of interest were designed using Oligo7 and Primer3Plus. Primer specificity was initially assessed in silico by searching the ENSEMBL, University of California Santa Cruz, and National Center for Biotechnology Information databases for non-target complementary sequences throughout the genome. The ability of transcript pairs to classify samples from males of proven fertility away from DFS was assessed.
Result(s): In conjunction with identifying four new stable transcript pairs, comparison of the DFS qPCR C(t) correlation coefficients revealed the disruption of four stable fertile sample transcript pairs. This suite of transcript pairs resolves DFS.
Conclusion(s): The results show that with effectively designed primers, qPCR may provide an affordable molecular assay to assess male fertility status. (Fertil Steril® 2012;97:1061–6. ©2012 by American Society for Reproductive Medicine.)
Key Words: Biomarker, male infertility, primer design, real-time PCR

Specific biomarkers of clinical significance can be identified based on gene expression analysis (1, 2). Several spermatozoal RNAs as candidate markers of male fertility status have been identified and tested using microarray-based and other genomic techniques. Previously published microarray data have shown that some pairs of transcripts are expressed in constant proportions in all sperm samples from males with proven fertility (3–5). However, such techniques have yet to be adopted clinically. One perceived barrier is the cost of analysis employing microarray technology because typically only a few probes are informative for a heterogeneous sample like spermatozoal RNAs. Perhaps a selection of paired transcripts could be of diagnostic value even when a single member of the pair is pathologically disrupted.

To date, only a handful of stable transcript pairs have been identified. Accordingly, qPCR may provide a reliable cost-effective solution to detect consistencies among fertile males. However, primer design to ensure specificity can be problematic, especially when considering the many testis-specific isoforms and the varied stability of spermatozoal transcripts (6).

Designing an effective primer set is a critical element to the success of any PCR project (7, 8). Important parameters to consider during primer design include the length of the primer sequence, melting temperatures, GC content, and specificity, as reflected in the quality of the sequence alignment of the primers to the template strand (9). A variety of bioinformatic tools have been developed to assist in the selection primers with these desired characteristics (10, 11). However, an effective means to initially assess specificity is required to ensure efficiency when one is developing a series of biomarkers. This could include a simple search of the public databases like ENSEMBL, University of

Received November 19, 2011; revised and accepted February 8, 2012; published online March 2, 2012.
A.L.-S. has nothing to disclose. E.A. has nothing to disclose. S.M. has nothing to disclose. W.J.H. has nothing to disclose. S.A.K. has nothing to disclose.
Supported in part by a visiting scholar fellowship from Brazilian Research Council (CAPES) (A.L.-S.), a mobility grant (JC 2010-0247) from the Ministerio de Educación (Gobierno de España) (E.A.), and by the Charlotte B. Failing Professorship (S.A.K.).
Reprint requests: Stephen A. Krawetz, Ph.D., 271 C.S. Mott Center, 275 E. Hancock Ave., Detroit, MI 48201 (E-mail: steve@compbio.med.wayne.edu).

Fertility and Sterility® Vol. 97, No. 5, May 2012 0015-0282/$36.00

doi:10.1016/j.fertnstert.2012.02.013

California Santa Cruz (UCSC) Genome Browser, and National Center for Biotechnology Information (NCBI) for alternative sites of priming within the now sequenced human genome.

Dysplasia of the fibrous sheath (DFS) is an anomaly found in spermatozoa of severe asthenozoospermic individuals. These severe structural sperm tail abnormalities are caused by a hyperplastic and disorganized fibrous sheath and the alteration of other cytoskeletal components of the flagellar axonemes (12). As a consequence, the gametes of these individuals present very short and thick tails (teratozoospermia) and have a severe reduction of their motility or are totally immotile (asthenozoospermia). According to the familial incidence of DFS, it has been suggested that these abnormal sperm features can have a genetic origin (12–15). The characteristic disorganization of the axonemal components present in these individuals can be traced to the protein level. Recent studies have used a candidate gene approach but this has proven to be a challenge. An infertile DSF patient presented with partial deletions in the Akap3 and Akap4 genes (16) but has not been consistently observed in others, highlighting our lack of understanding.

Microarray signal intensity analysis and preliminary RNA-Seq results suggested that at least the following six genes (ACSBG2, GPR137, RNF7, SPR54, TTC7A and UBAC1) were consistently present in mature spermatozoa. Some were present as full-length transcripts (unpublished observations). The efficacy of this combined in silico pretesting strategy for the clinical development of a qPCR-based male factor assessment assay for each combination of these six genes was tested. The utility of using a qPCR-based diagnostic of stable transcript pairs compared with a whole genome-based microarray assay (4) as biomarkers for both normal and DFS spermatozoa to assess male factor status is presented.

MATERIALS AND METHODS

Sample Selection and qPCR

Four sperm samples from fertile individuals and four from individuals affected with DFS were collected as previously described (3, 4) with Wayne State University institutional review board approval. RNAs were isolated as described (17) and subjected to cDNA synthesis for direct analysis. In brief, 10–500 ng of total RNA was incubated for 5 minutes at 65°C with oligo(dT) and 10 mM dNTP mix (Invitrogen) then placed in ice for at least 1 minute. The first strand synthesis reaction was performed by incubating these samples at 50°C during 60 minutes with a mixture containing 0.1 M dithiothreitol (DTT), RNase Block (Stratagene), 5X First-Strand Buffer, and SuperScript III RT (Invitrogen). This provided the unamplified transcript pair template. In addition, two rounds of mRNA amplification were carried out using MessageAmp II RNA Amplification Kit (Ambion, Austin, TX) according to the manufacturer's protocol. As previously described for microarray analysis, a biotinylated nucleotide analog was incorporated (3, 4). This amplified product also provided a source of transcript pair assay templates. The resultant products were quantified using PicoGreen (Invitrogen) (12), and then 10–100 ng of each product was evaluated for the presence of six transcripts (ACSBG2, GPR137, RNF7, SPR54, TTC7A, and UBAC1) in duplicate real-time polymerase chain reactions (Quantitec SYBR Green PCR, Qiagen Inc., Valencia, CA). Water, testis cDNA (2 ng/reaction; Ambion), and genomic DNA from HeLa cells (2 ng/reaction) served as qPCR controls. Genomic contamination was assessed using the DNAH12 intron spanning primers. The cycling parameters were 15 minutes at 95°C, and then 40 cycles of 30 seconds at 94°C denaturation, 60 seconds annealing at the primer-specific temperature, and 30 seconds at 72°C extension. The reaction was then terminated following a product completion cycle of 7 minutes at 72°C. Melting curve analysis was initiated at 60°C, with 1°C increments for 10 seconds to a final temperature of 95°C. Amplification products from transcripts RNF7 and UBAC1 were confirmed by agarose gel electrophoresis. The $C(t)$ value, the cycle number at which the fluorescence exceeded a threshold, were derived from the MJ Opticon Research Monitor output calculated using KLab PCR (13).

Paired Analysis

KLab PCR software (13) generated three sets of $C(t)$ values using different strategies: Template, Loglin, and $C(t)$ simple. Template is based on implementing two algorithms to quantify the starting template in PCR reaction. This approach can provide an accurate estimate of initial template concentration from both high- and low-quality data. The relative amount of initial template is automatically calculated as a function of amplification efficiency, independent of an arbitrarily selected fluorescent threshold. Loglin is based on a log-linear regression analysis of real-time PCR data (18). Both Loglin and $C(t)$ simple require low background noise to obtain accurate $C(t)$ values. In this analysis, $C(t)$ values from Template identify stable transcript pairs. A Pearson correlation coefficient (r) was calculated for every pairwise transcript combination based on the reported $C(t)$ values at each sample (19). The applied threshold for identifying a stable transcript pair was set at $r \geq |0.75|$.

RESULTS

Primer Design

The primer design strategy is outlined in Figure 1 along with the primer sequences, product length, and annealing temperatures, which are summarized in Supplemental Table 1. The sequence of the target gene of interest was retrieved using the ENSEMBL, NCBI, and the UCSC Genome Browsers. Upon acquiring the target gene structure, primer sequences were designed using Oligo7, purchased from Molecular Biology Insights, or Primer3Plus, which is freely available at http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi. The specificity of the primer was evaluated by whole genome alignment using the BLAST and BLAT similarity searches. ENSEMBL (http://useast.ensembl.org/index.html) provides both searches with a BLAST search specific for short oligonucleotides, whereas UCSC Genome Bioinformatics (http://genome.ucsc.edu/) provides access to BLAT searches only.