1 **SPERMATOZOAL TRANSCRIPT PROFILES AND FERTILITY**

2
3 **Oligo-dT selected spermatozoal transcript profiles differ among higher and lower**
4 **fertility dairy sires**

5
6 **C. J. Card,* K. E. Krieger,** M. Kaproth,** and B. L. Sartini*¹**

7 *Department of Fisheries, Animal and Veterinary Science, University of Rhode Island,
8 Kingston, 02881

9 **Genex Cooperative Inc., Shawano, WI 54166.

10
11 ¹Corresponding author: BL Sartini, 179 Center for Biotechnology and Life Sciences, 120
12 Flagg Road, University of Rhode Island, Kingston 02881, Phone: 401-874-2667, FAX:
13 401-874-7575, Email: blsartini@uri.edu

ABSTRACT

Spermatozoal messenger RNA (mRNA) has the potential as a molecular marker for sire fertility because this population can reflect gene expression that occurred during spermatogenesis and may have a functional role in early embryonic development. The goal of this study was to compare the oligo-dT selected spermatozoal transcript profiles of higher fertility (Conception Rate (CR) 1.8 to 3.5) and lower fertility (CR -2.9 to -0.4) sires using Ribonucleic Acid Sequencing (RNA-Seq). A total of 3,227 transcripts and 5,366 transcripts were identified in the higher and lower fertility populations, respectively. While common transcripts between the two populations were identified (2,422 transcripts), several transcripts were also unique to the fertility populations including 805 transcripts that were unique to the higher fertility population and 2,944 transcripts that were unique to the lower fertility population. From gene ontological analysis, the transcripts unique to each fertility population differed in Biological Processes (BP), including enrichment of regulatory transcripts for growth and protein kinase activity in the higher fertility bulls. Biological variation in transcript presence among individual sires was also found. Of the candidate fertility spermatozoal transcripts chosen from the RNA-Seq population analysis reported here and previous publications, *COX7C* was negatively correlated with sire fertility. Using high-throughput sequencing, candidate spermatozoal transcripts were identified for further study as potential markers for sire fertility.

Key Words: RNA-Seq, spermatozoa, bovine, mRNA

INTRODUCTION

Although selection for fertility has resulted in a population of dairy sires with an average fertility greater than the average cow fertility, the routine use of sires with below average fertility is a contributing factor to reduced pregnancy rates (Flowers, 2013). While the selection of above average fertility sires could improve the overall reproductive efficiency of dairy herds, sire fertility rankings are not optimized with current semen analysis methods (Gadea et al., 2004; Kastelic, 2013). Specifically, some functional characteristics of sperm, including non-compensable traits that cannot be overcome by increasing sperm dose, may not be evaluated in current assays highlighting the need to evaluate several semen parameters concurrently to more accurately assess sire fertility (Flowers, 2013; Omran et al., 2013).

Although transcription ceases during the spermatid stages of spermatogenesis, terminally-differentiated spermatozoa contain a subset of RNAs, including messenger RNAs (mRNA), non-coding RNAs (ncRNA) including microRNAs (miRNA), mitochondrial (mtRNA) and ribosomal RNA (rRNAs) that can be routinely isolated from the sperm of several species including bull, human, horse and pig (Kempisty et al., 2008; Das et al., 2010; Card et al., 2013; Sandler et al., 2013). While the function of specific spermatozoal RNAs is currently under investigation, the presence and/or abundance of some spermatozoal mRNAs have been positively correlated with sire fertility including *ADIPO*, *ADIOPR1*, *ADIPOR2*, *AK1*, *ITGB5*, *DOPPEL*, *NGF*, *TIMP2*, *PLCZ1*, *CRISP2*, and *PEBP1* while transcripts *CCT8* and *PRM2* are negatively correlated with fertility (Lalancette et al., 2008; Feugang et al., 2009; Arangasamy et al., 2011; Kasimanickam et

74 al., 2013). Comparison of spermatozoal transcripts using microarrays also demonstrates
75 differences in a large subset of these profiles among dairy sires (Lalancette et al., 2009;
76 Gilbert et al., 2007; Feugang et al., 2010). These results from previous studies suggest
77 the potential of including spermatozoal transcripts in *in vitro* assays to improve the
78 accuracy of sire fertility rankings.

79 The objective of this study was to compare the complete oligo-dT selected
80 spermatozoal mRNA profiles of higher and lower fertility dairy sires to determine if
81 variation in individual transcript presence and/or abundance could be a potential marker
82 for sire fertility. Global comparison of the entire transcript profiles can only now be
83 accomplished with the recent reports using high-throughput sequencing of spermatozoal
84 RNAs from bull, stallion and human (Card et al., 2013; Sendler et al., 2013; Das et al.,
85 2013). The oligo-dT primed spermatozoal mRNA population was explored by RNA-Seq
86 with the specific purpose to identify potentially full-length transcripts that may have a
87 biological function, including during the early stages of embryogenesis, and also differ
88 between the two fertility populations. We have previously demonstrated that
89 amplification of sperm RNA using oligo-dT priming does reveal a heterogeneous
90 population of degraded and full-length transcripts; several full-length transcripts were
91 confirmed by 5' and 3'RACE PCR (Card et al., 2013). We hypothesized that the
92 composition of the oligo-dT selected spermatozoal transcript profiles will differ between
93 the higher and lower fertility populations and select transcripts will correlate with sire
94 fertility scores. This study will provide a greater understanding of the potential utility of
95 spermatozoal mRNA as a diagnostic tool for assessing sire fertility.

MATERIALS AND METHODS

Spermatozoa

Cryopreserved semen straws were obtained from Genex Cooperative Inc. (Shawano, WI). Semen from four sires (-2.9 to -0.4 conception rate (CR; Genex's internal fertility ranking of a group of bulls relative to one another with the average bull at a 0 CR that is calculated based on direct records retrieved from cooperating dairy herds) was used to create the lower fertility amplified cDNA pool and four separate sires (1.8 to 3.5 CR) were used for the higher fertility cDNA pool that was submitted for RNA-Seq and subsequently used in qPCR validation. This batch of sires was selected for high reliability of conception rate estimate (over 90%). Spermatozoal RNA was isolated from 9 additional sires with varying CR scores (-6.7, -5.5, -3.5, -3.4, -2.7, 0.1, 0.8, 2.2, 3.6) to determine the biological variation of transcript presence and abundance.

RNA Isolation

For all sires, two straws were thawed in a 37° water bath for one minute and subsequently washed in 4 mL PBS (2 times at 600 x g). Spermatozoa RNA isolations were performed using TRIzol (Sigma Aldrich, St. Louis, MO) as previously described (Card et al., 2013). Genomic DNA removal by DNase treatment (RNA Cleanup protocol; RNeasy; Qiagen, Valencia, CA) was confirmed by lack of genomic DNA amplification by PCR as previously demonstrated (Card et al., 2013). This method of sperm processing and RNA isolation yields spermatozoal RNA that is free of contaminating RNA from leukocytes, testicular germ cells and epithelial cells as validated by lack of PCR amplification of cell specific transcripts as previously shown

(Card et al., 2013). The NanoDrop UV/Vis Spectrometer (Thermo Scientific; Wilimington, DE) was used to determine RNA concentrations and samples were stored at -80°C.

Double-Stranded cDNA (cDNA) Synthesis and Amplification

To ensure equal representation of all sires in the spermatozoal cDNA pools, equal amounts of spermatozoal RNA from individual sires were combined to generate a higher fertility pool and a lower fertility pool that were converted to ds-cDNA (referred to as cDNA in this manuscript) and linearly amplified (SMARTer Pico PCR cDNA Synthesis Kit; Clontech, Mountain View, CA) for RNA-Seq and qPCR validation. Amplification of RNA from 9 individual sires (CR -6.7 -3.6 as described above) was also conducted. This method of amplification has oligo-dT priming to specifically enrich for mRNA (full-length and fragmented) and reduced mtRNA and rRNA in the final transcript profile (Sendler et al., 2013; Mao et al., 2014). During cDNA amplification, cycle optimization was performed to maintain amplification in the linear phase. From the higher and lower fertility pooled cDNA samples, 5 µg of each sample was submitted for RNA-Seq. Additional cDNA from each pool was used for qPCR validation of RNA-Seq described below. The cDNA from individual bulls was used for qPCR analysis of transcript abundance with sire fertility scores.

RNA-Seq and Analysis

Higher and lower fertility spermatozoalcDNA samples were sequenced on a Illumina HiSeq 2000 to generate paired-end 100 bp reads (Otogenetics; Norcross, GA). All read processing and alignment to the bovine genome (UMD 3.1/bosTau6; iGenomes)

via Tophat was performed using Galaxy (Giardine et al., 2005; Langmead et al., 2009; Blankenberg et al., 2010; Goecks et al. 2010; Trapnell et al., 2012). The sequencing adapter (AGATCGGAAGAGC) was removed from all reads which resulted in the removal of 1,594,647 reads from high fertility file 1 (HF1), 64,282 reads from high fertility file 2 (HF2), 1,533,238 reads from low fertility file 1 (LF1), and 125,279 reads from low fertility file 2 (LF2) and 61,059 reads were removed due to being adapter only reads, reads < 15 nucleotides, or reads with unknown bases (N). Over-expressed sequences due to concatemer formation during the amplification of the SMARTer II A Oligonucleotide (AAGCAGTGGTATCAACGCAGAGTAA) were removed from all file reads (5,406,296 reads from HF1, 5,041,856 reads from HF2, 4,905,729 reads from LF1, and 5,488,985 reads from LF2). The maximum number of mismatches allowed during alignment was two. Read and post-alignment statistics were obtained using analysis from RSeQC (Wang et al., 2012). Cufflinks analysis was performed on aligned reads to determine transcripts present and respective expression levels, denoted as Fragments Per Kilobase of exon per Million fragments mapped (FPKM) [Giardine et al., 2005]. All raw read files are stored in the NCBI SRA database under SRP051725 (<http://www.ncbi.nlm.nih.gov/sra>).

Gene Ontology Analysis

Gene ontology analysis was performed using the DAVID Bioinformatics Database (<http://david.abcc.ncifcrf.gov/>) using three main categories: Biological Processes (BP), Molecular Function (MF), and Cellular Component (CC). Three

different populations were analyzed: transcripts common between higher and lower fertility populations, unique to higher fertility, and unique to lower fertility. It is important to note that an individual transcript can be classified into multiple ontological categories during this analysis. These ontological categories are ranked based on enrichment score and not raw transcript number in each category.

Differential Abundance Analysis (Ratio Analysis)

Differential abundance analysis was performed by examining FPKM ratios between lower:higher and higher:lower fertility groups. The parameters that were required for a transcript to be considered differentially present between the populations include FPKM of ≥ 5 in both populations and a minimum 2:1 ratio between the FPKMs of the two populations.

Polymerase Chain Reaction (PCR) for Transcript Presence among Individual Sires

Select fertility candidate transcripts (N=27) were analyzed via PCR using the cDNA from nine individual sires with a range of CR scores to determine the variation in transcript presence among individual sires. These transcripts were chosen for PCR analysis from (1) differential abundance analysis results (ratio analysis, N=16; see above), (2) previously published spermatozoal transcripts that demonstrated a correlation with fertility (N=5; (Feugang et al., 2009; Arangasamy et al., 2011; Kasimanickam et al., 2013), and (3) six transcripts from previous RNA-Seq of bovine spermatozoal transcripts (Card et al., 2013). The PCR cycling conditions were 94°C for 5 min, followed by 35

cycles of 94°C for 30 sec, primer dependent annealing temperature for 30 sec then 72°C for 2 min with a final extension step at 72°C for 10 min. PCR primers are listed in Table 1. Negative controls containing no template cDNA and no enzyme were run in parallel to ensure gene-specific amplification. Along with spermatozoal cDNA samples, a testis positive control and a no template negative control were included in each experiment. PCR products were separated by 1.0-2.0% agarose gel electrophoresis depending on the size of the amplicon.

Quantitative PCR (qPCR)

qPCR was performed using the cDNA as described above to (1) validate RNA-Seq expression levels (FPKM), (2) validate expression of transcripts identified by Differential Abundance analysis in higher fertility compared to the lower fertility pools, and (3) compare transcript levels among individual sires to test if spermatozoal transcript abundance is correlated with sire fertility. For validation of FPKM, seven transcripts with ranges of 12 - 42,275 FPKM for higher fertility; 58 - 28,826 FPKM for lower fertility were chosen to represent the range of FPKM and if validated primer sets were available. For absolute quantification of FPKM values by qPCR, standard curves were made for all transcripts by serial dilutions of purified cDNA from each target gene product into 8 standard concentrations from 1×10^9 to 1×10^2 copies which allowed for absolute quantification of copy number for each transcript being validated. Copy number for the standard curves was calculated based on DNA Nanodrop concentrations and product base pair size. Relative qPCR analysis was used to validate the differential

abundance analysis of specific transcripts between fertility groups as well as measuring spermatozoal transcript levels in individual sires of different fertility rankings. The qPCR reactions were run with amplified cDNA from the higher and lower fertility pools or the nine individual sires using LightCycler 480 SYBR Green I Master Mix Kits (Roche; Indianapolis, IN). Negative template controls were run with all reactions. All reactions were run in duplicate and performed on the Roche LightCycler 480 at the Genome Sequencing Center at the University of Rhode Island. All primers used for qPCR reactions can be found in Table 1. The dissociation curves were analyzed after each run to confirm homogeneity of amplified product. A reference gene was not used for this analysis because a transcript with consistent presence among samples has not been identified by this laboratory. The mean Cp values (cycle threshold equivalents) were compared among samples. Note, the Cp values were used for correlations with fertility and higher Cp value represents lower mRNA amount due to later amplification. Correlations were determined using nonlinear regression analysis using GraphPad Prism 6.0.

RESULTS

RNA-Seq

High-throughput sequencing of the higher fertility spermatozoal RNA pool resulted in 11,832,830 x 2 paired-end 100-bp reads while lower fertility spermatozoal RNA pool sequenced 11,434,525 x 2 paired-end 100-bp reads. The total number of reads obtained

for higher fertility and lower fertility after removal of artifact reads was 10,454,393 and 10,721,121 respectively. With stringent mapping parameters, the number of reads that mapped to the genome from the higher fertility population was 1,021,536 (9.77%) with 537,102 (52.57%) uniquely mapping. In the lower fertility population, a total of 1,336,844 (12.47%) reads mapped to the bovine genome with 940,778 (70.37%) uniquely mapping.

Higher and Lower Fertility Spermatozoal Transcript Profiles

The total number of transcripts expressed in the higher and lower fertility populations were 3,227 and 5,336 respectively. Of these transcripts, 2,422 were common between the two populations while 805 were unique to the higher fertility population and 2,944 were unique to the lower fertility population (Appendix 1 and 2). The FPKM values from RNA-Seq were validated by qPCR using the same higher and lower fertility spermatozoal cDNA pools. The FPKM values were highly correlated with qPCR expression levels for both higher fertility ($r^2 = .95$) and lower fertility ($r^2 = .97$) populations (Figure 1). Excluding mitochondrial transcripts, the top 10 unique transcripts expressed in each population were identified (Tables 2 and 3).

Gene Ontology Analysis

Gene ontological analysis of the 2,422 spermatozoal transcripts common to both higher and lower fertility spermatozoal RNA populations identified 225 BP, 114 CC, and 93 MF categories (Table 4). The top three BP categories represented for this common

transcript population were translation, protein catabolic process, and protein localization. The top categories differed for spermatozoal transcripts unique to the higher and lower fertility transcript profiles. Transcripts unique to the higher fertility population (805 transcripts) included 89 BP, 23 CC, and 32 MF categories. The top three BP represented in the higher fertility population were MAPKKK cascade, RNA processing, and negative regulation of growth. Transcripts unique to the lower fertility population (2,944 transcripts) included 197 BP, 78 CC, and 75 MF categories. The top three BP categories represented were cell division, transcription, and protein transport.

Differential Abundance Analysis

Spermatozoal transcripts were differentially amplified between the higher and lower fertility populations. Transcripts with $\geq 2:1$ FPKM included 142 transcripts with higher levels in the higher fertility population and 432 transcripts that were more abundant in the lower fertility population (Appendix 1 and Appendix 2). A total of 4 transcripts were expressed at greater than 10:1 ratios for higher:lower while 37 transcripts were expressed at greater than 10:1 ratio for lower:higher populations. These results were validated by qPCR in six (*ATP5J*, *COX7C*, *GPX4*, *PABPC1*, *SPAG4*, *SPATA6*) of eight chosen transcripts.

Individual sire variation of presence of specific spermatozoal transcripts

Spermatozoal transcripts were tested for variation in abundance among individual sires. These transcripts were chosen for analysis from (1) differential abundance analysis

(N=16; see above), (2) previously published spermatozoal transcripts that demonstrated a correlation with fertility (N=5; Feugang et al., 2009; Arangasamy et al., 2011; Kasimanickam et al., 2013), and (3) six transcripts that were identified from a previous bull spermatozoal RNA-Seq transcript profile (Card et al., 2013. The presence of the transcripts varied among individual sires (N=9) with a range of CR scores from -6.7 to 3.6. Only one transcript, *PRM1*, was present in all sires examined. Two transcripts, *PRM2* and *TNP1*, were amplified in eight of the nine sires. Eleven transcripts (*CKS2*, *COX7C*, *CRISP2*, *EEF1G*, *G2E3*, *GSTM3*, *GTSF1*, *PRM1*, *PRM2*, *PSMA6*, *TNP1*) were present in five or more sires while another sixteen transcripts (*AKAP4*, *APOABP1*, *ATPIF1*, *ATP5J*, *BRP*, *CCT8*, *CETN2*, *CLGN*, *CMYC*, *COX7A2*, *DNAJB7*, *EIF1*, *GPX4*, *HEMGN*, *PCNA*, *SPATA6*) were present in less than half of the sires. The sire with the highest number of transcripts present was the sire with the lowest fertility (CR -6.7; 17 of the 27 transcripts). One transcript identified as differentially expressed in the two fertility populations, *COX7C*, was present in six of the nine bulls while *ATP5J* and *GPX4* were only present in one and two bulls, respectively.

Spermatozoal RNAs identified from previous publications

The spermatozoal transcript profiles for both fertility populations were examined for the presence of transcripts, 19 in total, previously reported to be associated with sire fertility (Feugang et al., 2009; Arangasamy et al., 2011; Kasimanickam et al., 2013). Ten previously identified transcripts were present in at least one of the fertility populations (Table 5). Nine transcripts that have been previously associated with sire fertility were

not present in either population's transcript profile reported here, including *TSSK6*,
ADAM5P, *CD36*, *CENPA*, *AKI*, *ITGB5*, *DOPPEL*, *LDC1*, and *ADIPO*.

Higher and lower fertility populations were also explored for the presence of
transcripts in which spermatozoal proteins have been associated with sire fertility in
previous literature (Table 6). A total of 11 transcripts were present in at least one of the
transcript profiles while four transcripts, including *ATPSB*, *ASPP2*, *AHSG*, and *ELSPBPI*,
were not present in either transcript profile. Among transcripts that were present in both
populations, there was a trend of lower transcript abundance in the higher fertility
population.

Individual spermatozoal transcripts and sire fertility

Of the six spermatozoal transcripts chosen for fertility evaluation from ratio
analysis and previously reported transcripts, the *COX7C* showed a strong negative
correlation with fertility by qPCR (Figure 2). The other five transcripts, *PRM1*, *GTSF1*,
PRM2, *TNPI1*, and *PSMA6*, were not correlated with sire fertility although some
transcripts had moderate correlations approaching significance.

DISCUSSION

While progress has been made in improving the ranking of dairy sires by fertility,
in vitro assays are currently not an accurate representation of *in vivo* fertility suggesting
that additional assessments of sperm function are necessary (Flowers, 2013).
Spermatozoal mRNA has been proposed as a potential diagnostic tool for fertility due to

its representation of gene expression during spermatogenesis and potential function post-fertilization (Platts et al., 2007). The consistent presence of specific transcripts among human spermatozoal samples suggests stable regulation of at least some spermatozoal mRNAs that may also be an indication of sperm quality (Lalancette et al., 2009). In this study, the oligo-dT selected spermatozoal transcript profiles of a higher fertility group (1.8 to 3.5 CR) and a lower fertility group (-2.9 to -0.4 CR) of dairy sires were compared to identify potential molecular markers for sire fertility. Using this approach, differences in the presence and abundance of spermatozoal transcripts were identified between the two fertility groups and the *COX7C* transcript had a negative correlation with sire fertility. This study demonstrates the potential utility of including spermatozoal transcripts in assays of dairy sire fertility.

The goal of this present study was to identify oligo-dT selected bovine spermatozoal RNAs that differ between higher and lower fertility sires to discover potential predictors of fertility. Spermatozoal RNA from individual bulls with a wide range of fertility scores was pooled for RNA-Seq and biological replicates were not sequenced. Due to a lack of replicates a statistical comparison of the global FPKMs between the higher and lower fertility populations was not conducted. Despite this, the RNA-Seq FPKM and differential abundance of transcripts were validated by qPCR between the two fertility populations demonstrating that this approach was valuable for identifying candidate fertility transcripts.

The spermatozoal RNA isolated in this study was amplified with oligo-dT selection to enrich for mRNAs and to reduce the amount of mtRNA and rRNA in the final transcript

338 profile (Card et al., 2013; Sendler et al., 2013; Mao et al., 2014). Pseudogene-derived
339 PIWI-interacting RNAs (piRNAs) may also be present in these spermatozoal transcript
340 profiles although not specifically examined in this study (Pantano et al., 2015). Although
341 sperm RNA is known to be fragmented (Sendler et al., 2013) and this RNA amplification
342 method does result in 3' bias of RNA-Seq reads, both fragmented and full-length
343 transcripts with a poly(A)+ tail are amplified with oligo-dT priming (Card et al., 2013;
344 Mao et al. 2014). The oligo-dT amplification was specifically chosen to identify full-
345 length spermatozoal transcripts that potentially have a biological function, including
346 during the early stages of embryogenesis. We have previously demonstrated that
347 amplification of spermatozoal RNA using oligo-dT priming does reveal a heterogeneous
348 population of fragmented and full-length transcripts (several full-length transcripts were
349 confirmed by 5' and 3' RACE PCR; (Sendler et al., 2013)). Significantly, 66% of the
350 transcripts with FPKM>1000 had reads aligned to all exons potentially indicating the
351 presence of full-length transcripts. Additionally, the estimates of sperm transcript profile
352 size (number of transcripts) and transcript identity sequenced from oligo-dT amplified
353 sperm mRNA are consistent with reports in other species using random primed or no
354 RNA amplification (Sendler et al., 2013; Das et al., 2013; Jodar et al., 2012). In a
355 previous publication from our laboratory using oligo-dT primed amplification of bovine
356 sperm RNA (Card et al., 2013), we sequenced reads for 6,166 individual sperm
357 transcripts. This estimate of sperm transcript profile size is consistent with other reports
358 using single primer isothermal amplification (SPIA) or random primed sperm RNA
359 amplification methods in other species. For example, greater than 3,000 sperm

transcripts have been identified in human sperm (unamplified RNA; (Jodar et al., 2012 and SPIA amplification; (Sendler et al., 2013) and 6,761 transcripts were identified in stallion sperm (random primed RNA amplification; (Das et al., 2013). While the extent that biological variation contributes to the number of transcripts in sperm is still under investigation, the number of transcripts reported in this current manuscript for the higher fertility sires (3,227 transcripts) and the lower fertility sires (5,366 transcripts) is also consistent with the estimates of transcript population size in these previous publications. The exact proportion of poly (A)+ tailed transcripts in the total sperm RNA (fragmented and intact) population remains to be determined.

As previously mentioned, the spermatozoal RNA population contains full length transcripts, that could potentially be translated, but the function of these transcripts, if any, remains unknown (Card et al., 2013; Sendler et al., 2013). As previously reported, several of the bovine spermatozoal transcripts amplified with this method are orthologous to transcripts in human and stallion that were amplified with random priming suggesting that oligo-dT amplified populations are similar to transcript profiles reported using different amplification methods (Jodar et al., 2012). While the focus of this study was not a comparison of amplification methods, a different RNA amplification methods may result in a modified spermatozoal RNA profile that would include fragmented transcripts that have not retained the 3' end (Mao et al., 2014).

While several common transcripts were sequenced between the higher and lower fertility groups that have a function in spermatogenesis, differences in the presence and/or amount of other transcripts were found between the two fertility groups suggesting

that these variations could potentially be used as a fertility assessment. Gene ontological analysis of the unique transcripts revealed that higher fertility bulls had enrichment for regulatory transcripts (including regulation of growth and protein kinase activity) while the lower fertility population is not as enriched globally and has a broader representation of cellular process. The role in spermatogenesis and fertility for several of the top 10 transcripts unique for each population is not known and requires further investigation.

The differences in the spermatozoal transcript profiles between the fertility groups found in this study could be a reflection of sperm function in several ways. The spermatozoal mRNA population may represent gene expression that has occurred during spermatogenesis and therefore may reflect inefficiencies in transcription and/or translation that contribute to decreased sperm quality (Platts et al., 2007). For example, spermatozoal transcripts associated with the ubiquitin-proteosome pathway, acrosome formation, and proper tail formation differed in abundance between RNA isolated in normal semen and teratozoospermic samples (Platts et al., 2007). Spermatid and spermatocyte-specific transcripts are also absent in teratozoospermatozoic samples, thus indicating errors in the later stages of spermatogenesis (Platts et al., 2007). If some spermatozoal transcripts were not efficiently translated during the later stages of spermatogenesis, they could remain within the spermatozoa resulting in greater transcript abundance in the fully differentiated cell. Infertile males have increased levels of protamine 1 (*PRM1*) and protamine 2 (*PRM2*) mRNA and decreased PRM1 and PRM2 protein levels as well as abnormal *PRM1/PRM2* ratios in their spermatozoa (Aoki et al., 2006a; Aoki et al., 2006b) suggesting that improper translation of *PRM1* and *PRM2* can impair fertility. For

example, an increase in transcripts involved in protein production in the lower fertility sires reported here may indicate inefficient translation of these transcripts leading to reduced protein production during spermatogenesis.

Additionally, a role of sperm-borne mRNAs in early embryonic development may also contribute to the utility of spermatozoal transcripts in fertility assessment based on their function. With the paternal genome, spermatozoa deliver coding and non-coding RNAs, including *CLU*, *PRM1*, *PRM2*, *DDX3Y* and *PLCz*, to the oocyte at fertilization (Ostermeier et al., 2004; Swann et al., 2006; Kempisty et al., 2007). As previously mentioned, the sperm RNA population also contains full length transcripts, that could potentially be translated, but the function of these transcripts, if any, remains unknown (Card et al., 2013; Sandler et al., 2013).

Variation in spermatozoal transcript presence was also found among the individual sires examined in this study. Of the transcripts tested, one transcript, *PRM1* was expressed in all individuals. Some transcripts were expressed in less than half such as *SPAG4*, *ATP5J* and *GPX4* while others (including *PSMA6*, *COX7C* and *PRM1*) were expressed in over half (≥ 5) of the sires. Specific human spermatozoal transcripts isolated from samples with a range of semen quality also demonstrate consistent expression, by ratio, suggesting that at least a subpopulation of spermatozoal transcripts is stably regulated (Lalancette et al., 2009; Jodar et al., 2012; Ostermeier et al., 2002). Further investigation and identification of uniformly present transcripts and transcript variation amongst individuals is necessary.

Spermatozoal transcript cytochrome oxidase subunit 7C (*COX7C*) had a negative correlation with sire fertility. The protein *COX7C* is a subunit involved in the terminal step in the electron transport chain leading to ATP synthesis in the inner mitochondrial membrane (Seelan et al., 1997). We hypothesize that the greater amounts of *COX7C* present in the lower fertility sires may represent inefficient translation of this transcript and therefore impairment of mitochondrial function during the later stages of spermatogenesis.

The other five transcripts examined (*PRM1*, *PRM2*, *TNPI*, *GTSF1* and *PSMA6*) did not have a strong correlation with sire fertility although there was a trend of higher for greater abundance in lower fertility sires for these transcripts. The *PRM1* and *PRM2* mRNA and protein levels are positively correlated with spermatozoa concentration, motility, fertilization potential, and embryo quality (Kempisty et al., 2007; Depa-Martynow et al., 2012) as these proteins are essential for proper DNA compaction in the spermatozoal head (Oliva, 2006). Infertile males have increased levels of *PRM1* and *PRM2* mRNA and decreased *PRM1* and *PRM2* protein levels as well as abnormal *PRM1/PRM2* ratios (Aoki et al., 2006a; Aoki et al., 2006b) suggesting that improper translation of *PRM1* and *PRM2* can impair fertility. The *PRM2* spermatozoal transcript has previously been shown to have greater abundance in lower fertility sires which is in agreement with our results in individual sires (Lalancette et al., 2008). Because the *PRM2* protein is not expressed in bulls, the significance of these results is not known. Transition protein 1 (*TNPI*) is one of two transition proteins that replace the histones during the later stages of spermatogenesis (Oliva, 2006). Gametocyte-specific factor 1

(*GTSF1*) is a retrotransposon suppressor essential for spermatogenesis (Yoshimura et al., 2009). Proteasome subunit alpha type 6 (*PSMA6*) is one of the seven alpha subunits of the proteasome, a complex that facilitates protein degradation (Gomes, 2013). Greater abundance of these spermatozoal mRNAs may also be an indication of inefficient translation and therefore reduced sperm quality.

In previous studies, individual spermatozoal transcripts *AK1*, *ITGB5*, *DOPPEL*, *NGF*, *TIMP2*, *PLCZ1*, *CRISP2*, *PEBP1*, *CD36*, *CENPA*, *ADIPO*, *ADIPOR1*, and *ADIPOR2* had positive correlations with sire fertility (Lalancette et al., 2008; Feugang et al., 2009; Arangasamy et al., 2011; Kasimanickam et al., 2013) while *CCT8* and *PRM2* had negative correlations with sire fertility (Lalancette et al., 2008; Arangasamy et al., 2011). Only three of these previously identified transcripts, *PLCZ1*, *CRISP2*, *PEBP1*, were expressed in the spermatozoal transcript profiles reported here, and were not correlated with fertility in this group of bulls. Previous studies of spermatozoal transcripts in subfertile males using microarray (Lalancette et al., 2008; Feugang et al., 2009) reported differences among samples from sires with a range of fertility although the transcripts and the gene ontology categories reported in this paper differ from those reported in previous studies. These inconsistencies could be due to the transcript identification methodology (RNA-Seq vs microarray), RNA isolation methods, or may be representative of biological variation in transcript presence among individuals. Studies directly comparing RNA isolation and sequencing methods while evaluating semen from a large number of sires is warranted.

Previous proteomic studies have also evaluated spermatozoal protein differences amongst sires of varying fertility (Peddinti et al., 2008). Higher fertility sires have an abundance of ATP5B, ENO1, ASPP2, GPX4, AHSX, AKI, and PEBP1 proteins, while CCT5, CCT8, ELSPBP1, PSMA6, BSP1, VDAC1, UQCRC2, and ROPORIN-1 are lower in higher fertility sires (D'Amours et al., 2010; Park et al., 2012). Spermatozoal transcripts for each of these proteins were identified in the transcript profiles reported here. Our results show a trend for higher transcript abundance in lower fertility sires for many proteins that have been positively correlated with fertility, thus potentially indicating inefficiencies in translation leading to lower fertility in these individuals. Further studies coupling RNA-Seq data with proteomic data can provide more insight into translation efficiency of spermatozoal transcripts.

CONCLUSIONS

The work presented here demonstrates the utility of high-throughput sequencing methods in improving livestock breeding protocols by the identification of differences in the presence and abundance of spermatozoal transcripts between the higher and lower fertility groups showing the potential utility of including oligo-dT selected spermatozoal transcripts in dairy sire fertility assays. Inclusion of spermatozoal RNA abundance in multi-parameter sire fertility assays may improve the accuracy of sire fertility rankings to ensure that the most fertile sires are used for AI to improve herd pregnancy rates. Further work needs to be conducted before spermatozoal transcripts are routinely used in fertility assays including testing of additional candidate spermatozoal transcripts for correlation with fertility as it is likely that a number of spermatozoal traits need to be evaluated

concurrently to get an accurate assessment of sire fertility (Kastelic et al., 2013).
Additionally, screening of a larger population of individual bulls will determine variation
in the spermatozoal transcript profile among bulls and within individuals over time. This
larger scale analysis should then identify candidate spermatozoal fertility transcripts that
can be assayed for in a cost-efficient test (RT-PCR) conducted during ejaculate analysis
prior to breeding.

**Funding: This project was supported by Genex Cooperative, Inc. and by the
Agriculture and Food Research Initiative Competitive Grant no. 2011-67016-20073
from the USDA National Institute of Food and Agriculture to BLS. This research is
based in part upon work conducted using the Rhode Island Genomics and
Sequencing Center which is supported in part by the National Science Foundation
under EPSCoR Grants Nos. 0554548 & EPS-1004057.**

REFERENCES

- Aoki, V.W., Liu, L., Jones, P., Hatasaka, H.H., Gibson, M., Peterson, C.M., Carrell, D.T., 2006. Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertil Steril.* 86,1408-1415.
- Aoki, V.W., Liu, L., Carrell, D.T., 2006. A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. *Mol Hum Reprod.* 1,41-50.
- Arangasamy, A., Kasimanickam, V.R., DeJarnette, J.M., Kasimanickam, R.K., 2011. Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. *Therio.* 76, 570-577.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Molec Biol.* Chapter 19: Unit 19.10.1-21.
- Card, C.J., Anderson, E.J., Zamberlan, S., Krieger, K.E., Kaproth, M., Sartini, B.L., 2013. Cryopreserved bovine spermatozoal transcript profile as revealed by high-throughput ribonucleic acid sequencing. *Biol Reprod.* 88, 1-9.
- D'Amours, O., Frenette, G., Fortier, M., Leclerc, P., Sullivan, R., 2010. Proteomic comparison of detergent-extracted sperm proteins from bulls with different fertility indexes. *Reproduction.* 139,545-556.
- Das, P.J., Paria, N., Gustafson-Seabury, A., Vishnoi, M., Chaki, S.P., Love, C.C., Varner, D.D., Chowdhary, B.P., Raudsepp, T., 2010. Total RNA isolation from stallion sperm and testis biopsies. *Therio.* 74,1099-1106.
- Das, P.J., McCarthy, F., Vishnoi, M., Paria, N., Gresham, C., Li, G., Kachroo, P., Sudderth, A.K., Teague, S., Love, C.C., Varner, D.D., Chowdhary, B.P., Raudsepp, T., 2013. Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-Seq. *PLoS ONE.* 8(2), e56535.
- Depa-Martynow, M., Kempisty, B., Jagodzinski, P., Pawelczyk, L., Jedrzejczak, P., 2012. Impact of protamine transcripts and their proteins on the quality and fertilization ability of sperm and the development of preimplantation embryos. *Reprod Biol.* 12,57-72.
- Feugang, J.M., Kaya, A., Page, G.P., Chen, L., Mehta, T., Hirani, K., Nazareth, L., Topper, E., Gibbs, R., Memili, E., 2009. Two-stage genome-wide association study

identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics*. 10,176-185.

Feugang, J.M., Rodriguez-Osorio, N., Kaya, A., Wang, H., Page, G., Ostermeier, G.C., Topper, E.K., Memili, E., 2010. Transcriptome analysis of bull spermatozoa: implications for male fertility. *Reprod Biomed Online*. 21, 312-324.

Flowers, W., 2013. Sperm characteristics that limit success for fertilization. *J Anim Sci*. 91(7), 3022-3029.

Gadea, J., Selles, E., Marco, M.A., 2004. The predictive value of porcine seminal parameters on fertility outcome under commercial conditions. *Reprod Domest Anim*. 39, 303-308.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A., 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res*. 15, 1451-1455.

Gilbert, I., Bissonnette, N., Boissonneault, G., Vallee, M., Robert, C., 2007. A molecular analysis of the population of mRNA in bovine spermatozoa. *Reproduction*. 133, 1073-1086.

Goecks, J., Nekrutenko, A., Taylor, J., and The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. 11(8), R86.

Gomes, A.V., 2013. Genetics of proteasome diseases. *Scientifica (Cairo)*. 637629, 1-30.

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

Kasimanickam, V., Kasimanickam, R., Arangasamy, A., Saberivand, A., Stevenson, J.S., Kastelic, J.P., 2012. Association between mRNA abundance of functional sperm function proteins and fertility of Holstein bulls. *Theriogenology*. 78, 2007-2019.

Kasimanickam, V.R., Kasimanickam, K.R., Kastelic, J.P., Stevenson, J.S., 2013. Associations of adiponectin and fertility estimates in Holstein bulls. *Theriogenology*. 79, 766-777.

Kastelic, J.P., 2013. Male involvement in fertility and factors affecting semen quality in bulls. *Animal Frontiers*. 3(4), 20-25.

- Kempisty, B., Antosik, P., Bukowska, D., Jackowska, D., Lianeri, M., Jaskowski, K.M., Jagodzinski, P.P., 2008. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. *Reprod Fertil Dev.* 20, 513-518.
- Kempisty, B., Depa-Martynow, M., Lianeri, M., Jedrzejczak, P., Darul-Wasowicz, A., Jagodzinski, P.P., 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. *Folia Histochem Cytobiol.* 45, 109-113.
- Lalancette, C., Thibault, C., Bachand, I., Caron, N., Bissonnette, N., 2008. Transcriptome analysis of bull semen with extreme nonreturn rate: use of suppression-subtractive hybridization to identify functional markers for fertility. *Biol Reprod.* 78, 618-635.
- Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., Krawetz, S.A., 2009. Identification of human sperm transcripts as candidate markers of male fertility. *J Mol Med.* 87, 735-748.
- Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3), R25.
- Mao, S., Sendler, E., Goodrich, R.J., Hauser, R., Krawetz, S.A., 2014. A comparison of sperm RNA-seq methods. *Sys Bio in Reprod Med.* 60, 308-315.
- Oliva, R., 2006. Protamines and male infertility. *Hum Reprod.* 12, 417-435.
- Omran, H.M., Bakhiet, M., Dashti, M.G., 2013. DNA integrity is a critical molecular indicator for the assessment of male infertility. *Mol Med Rep.* 7(5), 1631-1635.
- Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., Krawetz, S.A., 2002. Spermatozoal RNA profiles of normal fertile men. *Lancet.* 360, 772-777.
- Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., Krawetz, S.A., 2004. Delivering spermatozoan RNA to the oocyte. *Nature.* 429, 154.
- Pantano, L., Jodar, M., Bak, M., Ballesca, J.L., Tommerup, N., Oliva, R., Vavouri, T., 2015. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *RNA.* 21, 1085-1095.
- Park, Y.J., Kwon, W.S., Oh, S.A., Pang, M.G., 2012. Fertility-related proteomic profiling bull spermatozoa separated by Percoll. *J Proteome Res.* 11, 4162-4168.

Peddinti, D., Nanduri, B., Kaya, A., Feugang, J.M., Burgess, S.C., Memili, E., 2008. Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Systems Biol.* 2, 19.

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

Seelan, R.S., Grossman, L.I., 1997. Structural organization and promoter analysis of the bovine cytochrome c oxidase subunit VIIc gene a functional role for YY1. *J Biol Chem.* 272, 10175-10181.

Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., Krawetz, S.A. 2013. Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Res.* 1-14.

Swann, K., Saunders, C.M., Rogers, N.T., Lai, F.A., 2006. PLC ζ (zeta): A sperm protein that triggers Ca²⁺ oscillations and egg activation in mammals. *Cell Dev Biol.* 17, 264-273.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 7(3), 562-578.

Wang, L., Wang, S., Li, W., 2012. RSeQC: Quality Control of RNA-seq experiments. *Bioinformatics.* 28(16), 2184-2185.

Yoshimura, T., Toyoda, S., Kuramochi-Miyagaw, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T., Miyazaki, J., 2009. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testis. *Dev Biol.* 335, 216-227.

676
677
678
679
680
681
682
683
684
685
686

Table 1. Bovine primer sequences

Gene Symbol	Genbank ID	Assay	Primer Sequences		Product Size (bp)
			Forward	Reverse	
<i>ACRV1</i>	BC109966	qPCR	5'- TGAACAGGCCGTAGGTGAAC -3'	5'- CAGCTTAATGGTGGGCCTGA -3'	174
<i>AKAP4</i>	AF100170	Ind	5'- AGGGGTCAGTGTGCCTTTTC -3'	5'- TCCAGACGTAGGCTCTGAGG -3'	234
<i>APOABP1</i>	AY528250	Ind	5'- CCCTGACACGGAATGTGTCT -3'	5'- AGAGAGGCACTAAGTCTTTATTGA -3'	80
<i>ATPIF1</i>	NM_175816	Ind; Ratio	5'- CTTTGGCTCGGAATCGGGAG -3'	5'- GCCAGCTGTTCTTTAGCACG -3'	132
<i>ATP5J</i>	NM_174717	Ind; qPCR; Ratio	5'- CTCCTGTCCGGATCACCAT -3'	5'- TCCTCCAGATGTCTGTCGCT -3'	195
<i>BRP</i>	NM_001012682	Ind	5'- CCAGGCTTTAGGCATCACCA -3'	5'- GGCGCCTACTTTGTCTCCTGT -3'	94
<i>CCT8</i>	NM_001033609	Ind	5'- TGGATCAGATCATCATGGCAA -3'	5'- AGGGTGCAAGATACAAACACAAAA -3'	202
<i>CETN2</i>	NM_001038515	Ind; Ratio	5'- GTCCGGGATGGCCTCTAACT -3'	5'- TCCTGTTTCTGCTCTTCGGT -3'	102
<i>CLGN</i>	BC103401	Ind	5'- ATCCCTGATGCTTCTGCTGTC -3'	5'- ATTCACCACACCCAATCCGA -3'	243
<i>CKS2</i>	BC105331	Ind	5'- CCTCGCTTCGCTCTAGTCAG -3'	5'- AGGATGTGTGGTTCTGTTCA -3'	245
<i>CMYC</i>	BC109848	Ind	5'- CCACCTAGGAGGAGAAACG -3'	5'- ACAGTTAACGTGTGATAGGTGAAT -3'	168
<i>COX7A2</i>	DQ347636	Ind	5'- AACTGGCTGTGGCTTCGTTT -3'	5'- TGCTTTATTGGTGGCAGCTAA -3'	204
<i>COX7C</i>	NM_175831	Ind; qPCR; Ratio	5'- GAGGAGGGTCCAGGGAAGAATA -3'	5'- TTTACGACAGTTGGTGTCTTACT -3'	129
<i>CRISP2</i>	BC109478	Ind; qPCR	5'- CGGCCGCTCTGCAACAGAAG -3'	5'- GGAAGCAGCACAGCGGTCAGA -3'	120
<i>DNAJB7</i>	BC142060	Ind	5'- GCACTTAAATGGCACCCAGA -3'	5'- TGGCTTACGGAATGTGAAGC -3'	204
<i>EEF1G</i>	AB098752	Ind; Val	5'- ATCCAGTTTCCGCCATGTGT -3'	5'- GTTGCAACGCTCATCACTGG -3'	198
<i>EIF1</i>	BC103170	Ind; Val	5'- AAGGGTGATGATCTGCTTCCTGCT -3'	5'- AACTGGCATATGTTCTTGCGCTGG -3'	235
<i>G2E3</i>	NM_001038671	Ind	5'- TCACTGGCAATTTTGCCTCA -3'	5'- TGCACCTGCAAACAGTCTCT -3'	187
<i>GPX4</i>	NM_174770	Ind; qPCR; Ratio	5'- CAATGTGGCCTCGCAATGAG -3'	5'- CATTACTCCCTGGCTCCTGC -3'	140
<i>GSTM3</i>	BC112491	Ind	5'- TGTGCCGTTTTGAGGCTTTGGAG -3'	5'- GGGCCATCTTGTGTGACAGGCAT -3'	112
<i>GTSF1</i>	BC102713	Ind; qPCR; Ratio	5'- CAGGTTCTCGGGCTGAAAT -3'	5'- ACTATGTTGCTTGCAGGGCT -3'	239
<i>HEMGN</i>	BC109747	Ind	5'- CAGCAAGAAATGGCTGTGCC -3'	5'- CAGGTGCATAGGCTTCAGGT -3'	247
<i>HMGB4</i>	BC109790	Val	5'- AGCTGGTCGGTGGTGCAGGT -3'	5'- GCAAGCATGTCTTCCGGGC -3'	167
<i>PABPC1</i>	NM_174568	Ratio	5'- ACATGGATGATGAGCGCCTT -3'	5'- TCTCATCCACAGCTTTTGTGTC -3'	150
<i>PCNA</i>	BC103068	Ind	5'- GAACCTCACCAGCATGTCCA -3'	5'- ACGTGTCCGCGTTATCTTCA -3'	86
<i>PLCZ1</i>	BC114836	Val	5'- CGGGTGGTTCGGAATCCCACTCT -3'	5'- AATTCCCTGGCTGCCAACTTTGT -3'	194
<i>PRM1</i>	BC108207	Ind; qPCR; Val; Ratio	5'- AAGAAGATGTGCGAGACGAAGGAG -3'	5'- ACAGGTGGCATTGTTTCGTTAGCAG -3'	228
<i>PRM2</i>	NM_174157	Ind; qPCR	5'- ATGCGGAAGGCAGCTTAAG -3'	5'- CTCAAGATCTCGTGGGCTCC -3'	109

<i>PSMA1</i>	BC102216	qPCR	5'- TGGTTAAGCATGGTCTGCGT -3'	5'- GGTGAGTAGGCTGTGCCTT -3'	187
<i>PSMA6</i>	BC110260	Ind; qPCR	5'- ACAGTGGAAGTGGCGTCACGG -3'	5'- ACAGGCAAGTGGCGTCACGG -3'	205
<i>SEC61G</i>	BC102186	Val	5'- GCAGACGCGGAGCAGACATCA -3'	5'- AGCGAATCCTATTGCTGTTGCCA -3'	155
<i>SPAG4</i>	NM_001076507	qPCR; Ratio	5'- CCACGGTTATCTTGAGCCAG -3'	5'- GATGCTGCAGAGTGATGTCG -3'	123
<i>SPATA6</i>	NM_001046371	Ind; Ratio	5'- GATGAGATCCATGACCGGGTA -3'	5'- TAAGAGGCTGCCCTGTTGGA -3'	230
<i>TNP1</i>	NM_174199	Ind; qPCR	5'- AAACGCTGTGACGATGCCAA -3'	5'- CTTGGCAGTCCCCTTCTGTT -3'	118

Ind = individual variation PCR; qPCR = qPCR 9 individuals; Ratio = Ratio diff validation; Val = qPCR copy/FPKM validation

Table 2. Top 10 spermatozoal transcripts unique to higher fertility population based on FPKM

Transcript	Gene Name	Accession Number	FPKM
<i>MIR708</i>	microRNA 708	n/a	200
<i>C4H7orf55</i>	chromosome 4 open reading fram, human C7orf55	NM_001076997	122
<i>VSNL1</i>	visinin-like 1	NM_174490	106
<i>AKR1B1</i>	aldo-keto reductase family 1, member B1 (aldose reductase)	NM_001012519	95
<i>APOA1BP</i>	apolipoprotein A-I binding protein	NM_205796	91
<i>SQRDL</i>	sulfide quinone reductase-like (yeast)	NM_001040511	91
<i>CD28</i>	CD28 molecule	NM_181004	87
<i>MPI</i>	mannose phosphate isomerase	NM_001035284	65
<i>RNGTT</i>	RNA guanylyltransferase and 5'-phosphatase	NM_001046085	64
<i>FCF1</i>	FCF1 small subunit (SSU) processome component homolog	NM_001037452	63

Table 3. Top 10 spermatozoal transcripts unique to lower fertility population based on FPKM

Transcript	Gene Name	Accession Number	FPKM
<i>RN5-8S1</i>	5.8S ribosomal RNA	NR_036643	348
<i>VDAC3</i>	voltage-dependent anion channel 3	NM_174729	300
<i>DEFB122A</i>	beta-defensin 122a	NM_001102339	247
<i>ATP6V1D</i>	ATPase, H ⁺ transporting, lysosomal 34kDa, V1 subunit D	NM_001075141	228
<i>LOC526524</i>	FK506 binding protein 1A, 12kDa	NM_001035456	203
<i>UQCRI0</i>	ubiquinol-cytochrome c reductase complex 7.2 kDa protein	NM_001113723	195
<i>PSMA1</i>	proteasome (prosome, macropain) subunit, alpha type, 1	NM_001035310	171
<i>INSL6</i>	insulin-like 6	NM_001077521	162
<i>C10H14orf166</i>	chromosome 14 open reading frame 166 ortholog	NM_001035280	161
<i>METAP2</i>	methionyl aminopeptidase 2	NM_001040493	157

743
744
745

Table 4. Top 10 Biological Processes (BP) for spermatozoal transcripts common and unique to the higher and lower fertility populations

	Biological Processes	% of Transcripts
Unique to Higher Fertility	MAPKKK cascade	9
	RNA processing	23
	negative regulation of growth	7
	regulation of growth	14
	mRNA processing	14
	negative regulation of gene expression	16
	macromolecule catabolic process	25
	bone development	8
	regulation of protein kinase activity	12
	epithelium development	10
Unique to Lower Fertility	cell division	39
	transcription	112
	protein transport	94
	cellular macromolecule catabolic process	71
	intracellular transport	69
	protein localization	98
	multicellular organism reproduction	41
	proteolysis involved in cellular protein catabolic process	60
	intracellular protein transport	50
	generation of precursor metabolites and energy	52
	cell cycle	61
Common Transcripts	translation	94
	protein catabolic process	77
	protein localization	109
	modification-dependent protein catabolic process	68
	protein folding	44
	protein transport	99
	cellular macromolecule catabolic process	77
	spermatogenesis	34
	macromolecule catabolic process	80
	intracellular transport	71

Table 5. Expression levels in higher and lower fertility transcript profiles of spermatozoal transcripts previously correlated with sire fertility

	Higher Fert. FPKM	Lower Fert. FPKM	Publication
<i>CCT8</i>	0	18	Arangasamy et al. 2011
<i>CRISP2</i>	661	1366	"
<i>PEBP1</i>	5	41	"
<i>NGF</i>	11	2	Kasimanickam et al. 2012
<i>ODF2</i>	50	35	"
<i>PLCZ1</i>	27	59	"
<i>SNRPN</i>	0.0001	3	"
<i>TIMP2</i>	4	14	"
<i>ADIPOR1</i>	4	3	Kasimanickam et al. 2013
<i>ADIPOR2</i>	0	0.66	"

Table 6. Expression levels in higher and lower fertility transcript profiles of spermatozoal proteins previously correlated with sire fertility

	Higher Fert. FPKM	Lower Fert. FPKM	Publication
<i>AKI</i>	10	20	D'Amours et al. 2010
<i>BSP1</i>	11	67	"
<i>CCT5</i>	3	9	"
<i>CCT8</i>	0	18	"
<i>PEBP1</i>	5	41	"
<i>PSMA6</i>	537	1296	"
<i>ENO1</i>	6	15	Park et al. 2012
<i>GPX4</i>	18	138	"
<i>ROPN1</i>	85	409	"
<i>UQCRC2</i>	7	18	"
<i>VDAC1</i>	0	0.62	"

Figure Captions

1. Correlation of qPCR transcript copy number and FPKM for higher and lower fertility populations by nonlinear regression analysis with the log-log curve fit. Axes are base 10 log scale.

2. Correlation between spermatozoal transcript abundance and fertility among individual sires. Note, the Cp values were used for correlations with fertility and higher Cp value represents lower mRNA amount due to later amplification.

839

840

841 Appendix 1. Spermatozoal transcripts with an abundance of > 2:1 in lower:higher fertility

842 populations based on FPKM

843

Official Gene Symbol	Gene Long Name	Higher FPKM	Lower FPKM	Lower: Higher Ratio
<i>RPL30</i>	#N/A	8.06	420.66	52.22
<i>CETN1</i>	centrin, EF-hand protein, 1	9.31	433.41	46.57
<i>MRPS33</i>	#N/A	6.45	209.46	32.49
<i>PTTG1</i>	protein tyrosine phosphatase type IVA, member 1	6.62	194.87	29.43
<i>C28H10orf53</i>	UPF0728 protein C10orf53 homolog	12.68	333.48	26.30
<i>LOC781895</i>	hypothetical protein LOC768323	39.07	980.19	25.09
<i>FANCD2OS</i>	#N/A	7.00	135.75	19.39
<i>HSFY2</i>	#N/A	13.56	259.74	19.16
<i>COPS5</i>	similar to COP9 signalosome subunit 5	8.57	160.08	18.67
<i>DYNLL1</i>	#N/A	12.17	222.01	18.24
<i>TOMM7</i>	translocase of outer mitochondrial membrane 5 homolog (yeast)	6.41	116.32	18.15
<i>NDUFA7</i>	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6	5.80	93.32	16.09
<i>H2AFZ</i>	H2A histone family, member Z	11.16	178.18	15.97
<i>CDO1</i>	cysteine dioxygenase, type I	8.56	134.59	15.72
<i>HINT1</i>	#N/A	6.31	98.05	15.55
<i>RPS18</i>	ribosomal protein S17	14.02	216.93	15.47
<i>SRR</i>	signal recognition particle 72kDa	8.17	124.32	15.23
<i>MGC148328</i>	hypothetical protein MGC137055	1625.53	23928.30	14.72
<i>RPL34</i>	ribosomal protein L32	14.58	191.73	13.15
<i>ERH</i>	similar to enhancer of rudimentary homolog	7.77	99.55	12.80
<i>IFT57</i>	intraflagellar transport 57 homolog (Chlamydomonas)	5.67	65.99	11.64
<i>RPL14</i>	ribosomal protein L13a	21.60	250.16	11.58
<i>SUMO2</i>	sulfotransferase family, cytosolic, 1A, phenol-preferring	8.71	95.97	11.02
<i>CD59</i>	CD59 molecule, complement regulatory protein	12.22	134.27	10.99
<i>MRFAP1</i>	#N/A	13.90	152.68	10.99
<i>CEP57</i>	centrosomal protein 57kDa	6.12	66.63	10.89
<i>ATP5B</i>	ATP synthase, H ⁺ transporting, mitochondrial F1 complex, beta	8.10	84.88	10.48
<i>NDUFA5</i>	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4	9.05	93.86	10.37
<i>PGK2</i>	FGF receptor activating protein 1	35.68	369.93	10.37
<i>PABPC1</i>	proliferation-associated 2G4, 38kDa	22.22	228.22	10.27
<i>CCDC23</i>	coiled-coil domain containing 23	8.50	85.90	10.11
<i>FAIM2</i>	Fas apoptotic inhibitory molecule 2	8.33	83.50	10.03
<i>KLHL10</i>	#N/A	14.12	140.42	9.94
<i>LOC788142</i>	#N/A	7.26	70.41	9.70
<i>ASB17</i>	ankyrin repeat and SOCS box-containing 17	21.52	205.35	9.54
<i>UBE2K</i>	ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast)	7.65	72.61	9.49

<i>C8H9orf135</i>	#N/A	12.17	115.21	9.47
<i>CHPT1</i>	choline phosphotransferase 1	6.65	63.00	9.47
<i>CA6</i>	carbonic anhydrase VI	5.65	52.50	9.28
<i>PSENEN</i>	#N/A	8.19	76.04	9.28
<i>RPS14</i>	ribosomal protein S12	8.73	81.05	9.28
<i>COX7C</i>	cytochrome c oxidase subunit VIIc	52.63	484.73	9.21
<i>GHITM</i>	#N/A	6.74	61.96	9.20
<i>SLIRP</i>	solute carrier family 30 (zinc transporter), member 3	53.19	482.84	9.08
<i>UCHL3</i>	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	9.85	88.44	8.97
<i>C1H21orf59</i>	chromosome 21 open reading frame 59 ortholog	14.51	128.88	8.88
<i>CLIC4</i>	chloride intracellular channel 4	14.66	129.83	8.86
<i>RPS4Y1</i>	ribosomal protein S4, Y-linked 2	11.98	105.73	8.82
<i>ATP5J</i>	ATP synthase, H+ transporting, mitochondrial F0 complex	50.28	441.92	8.79
<i>PPP1R2</i>	protein phosphatase 1G (formerly 2C), magnesium-dependent	47.53	414.36	8.72
<i>PLK1S1</i>	phospholipase D1, phosphatidylcholine-specific	5.28	45.53	8.62
<i>ATP5O</i>	ATP synthase, H+ transporting, mitochondrial F1 complex, O	20.64	175.16	8.49
<i>TRIM59</i>	tripartite motif-containing 36	13.99	116.44	8.32
<i>RPS24</i>	ribosomal protein S23; similar to ribosomal protein S23	54.06	449.04	8.31
<i>RPL10L</i>	#N/A	27.93	229.88	8.23
<i>PHYHIPL</i>	phosphoglycolate phosphatase	18.77	153.54	8.18
<i>GNAS</i>	GNAS complex locus	6.58	53.73	8.17
<i>NUPL2</i>	nucleobindin 1	13.63	110.49	8.11
<i>C21H15orf63</i>	#N/A	6.28	50.53	8.05
<i>LZTFL1</i>	lysozyme-like 2	6.47	52.02	8.05
<i>PEBP1</i>	#N/A	5.28	40.88	7.74
<i>SAMD13</i>	serum amyloid A4, constitutive	27.50	212.74	7.74
<i>RPS19</i>	ribosomal protein S18	10.32	79.29	7.68
<i>CSDE1</i>	cold shock domain containing E1, RNA-binding	22.61	173.34	7.67
<i>LOC783012</i>	hypothetical protein LOC781895	121.35	926.74	7.64
<i>NUPR1L</i>	nucleoporin like 2	13.20	100.80	7.63
<i>SPCS1</i>	spermatogenesis associated, serine-rich 2	5.25	39.83	7.58
<i>GPX4</i>	glutathione peroxidase 4 (phospholipid hydroperoxidase)	18.29	138.68	7.58
<i>LOC768323</i>	testis-specific histone 2a-like	194.47	1472.79	7.57
<i>TXNDC17</i>	thioredoxin	7.71	57.29	7.43
<i>PSMA4</i>	presenilin enhancer 2 homolog (C. elegans)	6.85	50.87	7.43
<i>ATPIF1</i>	ATPase inhibitory factor 1	6.31	46.83	7.43
<i>SELK</i>	Sec61 gamma subunit	36.11	268.22	7.43
<i>MS4A14</i>	membrane-spanning 4-domains, subfamily A, member 13	8.49	62.32	7.34
<i>RPS23</i>	ribosomal protein S21	11.43	82.94	7.26
<i>PCBP1</i>	poly (ADP-ribose) polymerase family, member 11	5.62	40.56	7.22
<i>ADAM32</i>	ADAM metalloproteinase domain 32	17.83	128.51	7.21
<i>SOD1</i>	sine oculis binding protein homolog (Drosophila)	9.69	68.96	7.12
<i>C17H5orf52</i>	chromosome 5 open reading frame 52 ortholog	10.92	76.04	6.96

<i>KHDRBS3</i>	potassium channel modulatory factor 1	9.23	64.28	6.96
<i>SH3GLB1</i>	SH3 domain binding glutamic acid-rich protein like 3	8.10	55.98	6.91
<i>C7H5orf48</i>	#N/A	56.24	385.08	6.85
<i>NEDD8</i>	NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa	10.63	72.36	6.81
<i>STK17A</i>	signal transducer and activator of transcription 1	13.08	88.03	6.73
<i>ARL3</i>	ADP-ribosylation factor-like 3	18.51	122.73	6.63
<i>RNF181</i>	ring finger protein 138	12.51	82.60	6.60
<i>PSMB3</i>	proteasome (prosome, macropain) subunit, alpha type, 6	8.70	57.43	6.60
<i>DYDC1</i>	DPY30 domain containing 1	44.45	292.98	6.59
<i>FTH1</i>	#N/A	13.19	85.70	6.50
<i>STT3B</i>	STT3, subunit of the oligosaccharyltransferase complex,	13.94	90.19	6.47
<i>RNF138</i>	similar to ring finger protein 103	15.15	97.93	6.46
<i>HMGN3</i>	high mobility group nucleosomal binding domain 3	5.19	32.14	6.19
<i>OXCT2</i>	OTU domain, ubiquitin aldehyde binding 2	6.54	40.46	6.19
<i>DCUN1D1</i>	DCN1, defective in cullin neddylation 1, domain containing 1	165.17	1022.28	6.19
<i>HSBP1</i>	heat shock factor binding protein 1	52.82	324.18	6.14
<i>UBC</i>	#N/A	9.04	55.19	6.10
<i>MGC134066</i>	hypothetical protein MGC133647	5.33	32.18	6.03
<i>SERP2</i>	stress-associated endoplasmic reticulum protein 1	20.26	121.81	6.01
<i>BSP1</i>	seminal vesicle secretory protein 109	11.27	67.42	5.98
<i>USP50</i>	ubiquitin specific peptidase 44	18.49	109.18	5.90
<i>C8H9orf24</i>	#N/A	6.82	40.10	5.88
<i>MEA1</i>	#N/A	30.55	179.29	5.87
<i>GLUL</i>	glutamate-ammonia ligase (glutamine synthetase)	11.52	66.94	5.81
<i>RPS12</i>	ribosomal protein S11	42.29	243.05	5.75
<i>ARRDC4</i>	arrestin domain containing 4	20.37	117.05	5.75
<i>PCYT1A</i>	protein-L-isoaspartate (D-aspartate) O-methyltransferase	12.30	70.59	5.74
<i>AKAP4</i>	A kinase (PRKA) anchor protein 4	31.55	179.45	5.69
<i>TEX26</i>	testis expressed 101	11.31	64.16	5.67
<i>C10H15orf23</i>	chromosome 15 open reading frame 23 ortholog	50.81	284.82	5.61
<i>AGPAT3</i>	1-acylglycerol-3-phosphate O-acyltransferase 3	6.49	36.13	5.57
<i>ARGLU1</i>	arginine and glutamate rich 1	40.61	226.22	5.57
<i>ARGLU1</i>	arginine and glutamate rich 1	40.61	226.22	5.57
<i>CKS2</i>	CDC28 protein kinase regulatory subunit 2	86.19	477.80	5.54
<i>GLRX2</i>	glutaredoxin 2	11.83	65.56	5.54
<i>CDC5L</i>	CDC5 cell division cycle 5-like (S. pombe)	5.43	29.85	5.50
<i>RPS6</i>	ribosomal protein S4, Y-linked 1	92.63	506.01	5.46
<i>RPS6</i>	#N/A	92.63	506.01	5.46
<i>LDHC</i>	lysosomal protein transmembrane 4 alpha	62.50	341.40	5.46
<i>MS4A5</i>	MS4A13 protein	23.81	128.92	5.42
<i>ODF1</i>	OCIA domain containing 1	16.65	87.01	5.23
<i>NPC2</i>	#N/A	9.09	46.88	5.16
<i>C21H14orf14</i>	Uncharacterized protein C14orf142 homolog	7.87	40.57	5.16

<i>SHCBP1L</i>	SH3-domain GRB2-like endophilin B1	19.92	102.37	5.14
<i>PSMG1</i>	proteasome (prosome, macropain) inhibitor subunit 1 (PI31)	7.69	39.24	5.11
<i>CSDA</i>	cold shock domain protein A	22.72	115.67	5.09
<i>CCNB1</i>	similar to Cyclin B1; cyclin B1	6.27	31.80	5.08
<i>SRP72</i>	signal recognition particle 54kDa	17.49	88.69	5.07
<i>COX7A2</i>	cytochrome c oxidase subunit VIIa polypeptide 2 (liver)	46.39	235.08	5.07
<i>CAPZA1</i>	capping protein (actin filament) muscle Z-line, alpha 1	8.33	42.17	5.06
<i>CLGN</i>	calmegin	82.69	417.94	5.05
<i>C18H16orf87</i>	#N/A	17.27	87.05	5.04
<i>RPS11</i>	ribosomal protein S10; similar to 40S ribosomal protein S10	177.02	886.91	5.01
<i>BCL2L14</i>	Bcl2-like 14 (apoptosis facilitator)	27.05	135.40	5.01
<i>SNHG4</i>	#N/A	6.88	34.06	4.95
<i>RPL18</i>	#N/A	7.37	36.51	4.95
<i>KRTCAP2</i>	karyopherin alpha 4 (importin alpha 3)	6.51	32.24	4.95
<i>KIF27</i>	#N/A	9.88	48.75	4.94
<i>EEF1A1</i>	#N/A	59.09	290.46	4.92
<i>TFAM</i>	#N/A	18.14	88.90	4.90
<i>PAFAH1B1</i>	poly(A) binding protein, cytoplasmic 1	13.95	68.36	4.90
<i>TMBIM4</i>	#N/A	14.07	68.86	4.89
<i>HIST1H2BN</i>	#N/A	6.69	32.44	4.85
<i>PPP1R42</i>	protein phosphatase 1, regulatory (inhibitor) subunit 2	39.21	189.94	4.84
<i>ROPN1</i>	reactive oxygen species modulator 1	84.95	408.68	4.81
<i>SS18L2</i>	#N/A	6.65	31.90	4.80
<i>COX6B1</i>	#N/A	17.02	80.78	4.75
<i>EIF4G2</i>	eukaryotic translation initiation factor 4 gamma, 2	11.49	54.34	4.73
<i>MMADHC</i>	myeloid/lymphoid or mixed-lineage leukemia	15.80	74.69	4.73
<i>FXR1</i>	fragile X mental retardation, autosomal homolog 1	46.08	217.73	4.72
<i>RPL4</i>	ribosomal protein L39	13.38	62.92	4.70
<i>SRSF2</i>	#N/A	5.49	25.83	4.70
<i>MESDC2</i>	maternal embryonic leucine zipper kinase	5.53	25.99	4.70
<i>BLZF1</i>	basic leucine zipper nuclear factor 1	16.20	75.93	4.69
<i>VDAC2</i>	vasohibin 2	16.24	75.77	4.67
<i>RPL5</i>	similar to ribosomal protein L4; ribosomal protein L4	7.67	35.49	4.63
<i>CYLC1</i>	cylicin, basic protein of sperm head cytoskeleton 1	33.50	150.68	4.50
<i>FRG1</i>	#N/A	16.15	72.48	4.49
<i>FAM81B</i>	family with sequence similarity 81, member B	42.17	187.90	4.46
<i>PICALM</i>	phytanoyl-CoA 2-hydroxylase interacting protein-like	15.76	69.65	4.42
<i>RPS26</i>	#N/A	15.98	70.29	4.40
<i>FKBP2</i>	#N/A	5.40	23.41	4.33
<i>AMZ2</i>	archaelysin family metallopeptidase 2	11.71	50.73	4.33
<i>EID3</i>	EP300 interacting inhibitor of differentiation 3	6.21	26.89	4.33
<i>CHCHD7</i>	coiled-coil-helix-coiled-coil-helix domain containing 7	10.27	44.51	4.33

<i>CETN4</i>	centrin 4	77.47	329.77	4.26
<i>IGFBP7</i>	insulin-like growth factor binding protein 7	5.16	21.29	4.13
<i>STMN1</i>	serine/threonine kinase 17a	12.21	50.37	4.13
<i>MORF4L1</i>	meiosis-specific nuclear structural 1	52.56	216.69	4.12
<i>NDUFB4</i>	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3	81.78	336.24	4.11
<i>SBDS</i>	similar to Cytokine induced protein 29 kDa	11.24	46.09	4.10
<i>DYDC2</i>	DPY30 domain containing 2	19.33	78.96	4.08
<i>DAD1</i>	defender against cell death 1	99.92	406.94	4.07
<i>MFF</i>	methyltransferase like 13	44.85	181.10	4.04
<i>GML</i>	#N/A	5.78	23.25	4.02
<i>GGNBP2</i>	gametogenetin binding protein 2	19.12	76.33	3.99
<i>RPGRIP1</i>	RAR-related orphan receptor A	20.51	80.71	3.94
<i>COX7B2</i>	#N/A	58.19	228.08	3.92
<i>ADORA3</i>	adenosine A3 receptor	35.26	137.83	3.91
<i>C16H1orf100</i>	chromosome 1 open reading frame 100 ortholog	61.96	241.06	3.89
<i>YBX2</i>	Y box binding protein 1	34.08	132.03	3.87
<i>ADAM3A</i>	ADAM metalloproteinase domain 3A (cyritestin 1)	36.65	141.01	3.85
<i>CCT2</i>	chaperonin containing TCP1, subunit 2 (beta)	18.39	70.57	3.84
<i>NDUFB3</i>	#N/A	35.05	133.77	3.82
<i>RPS27A</i>	#N/A	223.62	852.29	3.81
<i>SKP1</i>	signal-regulatory protein delta	39.11	149.05	3.81
<i>C23H6orf106</i>	#N/A	13.58	51.64	3.80
<i>HMGB1</i>	high-mobility group box 1	12.31	46.65	3.79
<i>RPL17</i>	#N/A	18.74	70.74	3.78
<i>GPI</i>	glucose phosphate isomerase	13.89	52.13	3.75
<i>SNRPG</i>	small nuclear ribonucleoprotein polypeptide F	29.55	110.44	3.74
<i>NDUFB1</i>	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex	22.98	85.34	3.71
<i>RPL39</i>	ribosomal protein L38	19.44	72.19	3.71
<i>CAPZB</i>	capping protein (actin filament) muscle Z-line, beta	30.03	111.53	3.71
<i>UBXN6</i>	UBX domain protein 4	5.51	20.45	3.71
<i>NDUFS5</i>	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9	11.39	42.30	3.71
<i>C3H1orf189</i>	chromosome 1 open reading frame 189 ortholog	70.91	261.32	3.69
<i>BCAP29</i>	B-cell receptor-associated protein 29	62.95	231.40	3.68
<i>GN5</i>	#N/A	18.11	64.44	3.56
<i>FAM71E1</i>	#N/A	11.19	39.82	3.56
<i>ATP1B3</i>	ATPase, Na+/K+ transporting, beta 3 polypeptide	56.01	198.96	3.55
<i>RPL38</i>	ribosomal protein L37a	180.85	639.59	3.54
<i>SPATA9</i>	#N/A	29.98	105.73	3.53
<i>HSP90B1</i>	tumor rejection antigen (gp96) 1	19.68	69.27	3.52
<i>ROPN1L</i>	roporin, rhophilin associated protein 1	14.79	51.87	3.51
<i>NDUFA6</i>	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	65.71	229.21	3.49
<i>PSMB7</i>	proteasome (prosome, macropain) subunit, beta type, 3	16.95	59.01	3.48
<i>RNF38</i>	ring finger protein 216	19.27	66.93	3.47

<i>TOMM70A</i>	translocase of outer mitochondrial membrane 7 homolog (yeast)	18.52	64.30	3.47
<i>PSMG2</i>	proteasome (prosome, macropain) assembly chaperone 1	28.63	98.73	3.45
<i>SHFM1</i>	#N/A	41.97	144.73	3.45
<i>C15H11orf71</i>	chromosome 11 open reading frame 71 ortholog	17.83	61.47	3.45
<i>TMBIM6</i>	transmembrane BAX inhibitor motif containing 4	26.85	92.08	3.43
<i>PPWD1</i>	protein phosphatase 2, regulatory subunit B', gamma isoform	17.56	60.04	3.42
<i>EIF3K</i>	eukaryotic translation initiation factor 3, subunit K	5.59	19.01	3.40
<i>CSRP3</i>	cysteine and glycine-rich protein 3 (cardiac LIM protein)	9.07	30.86	3.40
<i>ATP5E</i>	#N/A	50.85	173.08	3.40
<i>TMSB10</i>	transmembrane protein 60	265.87	897.20	3.37
<i>USP15</i>	ubiquinol-cytochrome c reductase hinge protein	7.42	25.05	3.37
<i>ELK4</i>	ELK4, ETS-domain protein (SRF accessory protein 1)	5.58	18.65	3.34
<i>ETF1</i>	eukaryotic translation termination factor 1	6.14	20.41	3.33
<i>TAF9</i>	TAF10 RNA polymerase II	28.71	95.44	3.32
<i>DNAJB3</i>	#N/A	67.47	224.17	3.32
<i>C4H7orf10</i>	#N/A	14.02	46.57	3.32
<i>PFN1</i>	prefoldin subunit 5	7.76	25.63	3.30
<i>EIF5</i>	eukaryotic translation initiation factor 5	36.95	121.20	3.28
<i>MELK</i>	meiosis expressed gene 1 homolog (mouse)	7.79	25.42	3.26
<i>ROMO1</i>	RNA (guanine-7-) methyltransferase	15.80	51.35	3.25
<i>RPL10A</i>	#N/A	19.30	62.21	3.22
<i>STAT4</i>	signal transducing adaptor molecule (SH3 domain and ITAM motif) 2	22.73	73.23	3.22
<i>NBN</i>	N-acyl phosphatidylethanolamine phospholipase D	6.88	22.13	3.22
<i>YPEL2</i>	YME1-like 1 (<i>S. cerevisiae</i>)	11.53	37.11	3.22
<i>SLC30A3</i>	#N/A	8.94	28.78	3.22
<i>SCAMP1</i>	Shwachman-Bodian-Diamond syndrome	24.97	79.99	3.20
<i>NME2</i>	NIPA-like domain containing 3	13.73	43.97	3.20
<i>CALU</i>	calumenin	5.34	17.06	3.20
<i>ARL2BP</i>	ADP-ribosylation factor-like 2 binding protein	13.28	42.46	3.20
<i>CISD1</i>	CDGSH iron sulfur domain 1	354.71	1133.38	3.20
<i>AIF1</i>	allograft inflammatory factor 1	95.89	305.21	3.18
<i>DNAJB4</i>	DnaJ (Hsp40) homolog, subfamily B, member 4	48.84	154.38	3.16
<i>SF3B1</i>	stress-associated endoplasmic reticulum protein	7.22	22.74	3.15
<i>HSP90AA1</i>	#N/A	43.10	135.28	3.14
<i>PGP</i>	phosphoglycerate kinase 2	11.81	36.94	3.13
<i>LAPTM4A</i>	#N/A	55.64	173.85	3.12
<i>CYLC2</i>	cylicin, basic protein of sperm head cytoskeleton 2	138.04	427.82	3.10
<i>C1D</i>	C1D nuclear receptor co-repressor	16.78	51.94	3.09
<i>PFN2</i>	#N/A	6.08	18.82	3.09
<i>CIAPIN1</i>	cytokine induced apoptosis inhibitor 1	10.35	32.02	3.09
<i>LOC100125913</i>	leukemia NUP98 fusion partner 1	9.85	30.50	3.09
<i>PDHB</i>	pyruvate dehydrogenase (lipoamide) alpha 2	6.73	20.82	3.09

<i>ACOT7</i>	acyl-CoA thioesterase 7	5.15	15.94	3.09
<i>NDUFB2</i>	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa	10.39	32.14	3.09
<i>CABYR</i>	calcium binding tyrosine-(Y)-phosphorylation regulated	21.15	65.21	3.08
<i>FAM92A1</i>	#N/A	27.53	83.42	3.03
<i>C5H12orf50</i>	chromosome 12 open reading frame 50 ortholog	28.76	86.78	3.02
<i>SPESP1</i>	spermatid maturation 1	11.12	33.56	3.02
<i>CUL3</i>	cullin 3	15.33	45.85	2.99
<i>IQCF1</i>	integrator complex subunit 12	36.67	109.69	2.99
<i>TMEM258</i>	transmembrane and coiled-coil domains 5B	101.28	301.68	2.98
<i>SERF1A</i>	SERPINE1 mRNA binding protein 1	162.97	485.41	2.98
<i>ARL4A</i>	ADP-ribosylation factor-like 4A	75.36	224.41	2.98
<i>COIL</i>	coilin	9.14	27.08	2.96
<i>BRD2</i>	bromodomain containing 2	6.47	19.13	2.96
<i>RABEPK</i>	RAB5A, member RAS oncogene family	28.56	84.21	2.95
<i>HSPA5</i>	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	20.32	59.86	2.95
<i>VTI1B</i>	very low density lipoprotein receptor	34.70	101.74	2.93
<i>FAM229B</i>	#N/A	965.33	2823.04	2.92
<i>PFDN5</i>	phosphatidylethanolamine-binding protein 4	108.54	317.07	2.92
<i>LYZL1</i>	Lym7 homolog (mouse)	33.52	97.51	2.91
<i>TMA16</i>	#N/A	19.54	56.43	2.89
<i>HSP90AB1</i>	heat shock 90kDa protein 1, beta	9.29	26.62	2.87
<i>MEIG1</i>	mediator complex subunit 6	117.34	334.29	2.85
<i>NDUFA4</i>	#N/A	93.47	265.58	2.84
<i>SPINK2</i>	sperm equatorial segment protein 1	192.07	545.75	2.84
<i>UBB</i>	#N/A	384.20	1088.56	2.83
<i>TOX4</i>	translocase of outer mitochondrial membrane 70 homolog A	9.26	26.20	2.83
<i>GSG1</i>	germ cell associated 1	24.89	70.19	2.82
<i>LELP1</i>	lactate dehydrogenase C	36.39	102.60	2.82
<i>CDV3</i>	CDV3 homolog (mouse)	27.46	77.42	2.82
<i>TERF2</i>	tektin 3	8.88	24.92	2.81
<i>RPS21</i>	#N/A	62.43	173.88	2.79
<i>H3F3A</i>	#N/A	16.81	46.23	2.75
<i>BANF2</i>	barrier to autointegration factor 2	251.77	692.54	2.75
<i>AKTIP</i>	AKT interacting protein	14.65	40.21	2.75
<i>WBP2NL</i>	Wiskott-Aldrich syndrome-like	18.68	51.20	2.74
<i>BAZ2B</i>	bromodomain adjacent to zinc finger domain, 2B	14.22	38.93	2.74
<i>DCUN1D4</i>	DCN1, defective in cullin neddylation 1, domain containing 4	61.63	166.89	2.71
<i>MS4A13</i>	mitochondrial ribosomal protein S9	151.90	411.21	2.71
<i>CCDC54</i>	coiled-coil domain containing 54	69.36	186.97	2.70
<i>SRP54</i>	signal recognition particle 19kDa	92.91	247.01	2.66
<i>DNAJA1</i>	DnaJ (Hsp40) homolog, subfamily A, member 1	15.92	42.16	2.65
<i>TPI1</i>	TOX high mobility group box family member 4	12.05	31.69	2.63

<i>RPS15</i>	#N/A	22.07	58.05	2.63
<i>KIF2B</i>	kinesin heavy chain member 2A	12.48	32.83	2.63
<i>PTGR1</i>	proteasome (prosome, macropain) assembly chaperone 2	5.99	15.76	2.63
<i>SAA4</i>	S100 calcium binding protein G	67.67	175.29	2.59
<i>ENO1</i>	enolase 1, (alpha)	6.00	15.47	2.58
<i>WASL</i>	vesicle transport through interaction with t-SNAREs	5.82	15.01	2.58
<i>FAM103A1</i>	#N/A	20.22	52.15	2.58
<i>PRPS1L1</i>	protein interacting with cyclin A1	6.03	15.45	2.56
<i>FAU</i>	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV)	100.22	256.54	2.56
<i>C15H11orf88</i>	#N/A	28.37	72.44	2.55
<i>AP2B1</i>	adaptor-related protein complex 2, beta 1 subunit	94.45	240.14	2.54
<i>TRIM36</i>	trafficking protein particle complex 3	20.97	53.24	2.54
<i>CNOT2</i>	CCR4-NOT transcription complex, subunit 2	8.19	20.73	2.53
<i>CLU</i>	clusterin	12.20	30.88	2.53
<i>MRPS18C</i>	#N/A	77.78	195.99	2.52
<i>RPL37A</i>	#N/A	375.70	945.04	2.52
<i>ZNF451</i>	zinc finger protein 432	9.83	24.70	2.51
<i>YWHAQ</i>	tyrosine 3-monooxygenase gamma polypeptide	82.64	206.93	2.50
<i>UQCRC2</i>	ubiquinol-cytochrome c reductase binding protein	7.02	17.52	2.50
<i>RANGRF</i>	#N/A	95.84	238.84	2.49
<i>UBE2V1</i>	ubiquitin-conjugating enzyme E2R 2	17.63	43.66	2.48
<i>C19H17orf98</i>	#N/A	6.02	14.90	2.48
<i>ARF4</i>	#N/A	21.19	52.45	2.48
<i>CAP1</i>	CAP, adenylate cyclase-associated protein 1	7.49	18.53	2.48
<i>LOC504599</i>	histone H2B variant PT15	7.46	18.47	2.48
<i>RPS27</i>	#N/A	120.56	298.47	2.48
<i>CCDC167</i>	#N/A	8.88	21.98	2.48
<i>C11H2orf43</i>	#N/A	8.07	19.97	2.48
<i>C15H11orf1</i>	#N/A	39.83	98.29	2.47
<i>FSCN3</i>	fascin homolog 3, actin-bundling protein, testicular	28.32	69.49	2.45
<i>FAM209B</i>	#N/A	328.88	805.05	2.45
<i>C21H14orf2</i>	#N/A	339.04	826.64	2.44
<i>ATP5L</i>	#N/A	92.45	225.29	2.44
<i>GALNTL5</i>	UDP-N-acetyl-alpha-D-galactosamine	42.73	103.89	2.43
<i>EIF1</i>	#N/A	87.71	212.72	2.43
<i>TEX101</i>	testis derived transcript (3 LIM domains)	23.63	57.28	2.42
<i>PSME4</i>	#N/A	5.55	13.46	2.42
<i>DNAJC21</i>	DnaJ (Hsp40) homolog, subfamily C, member 21	19.41	47.01	2.42
<i>C4H7orf62</i>	#N/A	95.34	230.41	2.42
<i>PSMA6</i>	proteasome (prosome, macropain) subunit, alpha type, 4	537.25	1296.79	2.41
<i>MGC134093</i>	hypothetical protein MGC134066	128.66	310.48	2.41
<i>PIGF</i>	protein interacting with PRKCA 1	58.77	141.32	2.40
<i>RAB5A</i>	RAB28, member RAS oncogene family	7.20	17.19	2.39

<i>KIF5C</i>	kinesin family member 5B	43.01	102.51	2.38
<i>TMEM30A</i>	#N/A	21.60	51.48	2.38
<i>TEX35</i>	#N/A	111.07	262.35	2.36
<i>HNRNPC</i>	heterogeneous nuclear ribonucleoprotein C (C1/C2)	11.00	25.95	2.36
<i>ARPM1</i>	actin related protein M1	21.94	51.76	2.36
<i>ERICH2</i>	#N/A	364.32	858.46	2.36
<i>HNRNPU</i>	heterogeneous nuclear ribonucleoprotein U	8.51	20.06	2.36
<i>FNDC3A</i>	fibronectin type III domain containing 3A	11.09	26.11	2.35
<i>TSACC</i>	hypothetical LOC540154	791.86	1857.74	2.35
<i>PRSS37</i>	protease, serine, 23	79.01	184.74	2.34
<i>CCDC101</i>	coiled-coil domain containing 101	24.12	56.39	2.34
<i>RPS20</i>	#N/A	82.82	193.11	2.33
<i>DBI</i>	diazepam binding inhibitor	245.96	572.66	2.33
<i>DKKL1</i>	dickkopf-like 1 (soggy)	8.27	19.18	2.32
<i>RPS7</i>	#N/A	39.81	92.39	2.32
<i>LY6G6C</i>	leucine zipper protein 1	18.64	43.26	2.32
<i>CALM2</i>	#N/A	233.28	539.14	2.31
<i>BPIFA3</i>	#N/A	119.02	272.68	2.29
<i>CDKL3</i>	cyclin-dependent kinase-like 3	10.09	23.10	2.29
<i>FBP1</i>	#N/A	49.85	114.05	2.29
<i>KLHL2</i>	kelch-like 10 (Drosophila)	8.41	19.23	2.29
<i>ALB</i>	albumin	6.09	13.91	2.28
<i>TMC02</i>	transmembrane BAX inhibitor motif containing 6	152.54	348.14	2.28
<i>RPSA</i>	ribosomal protein S8	11.98	27.18	2.27
<i>MED6</i>	mediator complex subunit 31	11.43	25.94	2.27
<i>HNRNPD</i>	heterogeneous nuclear ribonucleoprotein D	7.24	16.44	2.27
<i>RPL13</i>	#N/A	19.73	44.76	2.27
<i>RPL32</i>	similar to ribosomal protein L31; ribosomal protein L31	283.92	641.10	2.26
<i>RPS3A</i>	ribosomal protein S3	224.15	506.05	2.26
<i>ATP6V1C1</i>	ATPase, H ⁺ transporting, lysosomal 42kDa, V1 subunit C1	33.73	75.89	2.25
<i>RPL22L1</i>	#N/A	14.43	32.27	2.24
<i>UBXN1</i>	#N/A	9.07	20.21	2.23
<i>KIF5B</i>	kinesin family member 3A	5.39	11.81	2.19
<i>RPL35A</i>	ribosomal protein L35	238.98	519.68	2.17
<i>SPATA6</i>	spermatogenesis associated 4	314.58	681.94	2.17
<i>RBM4B</i>	RNA binding motif protein 46	30.40	65.85	2.17
<i>C9H6orf165</i>	chromosome 6 open reading frame 165 ortholog	6.14	13.30	2.17
<i>PLCZ1</i>	phospholipase C, delta 4	27.33	58.96	2.16
<i>C1H3orf38</i>	chromosome 3 open reading frame 38 ortholog	88.13	190.05	2.16
<i>KLHDC10</i>	kinesin family member 5C	19.54	42.01	2.15
<i>COPS4</i>	COP9 constitutive photomorphogenic homolog subunit 4	32.03	68.68	2.14
<i>GTF2H5</i>	general transcription factor IIH, polypeptide 5	50.85	108.57	2.14
<i>LPIN1</i>	hypothetical protein LOC788205	28.98	61.85	2.13

<i>UQCRB</i>	ubiquitin interaction motif containing 1	245.77	523.54	2.13
<i>CALM3</i>	#N/A	28.81	61.37	2.13
<i>TXN</i>	tumor suppressor candidate 3	36.83	78.15	2.12
<i>ATOX1</i>	ATX1 antioxidant protein 1 homolog (yeast)	92.30	195.86	2.12
<i>SAMD4A</i>	#N/A	6.02	12.73	2.11
<i>BEX4</i>	hypothetical protein LOC783451	16.67	35.25	2.11
<i>RPS25</i>	#N/A	184.22	388.38	2.11
<i>RPL36AL</i>	#N/A	190.90	401.36	2.10
<i>RPL23</i>	ribosomal protein L22-like 1	445.87	936.58	2.10
<i>SMCP</i>	SWI/SNF matrix associated, actin dep regulator of chromatin	458.59	961.76	2.10
<i>C16H1orf114</i>	#N/A	145.46	304.18	2.09
<i>DCTN6</i>	dynactin 6	41.67	86.91	2.09
<i>BAG4</i>	BCL2-associated athanogene 4	15.79	32.92	2.08
<i>TPT1</i>	#N/A	38.79	80.79	2.08
<i>CRISP2</i>	#N/A	661.17	1366.46	2.07
<i>SUGP1</i>	STT3, subunit of the oligosaccharyltransferase complex	15.21	31.39	2.06
<i>MGC134473</i>	hypothetical protein MGC134427	6.87	14.17	2.06
<i>FAM71D</i>	family with sequence similarity 71, member D ortholog	210.75	434.67	2.06
<i>KCMF1</i>	#N/A	49.08	101.01	2.06
<i>YBX1</i>	YY1 associated factor 2	104.98	215.97	2.06
<i>CCDC91</i>	coiled-coil domain containing 91	29.89	61.39	2.05
<i>GSTM3</i>	glutathione S-transferase mu 3 (brain)	1991.48	4088.10	2.05
<i>GDI2</i>	GDP dissociation inhibitor 2	9.53	19.52	2.05
<i>TES</i>	telomeric repeat binding factor 2	93.05	189.96	2.04
<i>BZW1</i>	basic leucine zipper and W2 domains 1	23.34	47.63	2.04
<i>MYCBP</i>	mitochondrial fission regulator 1	36.11	73.44	2.03
<i>RPL10</i>	retinitis pigmentosa GTPase regulator interacting protein 1	21.64	44.01	2.03
<i>BBX</i>	bobby sox homolog (Drosophila)	13.45	27.26	2.03
<i>AMD1</i>	#N/A	26.34	53.39	2.03
<i>MGC148714</i>	#N/A	54.96	111.33	2.03
<i>TSPAN6</i>	tetraspanin 5	105.27	213.03	2.02
<i>RPL21</i>	ribosomal protein L19	35.03	70.84	2.02
<i>EEF1D</i>	eukaryotic translation elongation factor 1 delta	24.78	50.09	2.02
<i>TXNL1</i>	thioredoxin domain containing 8 (spermatozoa)	50.57	101.98	2.02
<i>TUBB3</i>	#N/A	5.40	10.89	2.02
<i>CFL2</i>	cofilin 2 (muscle)	5.42	10.90	2.01
<i>TAOK1</i>	#N/A	6.21	12.49	2.01
<i>HIST2H2BE</i>	#N/A	5.99	12.04	2.01
<i>LOC513111</i>	#N/A	12.02	24.17	2.01
<i>DCAF12</i>	WD repeat domain 40A	13.54	27.16	2.01

844
845
846
847
848
849
850
851
852
853

Appendix 2. Spermatozoal transcripts with an abundance of > 2:1 in higher:lower fertility populations based on FPKM

Official Gene Symbol	Gene Long Name	Higher FPKM	Lower FPKM	Higher: Lower Ratio
<i>UCHL1</i>	UBX domain protein 6	142.15	5.23679	27.14
<i>ENY2</i>	enhancer of yellow 2 homolog (Drosophila)	113.50	5.66497	20.04
<i>PLAC1</i>	#N/A	119.24	8.02166	14.86
<i>CETN2</i>	centrin, EF-hand protein, 2	62.39	5.67822	10.99
<i>BCL2L11</i>	BCL2-like 11 (apoptosis facilitator)	141.55	15.5432	9.11
<i>EMC2</i>	#N/A	87.94	10.5031	8.37
<i>LSM12</i>	limbic system-associated membrane protein	43.79	5.5314	7.92
<i>NAPEPLD</i>	#N/A	40.84	5.20358	7.85
<i>CHRAC1</i>	chromatin accessibility complex 1 ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	72.12	9.76486	7.39
<i>UBE2R2</i>		143.09	21.255	6.73
<i>COPS3</i>	COP9 constitutive photomorphogenic homolog subunit 3	58.32	8.87618	6.57
<i>SH3BGR13</i>	#N/A	64.91	10.4808	6.19
<i>ASB12</i>	ankyrin repeat and SOCS box-containing 12	56.81	9.58963	5.92
<i>ANP32B</i>	acidic (leucine-rich) nuclear phosphoprotein 32 family, B	42.45	7.50689	5.66
<i>AHCY</i>	adenosylhomocysteinase	33.02	6.04565	5.46
<i>MANF</i>	maelstrom homolog (Drosophila)	69.42	12.7308	5.45
<i>GLYATL3</i>	#N/A	60.05	11.4356	5.25
<i>KPNA4</i>	kelch-like 24 (Drosophila)	59.37	11.5279	5.15
<i>C11H9orf9</i>	chromosome 9 open reading frame 9 ortholog	106.29	21.0511	5.05
<i>CDKL4</i>	cyclin-dependent kinase-like 4	93.75	18.686	5.02
<i>SPAG4</i>	sperm associated antigen 17	46.37	9.30735	4.98
<i>PTMA</i>	prostaglandin reductase 1	34.13	7.04153	4.85
<i>RPS28</i>	ribosomal protein S27-like	35.99	7.42415	4.85
<i>MIR1307</i>	missing oocyte, meiosis regulator, homolog (Drosophila)	420.45	86.7418	4.85
<i>MRPS16</i>	mitochondrial ribosomal protein L42	29.61	6.10836	4.85
<i>SNRPF</i>	small nuclear ribonucleoprotein D2 polypeptide 16.5kDa	24.60	5.37417	4.58
<i>LSM2</i>	LSM12 homolog (S. cerevisiae)	277.89	60.9775	4.56
<i>FAM76B</i>	#N/A	48.69	10.7459	4.53
<i>PRSS23</i>	proline rich 13	24.72	5.56424	4.44
<i>CEP76</i>	centrosomal protein 76kDa	25.96	5.91952	4.39
<i>AMAC1L3</i>	acyl-malonyl condensing enzyme 1-like 3	27.76	6.442	4.31
<i>TRAPPC3</i>	transformer 2 beta homolog (Drosophila)	74.24	17.5969	4.22
<i>SYNE4</i>	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)	21.54	5.12758	4.20
<i>UBE2J1</i>	ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast)	101.78	24.4097	4.17
<i>FBXL2</i>	F-box and leucine-rich repeat protein 2	33.65	8.12796	4.14
<i>PNN</i>	#N/A	23.51	5.69404	4.13
<i>CCNB2</i>	similar to cyclin B2; cyclin B2	34.12	8.2962	4.11
<i>LYRM4</i>	#N/A	52.41	12.9745	4.04

<i>C23H6orf25</i>	#N/A	39.98	9.99186	4.00
<i>SEC61A2</i>	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	70.32	17.871	3.94
<i>YAE1D1</i>	#N/A	69.71	17.8064	3.92
<i>PMPCB</i>	non-protein coding RNA 153	34.91	9.06111	3.85
<i>WDR53</i>	WBP2 N-terminal like	64.89	17.7455	3.66
<i>GALNT3</i>	UDP-N-acetyl-alpha-D-galactosamine	39.51	10.8672	3.64
<i>EIF3G</i>	eukaryotic translation initiation factor 3, subunit G	19.08	5.36828	3.55
<i>ASB8</i>	ankyrin repeat and SOCS box-containing 8	50.71	14.4483	3.51
<i>HIST1H2AG</i>	#N/A	23.82	6.80562	3.50
<i>RPS3</i>	#N/A	236.85	68.2526	3.47
<i>SSRP1</i>	#N/A	65.61	19.085	3.44
<i>BTBD10</i>	BTB (POZ) domain containing 10	44.81	13.1859	3.40
<i>DYNLRB2</i>	dynein, light chain, roadblock-type 2	260.34	77.7852	3.35
<i>IQCF2</i>	IQ motif containing F1	120.31	35.9479	3.35
<i>GOSR1</i>	golgi SNAP receptor complex member 1	61.50	18.5172	3.32
<i>REEP6</i>	RNA binding motif protein 4B	28.81	8.79414	3.28
<i>SDHA</i>	#N/A	17.34	5.36623	3.23
<i>C11H9orf16</i>	chromosome 9 open reading frame 16 ortholog	38.57	11.9349	3.23
<i>SRP19</i>	spermatogenic leucine zipper 1	24.08	7.45045	3.23
<i>ADK</i>	#N/A	23.34	7.22263	3.23
<i>SCP2D1</i>	#N/A	743.18	232.871	3.19
<i>GTSF1L</i>	gametocyte specific factor 1-like	862.80	276.158	3.12
<i>RFWD2</i>	receptor accessory protein 6	35.98	11.5462	3.12
<i>CDC27</i>	cell division cycle 27 homolog (S. cerevisiae)	20.91	6.86248	3.05
<i>WFDC15B</i>	#N/A	33.28	10.9419	3.04
<i>CCDC67</i>	coiled-coil domain containing 67	52.53	17.417	3.02
<i>AGPAT9</i>	1-acylglycerol-3-phosphate O-acyltransferase 9	19.66	6.5189	3.02
<i>SERP1</i>	small EDRK-rich factor 2	56.38	19.0346	2.96
<i>LOC520057</i>	#N/A	114.95	39.0595	2.94
<i>ZDHHC20</i>	zinc finger, CCHC domain containing 6	14.93	5.1975	2.87
<i>SERBP1</i>	SUMO1/sentrin specific peptidase 6	16.13	5.61632	2.87
<i>TMEM41B</i>	transmembrane protein 30A	16.67	5.80419	2.87
<i>ILF2</i>	interleukin enhancer binding factor 2, 45kDa	207.66	72.6433	2.86
<i>GDE1</i>	glycerophosphodiester phosphodiesterase 1	33.40	11.7081	2.85
<i>CYB5R1</i>	cytochrome b5 reductase 1	21.55	7.55951	2.85
<i>DNAJC7</i>	DnaJ (Hsp40) homolog, subfamily C, member 7	19.15	6.77096	2.83
<i>RHOBTB1</i>	ring finger and WD repeat domain 2	18.93	6.69607	2.83
<i>CFL1</i>	cofilin 1 (non-muscle)	60.20	21.4534	2.81
<i>MCFD2</i>	MARCKS-like 1	17.65	6.32279	2.79
<i>WDR96</i>	WD repeat domain 74	14.69	5.33539	2.75
<i>TAF10</i>	synaptophysin-like 1	354.99	129.035	2.75
<i>FAM76A</i>	family with sequence similarity 76, member A	42.42	15.6108	2.72
<i>BTG1</i>	B-cell translocation gene 1, anti-proliferative	30.42	11.2972	2.69
<i>GLOD4</i>	glyoxalase domain containing 4	64.59	24.1218	2.68
<i>AFF4</i>	AF4/FMR2 family, member 4	69.77	26.1796	2.67
<i>FAM98A</i>	#N/A	22.78	8.61676	2.64

<i>LIN7C</i>	lipoic acid synthetase	23.65	8.94574	2.64
<i>CCDC135</i>	coiled-coil domain containing 135	18.51	7.01963	2.64
<i>EIF5A</i>	eukaryotic translation initiation factor 5A	79.68	30.2392	2.63
<i>ATP5I</i>	ATP synthase, H ⁺ transporting, mitochondrial F0 complex	200.95	76.5374	2.63
<i>USP44</i>	ubiquitin specific peptidase 16	45.64	17.547	2.60
<i>UAP1</i>	thioredoxin-like 1	36.26	13.9643	2.60
<i>HNRNP1</i>	#N/A	22.54	8.71883	2.59
<i>CCDC70</i>	coiled-coil domain containing 70	25.22	9.93395	2.54
<i>ANKRD32</i>	ankyrin repeat domain 32	23.33	9.34499	2.50
<i>ACTL7A</i>	actin-like 7A	44.54	17.8814	2.49
<i>IQCG</i>	IQ motif containing F5	115.05	46.2528	2.49
<i>SSR1</i>	#N/A	23.27	9.37931	2.48
<i>TEKT3</i>	tektin 1	25.22	10.1809	2.48
<i>RNMT</i>	similar to ring finger protein 38; ring finger protein 38	28.31	11.478	2.47
<i>MCL1</i>	#N/A	19.43	8.01607	2.42
<i>NXN</i>	#N/A	16.39	6.76194	2.42
<i>SPATA6L</i>	spermatogenesis associated 6	89.91	37.4999	2.40
<i>LSAMP</i>	#N/A	14.72	6.14617	2.40
<i>VIMP</i>	vimentin	101.90	42.9584	2.37
<i>VGLL3</i>	voltage-dependent anion channel 2	94.26	39.7495	2.37
<i>POLB</i>	pinin, desmosome associated protein	63.50	26.7976	2.37
<i>FZD7</i>	frizzled homolog 7 (Drosophila)	13.86	5.88325	2.36
<i>FN3KRP</i>	fructosamine 3 kinase related protein	40.66	17.254	2.36
<i>APOPT1</i>	#N/A	291.65	124.488	2.34
<i>ACAT1</i>	acetyl-Coenzyme A acetyltransferase 1	99.09	42.3472	2.34
<i>LOC788205</i>	#N/A	15.40	6.60052	2.33
<i>BAZ1A</i>	bromodomain adjacent to zinc finger domain, 1A	17.47	7.5322	2.32
<i>RN45S</i>	RNA component of mitochondrial RNA processing	258.48	113.249	2.28
<i>TM9SF2</i>	T-cell immunoglobulin and mucin domain containing 4	31.74	14.211	2.23
<i>RAD21</i>	RAB interacting factor	41.18	18.4569	2.23
<i>ZNF706</i>	hypothetical protein LOC100125413	180.19	81.1021	2.22
<i>PSMD1</i>	proteasome (prosome, macropain) subunit, beta type, 7	18.26	8.30929	2.20
<i>CNBP</i>	CCHC-type zinc finger, nucleic acid binding protein	30.43	13.921	2.19
<i>CHMP5</i>	chromatin modifying protein 5	4359.8	1995.9	2.18
<i>GSTO2</i>	glutathione S-transferase omega 2	121.97	56.3501	2.16
<i>CSN3</i>	casein kappa	19.71	9.14845	2.15
<i>MND1</i>	methylmalonic aciduria (cobalamin deficiency) cblD type	58.49	27.1507	2.15
<i>RORA</i>	ropporin 1-like	53.29	24.7343	2.15
<i>CSPP1</i>	centrosome and spindle pole associated protein 1	12.16	5.64413	2.15
<i>ZNF33B</i>	zinc finger protein 24	23.93	11.3243	2.11
<i>HDGFRP3</i>	#N/A	54.71	26.1645	2.09
<i>FGA</i>	fibrinogen alpha chain	14.62	7.03925	2.08
<i>BCKDHB</i>	branched chain keto acid dehydrogenase E1, beta	20.81	10.1199	2.06
<i>SHOC2</i>	#N/A	23.88	11.6139	2.06
<i>MLL5</i>	myeloid leukemia factor 1	10.48	5.131	2.04
<i>YWHAE</i>	yippee-like 2 (Drosophila)	24.40	11.9553	2.04

<i>RPL23A</i>	ribosomal protein L23	34.30	16.8916	2.03
<i>PPM1G</i>	#N/A	40.24	19.8221	2.03
<i>APOO</i>	apolipoprotein O	59.03	29.2261	2.02
<i>TUBB</i>	testis-specific serine kinase 3	10.77	5.33413	2.02
<i>ZNF428</i>	zinc finger protein 33B	13.24	6.55637	2.02
<i>STRBP</i>	#N/A	21.34	10.5645	2.02
<i>MAP9</i>	mitogen-activated protein kinase kinase kinase kinase 3	40.52	20.1516	2.01

859

Card et al., Figure 1

Card Figure 2