2014

# THE BOVINE SPERMATOZOAL TRANSCRIPTOME AND SIRE FERTILITY

Christopher James Card
*University of Rhode Island*, ccard28@gmail.com

# THE BOVINE SPERMATOZOAL TRANSCRIPTOME AND SIRE FERTILITY

**BY**

**CHRISTOPHER JAMES CARD**

**A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOLOGICAL AND ENVIRONMENTAL SCIENCES**

**UNIVERSITY OF RHODE ISLAND**

**2014**

DOCTOR OF PHILOSOPHY DISSERTATION

OF

CHRISTOPHER J CARD

APPROVED:

Dissertation Committee:

Major Professor Becky L. Sartini

Katherine Petersson

Steven Irvine

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2014

**ABSTRACT**

The goal of this dissertation research was to characterize the spermatozoal transcript profile of bovine cryopreserved spermatozoa. The main goal was to sequence the complete transcript profile from a population of sires across a wide fertility range utilizing RNA-Sequencing (RNA-Seq). Sequencing was then conducted to compare the spermatozoal transcript profiles of higher and lower fertility sires in an effort to determine the presence or absence of spermatozoal transcripts that may aid in improving fertility testing in sires. In addition to RNA-Seq analysis, other semen parameters such as morphology, DNA fragmentation, and RNA amount were analyzed for relationships with fertility.

The first chapter of this dissertation is a literature review detailing the artificial insemination industry, male fertility, spermatogenesis, spermatozoal RNAs, and RNA-Seq high-throughput sequencing. The review details what is currently known about the spermatozoal RNA population as well as the potential application of RNA-Seq to advance the knowledge of this spermatozoal RNA population.

The second chapter is a manuscript that was published in the journal Biology of Reproduction in January 2013 and was co-first-authored by Elizabeth Anderson. This manuscript was the first study to sequence the spermatozoal transcript profile using RNA-Seq for any species. RNA-Seq analysis of pooled semen from multiple sires sequenced a bovine spermatozoal transcriptome consisting of 6,166 transcripts, including several previously identified and novel candidate transcripts for further functional study.

The third chapter of this dissertation is a manuscript being prepared for submission to the Journal of Dairy Science. In this manuscript, the spermatozoal transcript profiles of higher and lower fertile sires was sequenced. Target fertility candidate transcripts were selected on the basis of expression differences between the two populations and then correlation of these spermatozoal transcripts with sire fertility was examined.

Appendix 1 consists of unpublished data examining the relationship of sperm RNA amount with semen parameters such as morphology, DNA fragmentation index, motility, and sperm RNA quantity. Sperm morphology analysis was performed by the Parrish lab while DNA fragmentation index analysis was performed by the Evenson lab. This sperm RNA isolation method is the same procedure as described in Appendix One and therefore I was unable to validate this transcript population.

Appendix 2 reports an unpublished initial RNA-Seq study done with spermatozoal RNA isolated with a column based method and mRNA amplification that differs from these protocols reported in Chapter 2. I was unable to validate the spermatozoal transcript profile generated with these RNA isolation and amplification methods leading to the development of a second RNA isolation procedure that is reported in Chapter 2.

Through this work it is evident that there was much we did not know about the bovine spermatozoal transcript profile but the use of RNA-Seq in this study has helped us understand the population more. This new technology has allowed us to sequence the entire transcript profile while determining what

transcripts are full-length. Not only were we able to sequence a general population but we compared lower and higher fertility populations and found many differences between the two. Utilizing all of our data we were able to identify four individual transcripts that have correlations with sire fertility that may prove useful as alternatives for *in vitro* fertility tests.

**ACKNOWLEDGMENTS**

I would like to thank my principal investigator, Dr. Becky Sartini for her constant support throughout my tenure at URI. I would also like to thank fellow students in the Sartini lab for their continued support and friendships forged over the years, including: Sabrina Green, Erin Noblet, Rachel Gately, Mike Eerhart, Jazmin Zamberlan, Erin Aparicio, Caitlin Kozel, Krystle Schultz, Evan Hoffman, Holly Tran, and Justin Richard. Special thanks goes out to Elizabeth Anderson, as this work, especially the BOR publication, would not have been possible without her help of so I would like to especially thank her. A special thanks is owed to chair and committee members Dr. Katherine Petersson, Dr. Marta Gomez-Chiarri, Dr. Steven Irvine, Dr. Joel Chandlee, and Dr. Bovy for taking the time to aid me in developing and completing my dissertation over the years.

I have had the opportunity to serve as teaching assistant for many professors, and I would like to extend a thank you to Fred Launer and Dr. Darlene Jones for the invaluable experiences they made possible mentoring me as their teaching assistant.

Lastly I would like to thank my parents for the continued financial and emotional support throughout my many years as a graduate student at URI. Without them none of this would have been possible.

**PREFACE**

This dissertation is in manuscript format. Chapters Two and Three of this dissertation are two manuscripts for publication in peer-reviewed journals. Chapter Two titled, "Cryopreserved Bovine Spermatozoal Transcript Profile as Revealed by RNA-Seq" was published in Biology of Reproduction in January 2013 with Elizabeth Anderson and myself as co-first authors. My responsibilities in that manuscript include Figures 1, 2, and 4, along with Tables 1, 2, and 4. Liz was responsible for Figure 3 as well as Tables 3, 5, and 6. Editing responsibilities were shared equally.

Chapter Three titled, "Spermatozoal Transcript Profiles differ between Lower and Higher Fertility Sires; " is a second manuscript that is currently being edited for submission to the Journal of Dairy Science and is solely my work.

Appendix One involves data that I collected (RNA amount; motility) as well as data from two other labs, the Parrish (sperm morphology) and Evenson labs (DNA fragmentation).

## TABLE OF CONTENTS

## LIST OF TABLES

### CHAPTER 1:



### CHAPTER 2:



### CHAPTER 3:

**APPENDIX 1:**



**APPENDIX 2:**

## LIST OF FIGURES

### CHAPTER 1:



### CHAPTER 2:



### CHAPTER 3

**APPENDIX 2**

# CHAPTER 1

## LITERATURE REVIEW

### I. Livestock fertility

*A. Artificial insemination industry*

Worldwide, artificial insemination (AI) has become standard practice for breeding in the dairy industry. In 1944, there were approximately 25.6 million dairy cows that produced 53.1 billion kg of milk in contrast to 1997 when there were an estimated 9.2 million dairy cows that produce an estimated 70.8 billion kg of milk (Dejarnette et al, 2004). This increase in milk production from fewer cows is due to the widespread use of AI to enhance genetic selection of desirable production traits, including increased milk production. In addition to advances in cow nutrition and management practices, AI sires have been chosen for daughter milk production traits such as high milk yield, fat/protein yield, mastitis, non-return rate, and lameness, which have lead to significant overall increases in milk production (Cottle et al, 2013). As producers seek to continue selecting sires that produce highly productive daughters, the AI industry has continued to expand annually. Global semen sales have nearly doubled from 19.5 million units in 1994 to 38 million units in 2012 as the demand for milk continues to grow (National Association of Animal Breeders, NAAB). About 28 million units of semen are produced each year, with the exportation of approximately 9 million straws, totaling over $131 million in sales (NAAB).

While AI has allowed for marked increases in milk production, sire selection has also improved other parameters related to animal health and production. The selection of conformation traits beneficial to dairy herd production and health include sturdy feet and

leg conformation, udder development and overall body condition scores (Cottle et al, 2013). Other benefits of AI include a decreased incidence of venereal disease by eliminating animals or semen with diseases such as infectious bovine rhinotracheitis (IBR) and bovine viral diarrhea virus (BVDV; Eaglesome and Garcia, 1997). The export of semen straws has also decreased the incidence of disease from shipping potentially sick animals into new herds. Many dairy farms strictly use AI for their breeding programs, eliminating the need for farms to house their own bull, resulting in a safer environment for workers as well as reduction of production costs.

Despite these advances, currently only an average of 50% of inseminations (range 30-75%) result in a full-term pregnancy, as cow fertility rates have been steadily declining over the past 50 years (Peddinti et al, 2008; Flowers, 2013). Sires are not chosen for fertility, as only the most sub-fertile bulls are culled, but are mainly chosen for production traits displayed in subsequent offspring, mainly milk production (Dejarnette et al, 2004; Colenbrander et al, 2003). Milk production is negatively correlated with fertility, and milking frequency has been associated as a possible risk factor for the trending lower fertility rates (Dejarnette et al, 2004). However, omitting fertility as a selection trait has lead to an increase in the number of days a cow is not pregnant (days open), thus decreasing fertility while increasing producer costs associated with managing these animals (Kuhn et al, 2008).

Improving fertility rates can alleviate costs associated with production. For example, a 1% increase in the Estimated Relative Conception Rate (ERCR), one measure of cow fertility based on non-return to pregnancy rate (NRR), could lead to a $5.01 increase in profit per animal (Pecsok et al, 1994). By improving conception rates using

AI, producers avoid costs associated with repeat inseminations of the same cow and milking days lost. Although milk production has significantly increased with the use of AI, more accurate male fertility assays are needed to reverse declining cow fertility and to improve conception rates.

*B. Male fertility*

Male fertility is one of the most important confounders of overall pregnancy rate in current dairy production (Funk, 2006). Only a small number of sires are needed for AI breeding due to the ability of each mature sire to produce approximately 100,000 semen straws annually with the potential to inseminate thousands of cows each year. The range of dairy sire fertility (23-96%) is larger than the average herd fertility (27-74%), suggesting that male fertility accounts for low and high herd conception rates (Flowers, 2013). Therefore, the selection of the most fertile sires has the potential to increase conception rates, while the selection and removal of sub-fertile sires could also reduce low fertility rates (Flowers, 2013). Selection of sires for production traits as well as above average fertility is essential to economic stability and herd management of the dairy industry (Foote, 2003).

*C. Current sire fertility assays*

*1. In vivo fertility assays*

Direct assessment of semen fertilizing ability in the female reproductive tract (*in vivo* fertility) is the most accurate assay of sperm function and therefore the ideal fertility assay. After insemination of a large number of cows with different sires, a number of

fertility estimates, including non-return rate (NRR), estimated relative conception rate (ERCR), and sire conception rate (SCR), can be used to rank sires by *in vivo* fertility estimates. The NRR fertility measurement determines how many cows return to estrus post-insemination, after either 55 or 70 days, and are therefore not pregnant (D'Amours et al, 2010). To get an accurate NRR for each sire, at least 300 inseminations are necessary. ERCR is a fertility ranking similar to NRR, where pregnancy is assumed but not known for all first services but ERCR scores are normalized to 0 for the average of the population. The newer SCR fertility index improves upon previous fertility indices by confirming pregnancy status through ultrasound, palpation, or blood tests. The SCR scoring system also includes including multiple services, a stud year effect, sire age, and inbreeding adjustments and does not round to the nearest whole number as is done in the ERCR estimate (Clay and McDaniel, 2001). The addition of these variables with conception rates can improve the accuracy of sire *in vivo* fertility measurements by as much as 20% (Kuhn and Hutchinson, 2008). In these fertility scoring systems, all data is averaged and normalized to a score of 0. For example, a sire that is 2% above average in its fertility estimation would have a +2 SCR while a sire with 1% below average fertility would have an SCR of -1. The average fertility range includes sires within the range of -2 to 2 SCR (source: Semex).

*In vivo* measurements of sire fertility are time consuming, expensive, and require special resources, including large numbers of cows, AI technicians, in-depth record keeping, and cow heat detection equipment. The fertility estimates NRR, ERCR and SCR do not take into account variables such as reliable estrus control systems, season,

inseminator, breed, and parity (Al-Makhzoomi et al, 2008). An optimal fertility assay would mirror *in vivo* sperm performance in a more efficient and economical manner.

*2. In vitro sire fertility assays*

Conventional *in vitro* fertility assays often can only detect the most sub-fertile individuals and are insufficient at accurately predicting *in vivo* fertility as normal semen parameters do not necessarily correlate to acceptable fertility (Petrunkina et al, 2007; Colengrander et al, 2003). Despite a decline in sire fertility in the past 35 years, measures of scrotal circumference, sperm concentration, and sperm morphological traits have remained consistent, demonstrating the ineffectiveness of these parameters to determine fertility accurately (Dejarnette et al, 2004). Examining sperm functional competence can give a better understanding of sperm quality than current assays for physical, and observable sperm traits (Petrunkina et al, 2007).

Several sperm traits currently used to predict sire fertility include scrotal circumference, sperm concentration, motility, morphology, DNA fragmentation, sperm-zona pellucida penetration and sperm-oocyte penetration (Foote, 2003). Sperm traits can be characterized as compensable or non-compensable. Compensable traits can be overcome, to an extent, by increasing sperm numbers during AI and include sperm concentration, motility, morphology, and sperm-zona pellucida penetration (Flowers, 2013). Noncompensable traits, or those traits for which increased sperm numbers cannot increase fertility, include sperm DNA fragmentation and *in vitro* oocyte penetration ability. Sperm with defects in noncompensable traits can often reach the site of fertilization but fail to fertilize or cause embryonic or fetal death (Ostermeier et al, 2002).

Noncompensable traits are more accurate fertility predictability measures (Flowers, 2013). Other noncompensable sperm traits that hold potential for more accurate fertility assays could include protein or gene expression that reflect sperm functional competence.

It has become evident that the best way to predict *in vivo* fertility is by evaluating several semen parameters. For example, analyzing percentage of sperm DNA fragmentation with conventional semen parameters including motility, morphology, and sperm concentration can better predict fertilization potential than analysis based solely on conventional semen parameters (Phillips et al, 2004; Omran et al, 2013). The inclusion of DNA, RNA and protein expression level assays, and other semen parameters could develop an assay that could more accurately predict in vivo fertility.

The identification of genes associated with fertility is a promising area for development of more accurate male fertility assays. Select single-nucleotide polymorphisms (SNPs) are correlated to SCR score in AI sires (Penagaricio et al, 2012). Genomic testing can also select for production traits, as well as the prevalence of certain diseases (De Donato et al, 2013; Pryce et al, 2012). With this approach, fertility assaying of associated DNA markers could potentially be conducted prior to puberty. This early sire testing would reduce costs associated with maintaining less fertile sires as they currently need to be housed, with associated costs, past maturity to determine if their semen is fertile (Foote et al, 2003).

Transcriptome analysis studies have recently explored potential male germ cells mRNAs involved in spermatogenesis that may also be predictive of fertility. SNP studies offer insight into the effects of slight changes in the DNA code but looking at the RNA profile could show differences in transcript presence/amount that may alter protein

production though no errors occur at the DNA level. Many of these studies utilize microarray to determine the transcriptomes for testis, epididymis, and early stage developmental germ cells and identify transcripts that may be key to successful fertility (Chalmel et al, 2007; Dean et al, 2008; Guyonnet et al, 2009). For example, comparison of testicular transcriptomes from infertile and fertile patients found a decrease in transcripts associated with testicular mRNA storage (Gatta et al, 2010). Transcriptome differences in fertile and infertile patients have also shed light on gametogenesis-associated genes could be critical for proper fertility (Chalmel et al, 2012). Elucidation of gene expression that is responsible for sperm production, or the process of spermatogenesis, may provide more accurate fertility assays.

**II. Spermatogenesis**

Spermatogenesis takes place in the seminiferous tubules of the testes where billions of sperm are produced per day. Individual sperm are genetically unique to ensure genetic diversity of offspring. A series of stages and events are necessary to produce fertile spermatozoa (Holt et al, 2004). An overview of spermatogenesis is depicted in Figure 1.

At the onset of male germ cell development, undifferentiated spermatogonia initiate spermatogenesis at the basal membrane of the seminiferous tubules and undergo several mitotic divisions during the proliferation stage (Boerke et al, 2007). Undifferentiated spermatogonia differentiate into A1-A4, I, and B spermatogonia; A-spermatogonia have the ability to revert to undifferentiated spermatogonia through the process of stem-cell renewal (Senger, 2003). This self-renewal process ensures a constant supply of original cells to maintain spermatogenesis indefinitely, constantly

providing the male with a supply of mature spermatozoa. The B-spermatogonia undergo one final mitotic division resulting in an abundance of primary spermatocytes, which then enter the second phase of spermatogenesis (Holstein et al, 2003).

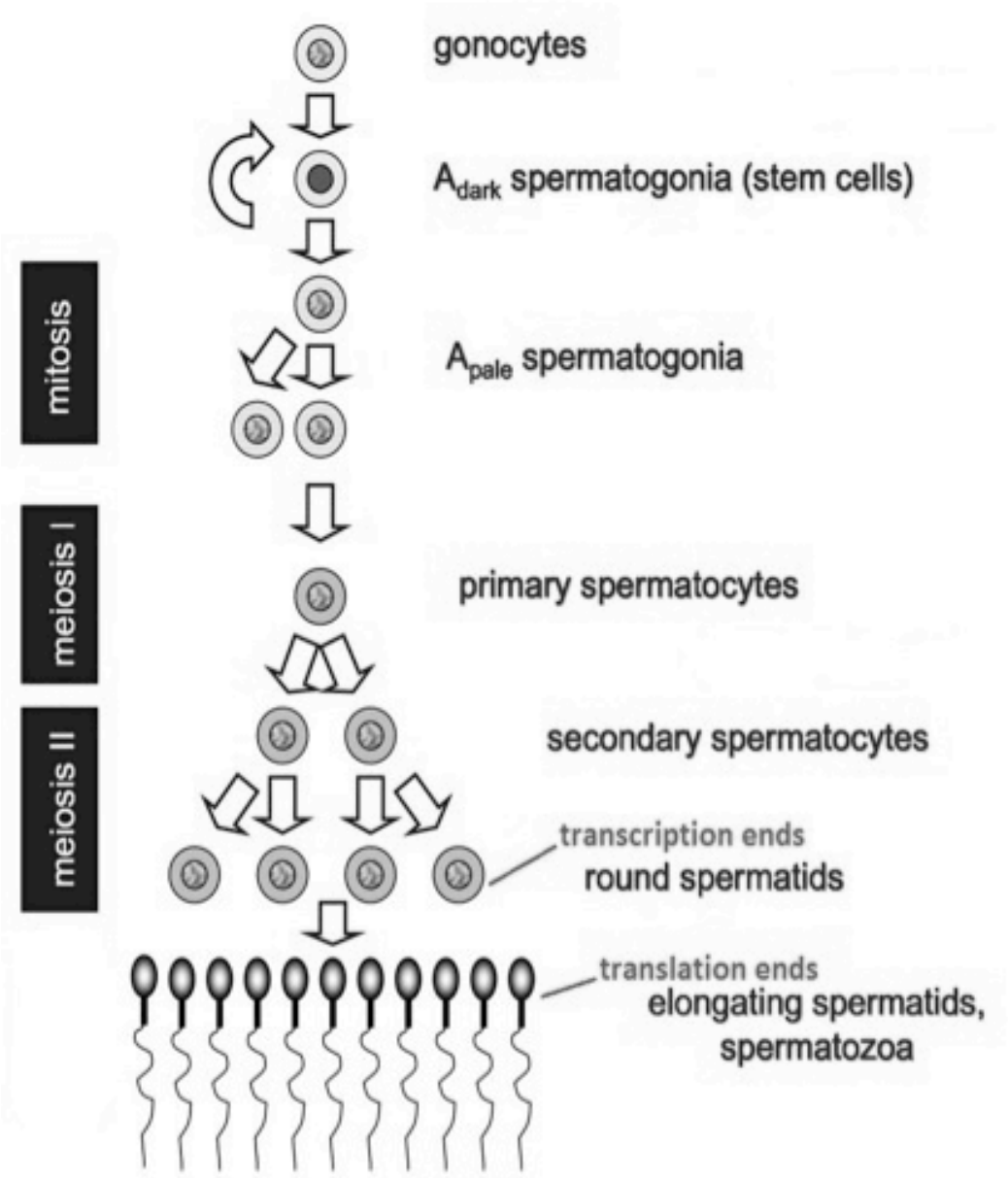



**Figure 1. Cellular divisions of spermatogenesis. The uncoupling of transcription and translation is highlighted. (www.springerimages.com)**

Two meiotic cell divisions occur during the primary and secondary spermatocyte stages of spermatogenesis (Holstein et al, 2003). Homologous recombination occurs during the first meiotic division of primary spermatocytes to maintain genetic diversity. Spermatocytes then complete meiosis II to form haploid round spermatids (Holstein et al, 2003). The haploid round spermatids move into the third and final phase of spermatogenesis, the differentiation phase.

During the differentiation phase, the newly formed round spermatids undergo a series of morphological changes to produce a functional head and tail for a fully formed spermatozoon. It is during this final morphological stage that the sperm head develops from the rearrangement of several vesicles from the Golgi apparatus fusing along the nucleus, forming the pro-acrosomic granules that eventually give rise to the acrosomal enzyme layer that is necessary for sperm-egg interaction. The head houses all of the genetic material with the acrosome spread across the anterior portion of the head. The acrosome consists of various hydrolytic enzymes that help the sperm penetrate the zona pellucida of the oocyte during fertilization (Eddy, 2002). The sperm tail, the motor of the sperm, develops from centriole elongation and localization of mitochondria along the sperm midpiece. Removal of excess cytoplasm during the final stages ensures that transcription and translation cease (Senger 2003). After release of the sperm from the seminiferous tubules in the testes, sperm acquire motility and fertilizing potential while maturing in the epididymis (Senger 2003).

Transcription is halted during the late round spermatid stage/early elongated spermatid stage of spermatogenesis while translation continues until the late elongated spermatid stage (Braun, 2000; Eddy, 2002; Figure 1). This uncoupling of transcription

and translation is necessary for protein production for formation of the late stage spermatids at the same time when the genomic DNA is being highly compacted, and therefore inactive, in the sperm head (Eddy, 2002). Some transcripts can be stored in an inactive state in male germ cells for several days before they are translated and used as functional protein in the later stages of spermatogenesis, as ribosomal RNAs are still present in the elongating spermatids (Braun, 2000). Translation ceases as excess cytoplasm and ribosomes are removed in the fully formed spermatozoa. If silenced transcripts are not translated as the elongated spermatid develops then a population of mRNAs, from the earlier stages of spermatogenesis can remain in the transcriptionally and translationally-inert spermatozoa (Miller et al, 2006; Boerke et al, 2007; Bissonnette et al, 2009). This spermatozoal mRNA population has potential as a fertility assay based on the ability to reflect past events of gene expression during spermatogenesis.

## III. Spermatozoal RNA

### *A. Spermatozoal RNA*

It has long been thought that the only function of spermatozoa was to deliver the paternal half of the genome to the egg for fertilization. However, more recently, spermatozoa have been shown to also deliver other components to the oocyte at fertilization that are necessary for successful early embryogenesis, including a population of RNAs (as reviewed in Hamatani, 2012; Jodar et al., 2013). Spermatozoa contain lower amounts of RNA (5-400 fg/sperm depending on species) in contrast to oocytes (0.5-1.5 ng) and other cell types (10-30 pg; Boerke et al, 2007; Das et al, 2010; Hamatani 2012). The sperm RNA population includes coding and non-coding RNAs, such as messenger

RNA (mRNA), microRNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), and long non-coding (lncRNA) (Jodar et al, 2013). While the focus of this literature review is the spermatozoal mRNA population, the non-coding RNAs, long non-coding RNAs, chromatin-associated RNAs, and small-nuclear RNAs may aid processes critical for fertilization and early embryonic development demonstrating a functional role for sperm RNA (Jodar et al, 2013). For example, in human spermatozoa, the microRNA miR-34C is critical for early zygotic cell division (Liu et al, 2012). piRNAs regulate stability of RNAs and protect the genome by binding to DNA and preventing the action of transposable elements (Kawano et al, 2012).

*B. Spermatozoal mRNA*

In humans, 3,000-7,000 transcripts have been identified in sperm; while in the bovine over 6,000 transcripts have been identified (Boerke et al, 2007; Chapter 2 Card and Anderson et al, 2013). As stated previously, because spermatozoa are transcriptionally-inactive, these mRNAs must originate from the stages of spermatogenesis when transcription occurs as all spermatozoal transcripts identified by microarray are expressed in round spermatids (Gilbert et al., 2007). The spermatozoal mRNA population contains degraded as well as intact transcripts capable of translation into functional proteins (Gilbert et al, 2007; Chapter 2 Card and Anderson et al, 2013; Anderson et al., in prep) with potential functional and non-functional uses as a fertility assay.

*C. Potential role of spermatozoal mRNAs in fertilization and embryonic development*

A small portion of the spermatozoal mRNA population consists of full-length transcripts with poly-A tails, and therefore has the potential to be translated into protein post-fertilization (Gilbert et al, 2007; Chapter 2 Card and Anderson et al., 2013, Anderson et al, 2014 in prep). Spermatozoal mRNA is delivered to the oocyte during fertilization and therefore has been hypothesized to have a role in fertilization and early embryonic development such as the processes of morphogenesis and implantation (Ostermeier et al, 2004). A significant amount of spermatozoal mRNA is rapidly degraded but some spermatozoal transcripts remain intact during early stages of embryogenesis (Boerke et al, 2007). Not only have mRNAs been shown to persist in zygotes, but can also be found post-fertilization until the blastocyst stage, which has lead to speculation that they may function to aid early embryonic development (Hata et al, 2007). Spermatozoal transcripts delivered to the oocyte at fertilization and present at 24 hours post-fertilization in the zygote include *INTS1, PLCZ1*, *PRM1, PRM2, CLU, PSG-1,* and *HLA-E* (Ostermeier et al, 2004; Kempisty et al, 2008; Avendano et al, 2009; Swann et al., 2006; Vassena et al, 2011). The mouse spermatozoal transcript, *Ints1,* is transcribed in the round spermatid stage, selectively silenced and retained in sperm then and detected post-fertilization prior to activation of the embryonic genome. The *INTS1* knockout embryos arrest at the blastocyst stage, thus demonstrating the importance of this sperm-borne transcript in early embryonic development (Vassena et al, 2011). The sperm RNA *PLCZ1* is translated in the oocyte and initiates oocyte activation at fertilization (Swann et al, 2006). These transcripts are examples of a function for sperm-specific mRNA and may indicate that other spermatozoal mRNAs are necessary for

successful fertilization and embryogenesis. It is evident that spermatozoal RNAs, once an afterthought, are likely critical for successful reproduction and could aid many processes necessary for fertility.

*D. Spermatozoal RNAs, sperm motility and DNA fragmentation*

The amount of individual spermatozoal RNAs is correlated with sperm functional parameters such as sperm concentration, motility, morphology and DNA fragmentation (Lambard et al, 2004; Aoki et al, 2005; Aoki et al, 2006; Bissonnette et al, 2009). Specifically, sperm motility levels are correlated with the ratio of protamine transcripts, *Prm1/Prm2* ratios, in human sperm; high motility samples had a low ratio of *Prm1/Prm2* (Lambard et al, 2004). Increased *Prm1*/*Prm2* transcript levels and corresponding low protein levels in an individual could be an indicator of inefficient translation during spermatogenesis (Aoki et al, 2006). The protamines (PRM1 and PRM2) have an essential role in spermatogenesis by replacing DNA histones, acquired during the early stages of spermatogenesis, to tightly compact the DNA (Braun, 2001). Protamines are small, arginine-rich nuclear proteins with a strong positive charge, which allows them to bind tightly to the negatively charged DNA strands, protecting sperm genomic DNA before fertilization while also making the sperm nucleus more compact and hydrodynamic (Balhorn, 2007). This new tight packaging of the chromatin via protamines also makes the DNA inaccessible to transcription factors, thus causing the cessation of transcription (Kimmins et al, 2004; Aoki et al, 2005). In human spermatozoa, 85% of chromatin is bound to protamines and 15% remains histone bound while in other mammalian species it is estimated that 98% of chromatin is bound to

protamines (Oliva, 2006). Specific measurements of fertility, including *in vitro* fertilization, sperm concentration, motility, and morphology, are all negatively affected when the PRM1/PRM2 protein ratios are abnormally high or low (Aoki et al, 2006). DNA fragmentation, or the extent of damaged chromatin in the sperm nucleus, is significantly higher in sperm with low PRM1/PRM2 protein ratios (Aoki et al, 2005). Other transcripts that have been correlated with motility in bovine spermatozoa include *TSSK6* and *ADAM5P* (Bissonnette et al, 2009).

*E. Sperm mRNAs as a potential predictor of fertility*

Due to the lack of transcription, the transcripts present in spermatozoa represent spermatogenic gene expression, potentially indicating efficiency of the process as a whole, possibly useful as a "fingerprint" of past events during spermatogenesis (Ostermeier et al., 2002; Lambard et al., 2004; Ostermeier et al., 2005; Platts et al., 2007; Lalancette et al., 2009). Spermatogenic defects could potentially be represented by variation in transcript profiles between normal and abnormal spermatogenic samples while also predicting fertility (Jodar et al, 2013). For example, spermatozoal transcripts associated with the ubiquitin-proteosome pathway, acrosome formation, and proper tail formation have different expression levels between normal semen and teratozoospermic (abnormal morophology) samples (Platts et al., 2007). Spermatid and spermatocyte specific transcripts are also not present in teratozoospermatozic samples thus indicating errors in the later stages of spermatogenesis (Platts et al., 2007). Both degraded and full-length transcripts may be of importance as degraded transcripts may have been critical

during earlier stages of spermatogenesis while the full-length transcripts may be critical for the potential function during fertilization or early embryonic development.

*1. Global expression of sperm mRNAs in fertility*

In order to identify the most useful candidate fertility spermatozoal transcripts, global comparison of spermatozoal transcript expression is needed by comparing complete transcript profiles of fertile with sub-fertile or infertile sires. To date, sperm transcript profile comparison studies have been conducted using hybridization-based methods, primarily microarrays that do not probe complete transcriptomes as described below. Overall, microarray studies of spermatozoal transcript profiles for higher and lower fertility individuals have shown several common transcripts between populations and differences in the presence of some transcripts as well as transcript abundances. In humans, fertile and infertile patients differ in transcript presence associated with cellular remodeling pathways (Platts et al, 2007; Ostermeier et al, 2005). Over 1,700 transcripts are common in spermatozoal RNA populations amongst 23 fertile human subjects while 96 and 37 transcripts are common between high and medium quality semen samples and medium and poor quality semen samples respectively (Lalancette et al., 2009; Ostermeier et al., 2005). Comparison of spermatozoal RNA from an individual with pooled spermatozoal RNA has shown conservation of transcripts across individuals, potentially stably regulated transcripts, while some individual variation is also present (Ostermeier et al, 2002).

Comparison of spermatozoal transcript profiles in bovine has also been conducted using microarrays. Significant differences in transcript number and amount were

reported between high and low NRR fertility groups (high-NRR ≥71%; low-NRR ≤65%; Lalancette et al, 2008b). The high-NRR group contained 17% transcripts that aligned to the bovine genome and 29% of these transcripts had a known function, while the low-NRR group only 3% aligned to the bovine genome and 10% had a known function. Specifically, the *PRM2* transcript was at higher levels in the low-NRR group (Lalancette et al, 2008b). In a separate microarray study, the expression levels of 415 sperm transcripts differed between high and low fertility sires; 211 transcripts were detected at levels at least twice as high in the higher fertility population, while 204 transcripts were detected at least twice as high in the low fertility population (Feugang et al, 2010). It is apparent from global expression studies that there are marked expression differences of spermatozoal RNAs between individuals with different fertility. These differential expression studies have shown a sub-population of transcripts that would be candidates for further study on an individual transcript basis. Exploring individual transcripts to attempt to elucidate those that may have correlations could lead to the potential for transcripts being used as predictors of fertility.

*2. Individual spermatozoal transcripts as fertility markers*

Individual spermatozoal transcripts that are candidate fertility markers have been identified in sires using microarrays. Many transcripts, including, *CRISP2*, *CCT8*, *IB5*, *TIMP2*, are positively correlated with fertility while *PEBP1* is negatively correlated with fertility (Table 1). In humans, a number of additional transcripts have shown correlations with fertility including *Prm1*, *Prm2*, *Bcl2*, *Psg1*, and *Hlae* (Aoki et al, 2006; Steger et al, 2008; Avendano et al, 2009). A correlation of spermatozoal transcript level with sire

fertility may exist because amount of transcript remaining in the terminally-differentiated cells could be due to insufficient copies of a transcript synthesized due to transcriptional errors and inefficient translation of mRNAs, as discussed above for Prm1/Prm2 ratios and protein levels, resulting in inefficient spermatogenesis and subfertile spermatozoa. Additionally, a functional role for spermatozoal transcripts after fertilization, as discussed previously, may exist and contribute to sperm fertility.

**Table 1. Individual spermatozoal transcripts with correlations to bovine fertility.**

| Transcripts | Parameter | mRNA Abundance | Reference |
|---|---|---|---|
| *PEBP1, CRISP2* | high SCR | increase | Arangasamy et al. 2011 |
| *CCT8* | high SCR | decrease | |
| *CD36, CENPA* | high fertility | increase | Feugang et al. 2010 |
| *AK1, IB5, DOPPEL, NGF, TIMP2, LDC1, SNRPN, ODF2, PLCz1* | high SCR | increase | Kasimanickam et al. 2012 |
| *ADIPONECTIN, ADIPOR1, ADIPOR2* | high SCR | increase | Kasimanickam et al. 2013 |

## IV. RNA-Seq

### *A. Comparison of microarray and RNA-Seq methods*

Although microarrays have been powerful tools for comparative transcriptome studies, deficiencies in the complete transcriptome coverage and the reliance on hybridization methods make this data incomplete. The newer direct sequencing technology of RNA-Seq offers many benefits over previous hybridization-based techniques (Werner, 2011). RNA-Seq, including the Illumina sequencing which was used for our studies, has the capability of directly sequencing the entire expressed transcript population by amplifying short cDNA and then fluorescently tagging a single base at a time generating tens of millions of short reads from 30-400 bp in length (Wang et al, 2009; www.illumina.com). The sequencing reads can then be assembled *de novo* or

against a reference genome for a complete transcriptome (Wang et al, 2009). Direct sequencing is an advantage over the short sequence of cDNA probes used for microarray chips that do not cover entire transcripts potentially resulting in misrepresentation of truncated transcripts and transcript isoforms. Also, the lack of transcript coverage in microarrays compromises the ability to discriminate between fully-intact and degraded transcripts therefore RNA-Seq has the potential for novel discovery of transcripts, exons, and splicing junctions (Wang et al, 2009; Costa et al, 2010; Werner, 2011). Another benefit of RNA-Seq is the ability to detect single nucleotide polymorphisms (SNPs) as well as exon-exon boundaries for all expressed transcripts in the sample (Costa et al, 2010). Due to increased sensitivity with decreased background noise, accurate quantification of transcripts can also be obtained with RNA-Seq (Werner, 2011). With great dynamic range of sequencing, this quantification can be accurate at both very low and high levels (Wang et al, 2009). Using RNA-Seq will allow for more complete sequencing and comparison of transcriptomes for more accurate identification of the role of gene expression in mammalian reproduction and sire fertility.

*B. RNA-Seq in reproduction*

In mammalian reproduction, RNA-Seq has been used recently to reveal new insights into the presence, function, and importance of transcripts at various reproductive stages with a primary focus on acquisition of reproductive function in adults and embryo quality. In non-human primates, comparison of transcript profiles of juvenile and adult ovarian tissues revealed a population of up-regulated transcripts in adults that are targets for developmental studies (Babbitt et al, 2011). Using high throughput sequencing, the

infant (6 day postnatal) testicular tissue expression profiles were found to differ between the juvenile (4 week) and adult (10 week) profiles, especially for transcripts associated with spermatogenesis, thus showing the onset of spermatogenesis for the juvenile stage (Wei et al, 2013). These developmental studies have offered some insight into the timing of maturity as well as critical transcripts in mature animals. At the adult level, the testis transcript profiles of two different pig breeds, Iberian and Large White, have been sequenced and differential expression between the two breeds was prevalent. Many of these transcripts were associated with spermatogenesis and lipid metabolism (Esteve-Codina et al, 2011).

RNA-Seq studies in embryos have shed light on important gene regulatory pathways. Porcine embryonic germ cells have also been sequenced, identifying genes associated with pluripotency (Petkov et al, 2011). In rhesus macaques, transcriptome comparisons of embryos resulting from fertilization of normal or reactive oxidative species (ROS) damaged sperm identified 40 transcripts that were differentially expressed between the two embryo types (Burruel et al, 2013). Single cell blastomeres from mouse embryos have also been sequenced to examine allele specific gene expression patterns (Tang et al, 2011). The high throughput sequencing of various placentation sites in rats showed differential gene expression based upon the different compartments (Shankar et al, 2012).

In the bovine, RNA-Seq has been recently used to investigate proper embryonic development. In a single bovine blastocyst, RNA-Seq analysis uncovered 9,489 transcripts and a high number of novel transcriptional units providing evidence of an incomplete bovine genome (Chitwood et al, 2013). Unique transcript expression was

also reported when comparing the transcriptomes of normal and abnormal bovine blastocysts sequenced using RNA-Seq (Huang and Khatib, 2010). One transcript was even found almost exclusively expressed in the degenerative embryo population, indicating potential for it to be used as a marker of abnormal embryo development (Huang and Khatib, 2010). *In vitro* fertilization alters the transcriptome of embryos versus *in vivo* counterparts, as 793 genes were differentially expressed, though both pools of embryos were morphologically comparable (Driver et al, 2012). Y-chromosome specific reads have also been reported from RNA-Seq of testis tissue (Chang et al, 2013).

RNA-Seq has just recently been applied to spermatozoal RNA populations as well. To date, spermatozoa from two species, the human and horse have been sequenced utilizing this technology (Das et al, 2013; Sendler et al, 2013; Note: these manuscripts were published after the Chapter 2 Biology of Reproduction manuscript of this dissertation). The total human spermatozoal RNA population contains many coding RNAs as well as non-coding RNAs such as snRNAs, miRNAs, and pri-miRNAs. The population was also found to include fully intact RNAs, as also reported in Chapter 2 of this dissertation, which is crucial for potential downstream significance (Sendler et al, 2013). In the horse, over 4,500 transcripts were detected in the spermatozoal population, including rRNAs and 82 miRNAs. Gene ontology analysis elucidated that some processes such as chemoattractant-activated signal transduction and cellular components related to sperm functions at fertilization were prominent (Das et al, 2013). To date, no RNA-seq studies have been conducted using bovine spermatozoa to characterize the spermatozoal transcript profile for this species.

## V. Summary

Without transcription occurring, mRNA in the spermatozoa must originate from earlier stages of spermatogenesis (Hamatani, 2012). Consequently, spermatozoal mRNA could be considered a "fingerprint", or indicator of the efficiency of spermatogenesis. By examining the spermatozoal mRNA population and comparing levels of transcripts that are critical for sperm function, it may be possible to determine the success and efficiency of spermatogenic events. Specific transcripts that have roles during spermatogenesis could, in turn, become markers for sire fertility (Li et al, 2012). Assaying spermatozoal transcripts has great potential as a new non-invasive measure of sire fertility, as testis biopsies to measure gene expression typically are not possible (Miller et al, 2006). While individual spermatozoal mRNAs have been identified, the complete bovine spermatozoal transcript profile has not been sequenced and the correlation of sperm mRNA levels with sperm functional assays including fertility, morphology and DNA fragmentation have not been explored.

## VI. Aims

**Aim 1:** Sequence and characterize the transcriptome of cryopreserved bull spermatozoa using next-generation RNA sequencing.

**Aim 2:** Compare transcriptomes of two pools of cryopreserved bull spermatozoa, low and high fertility, based on CR score.

**Aim 3:** Correlate total RNA/10 million spermatozoa from cryopreserved bull spermatozoa with DNA fragmentation, and sperm head morphology characteristics.

Al-Makhzoomi A, Lundeheim N, Haard M, Rodriguez-Martinez H (2008): Sperm morphology and fertility of progeny-tested AI dairy bulls in Sweden. Theriogenology, 70(4): 682-691.

Aoki VW, Moskovtsev SI, Willis J, Liu L, Mullen BM, Carrell DT (2005): DNA integrity is compromised in protamine-deficient human sperm. Journal of Andrology, 26: 741-748.

Aoki VW, Liu L, Jones KP, Hatasaka HH, Gibson M, Peterson MC, Carrell DT (2006): Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. Fertility and Sterility, 86: 1408-1415.

Arangasamy A, Kasimanickam VR, DeJarnette JM, Kasimanickam RK (2011): Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. Theriogenology, 76: 570-577.

Avendano C, Franchi A, Duran H, Oehninger S (2009): DNA fragmentation of normal spermatozoa negatively impacts embryo quality and intracytoplasmic sperm injection outcome. Fertility and Sterility, 94: 549-557.

Babbitt CC, Tung J, Wray GA, Alberts SC (2011): Changes in gene expression associated with reproductive maturation in wild babboons. Genome Biology and Evolution, 4(2): 102-109.

Balhorn R (2007): The protamine family of sperm nuclear proteins. Genome Biology, 8:227.

Bissonnette N, Levesque-Sergerie JP, Thibault C, Boissonneault G (2009): Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality. Reproduction, 138(1): 65-80.

Boerke A, Dieleman SJ, Gadella BM (2007): A possible role for sperm RNA in early embryo development. Theriogenology, 68S: 147-155.

Braun RE (2000): Temporal control of protein synthesis during spermatogenesis. International Journal of Andrology, 23: 92-94.

Braun RE (2001): Packaging paternal chromosomes with protamine. Nature Genetics, 28: 10-12.

Burruel V, Klooster KL, Chitwood J, Ross PJ, Meyers SA (2013): Oxidative damage to rhesus macaque spermatozoa results in mitotic arrest and transcript abundance changes in early embryos. Biology of Reproduction, in press.

Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J, Wolgemuth DJ, Jegou B, Primig M (2007): The conserved transcriptome in human and rodent male gametogenesis. Proceedings of the National Academy of Science (PNAS), 104(20): 8346-8351.

Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jegou B, Kirchhoff C, Primig M (2012): Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male fertility. Human Reproduction, 27(11): 3233-3248.

Chang T, Yang Y, Retzel EF, Liu W (2012): Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. PNAS, 110(30): 12373-12378.

Chitwood JL, Rincon G, Kaiser GG, Medrano JF, Ross PJ (2013): RNA-Seq analysis of single bovine blastocysts. BMC Genomics, 14: 350-364.

Clay JS and McDaniel BT (2001): Computing mating bull fertility from DHI nonreturn data. Journal of Dairy Science, 84(5): 1238-1245.

Colenbrander B, Gadella BM, Stout TA (2003): The predictive value of semen analysis in the evaluation of stallion fertility. Reproduction in Domestic Animals, 38(4): 305-311.

Costa V, Angelini C, De Feis I, Ciccodicola A (2010): Uncovering the complexity of transcriptomes with RNA-Seq. Journal of Biomedicine and Biotechnology.

Cottle DJ and Coffey MP (2013): The sensitivity of predicted financial and genetic gains in Holsteins to changes in the economic value of traits. Journal of Animal Breeding and Genetics, 130: 41-54.

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 PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, Varner DD, Chowdhary BP, Raudsepp T (2010): Total RNA isolation from stallion sperm and testis biopsies.Theriogenology, 74: 1099-1106

Das PJ, McCarthy F, Vishnoi M, Paria N, Gresham C, Li G, Kachroo P, Sudderth AK, Teague S, Love CC, Varner DD, CHowdhary BP, 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

De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG (2013): Genotyping-by-Sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next generation sequencing. PLOS One, 8(5): e62137.

Dean MD, Good JM, Nachman MW (2008): Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. Molecular Biology and Evolution, 25(2): 383-392.

Dejarnette J, Marshall C, Lenz RW, Monke DR, Ayars WH, Sattler CG (2004): Sustaining the fertility of artificially inseminated dairy catle: the role of the artificial insemination industry. Journal of Dairy Science, 87: 93-104.

Driver AM, Penagaricano F, Huang W, Ahmad KR, Hackbart KS, Wiltbank MC, Khatib H (2012): RNA-Seq analysis uncovers transcriptomic variations between morophologically similar in vivo- and in vitro- derived bovine blastocysts. BMC Genomics, 13:118.

Eaglesome MD and Garcia MM (1997): Disease risks to animal health from artificial insemination with bovine semen. Scientific and Technical Review of the Office International des Epizooties, 16(1): 215-225.

Eddy EM (2002): Male germ cell gene expression. Recent Progress in Hormone Research, 57: 103-128.

Esteve-Codina A, Kofler R, Palmieri N, Bussotti G, Notredame C, Perez-Enciso M (2011): Exploring the gonad transcriptome of two extreme male pigs with RNA-seq. BMC Genomics, 12: 552-565.

Feugang JM, Kaya A, Page GP, 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 a potential role in bull fertility. BMC Genomics, 10: 176-185.

Feugang JM, Rodriguez-Osorio N, Kaya A, Wang H, Page G, Ostermeier GC, Topper EK, Memili E (2010): Transcriptome analysis of bull spermatozoa: implications for male fertility. Reproductive BioMedicine, 21: 312-324.

Flowers W (2013): Sperm characteristics that limit success for fertilization. Journal of Animal Science, 2010.

Foote RH (2003): Fertility estimation: a review of past experience and future prospects. Animal Reproduction Science, 75: 119-139.

Funk DA (2006): Major advances in globalization and consolidation of the artificial insemination industry. Journal of Dairy Science, 89(4): 1362-1368.

Gatta V, Ralcu F, Ferlin A, Antonucci I, Scioletti AP, Garolla A, Palka G, Foresta C, Stuppla L (2010): Testis transcriptome analysis in male infertility: new insight on the pathogenesis of oligo-azoospermia in cases with and without AZFc microdeletion. BMC Genomics, 11: 401.

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.

Guyonnet B, Marot G, Dacheux J, Mercat M, Schwob S, Jaffrezic F, Gatti J (2009): The adult boar testicular and epididymal transcriptomes. BMC Genomics, 10: 369.

Hamatani, T (2012): Human spermatozoal RNAs. Fertility and Sterility, 97(2): 275-281.

Hata T and Nakayama M (2007): Targeted disruption of the murine large nuclear KIAAI440/Ints1 protein causes growth arrest in early blastocyst stage embryos and eventual apoptotic cell death. Biochim Biophys Acta, 1773: 1039-1051.

Holstein AF, Schulze W, Davidoff M (2003): Understanding spermatogenesis is a prerequisite for treatment. Reproductive Biology and Endocrinology, 1: 107-123.

Holt WV and Van Look KJW (2004): Concepts in sperm heterogeneity, sperm selection and sperm competition as biological foundations for laboratory tests of semen quality. Reproduction, 127: 527-535.

Huang W and Khatib H (2010): Comparison of transcriptomic landscapes of bovine embryos using RNA-Seq. BMC Genomics, 11: 711.

Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA (2013): The presence, role and clinical use of spermatozoal RNAs. Human Reproduction Update, 0(0): 1-21.

Kasimanickam V, Kasimanickam R, Arangasamy A, Saberivand A, Stevenson JS, Kastelic JP (2012): Association between mRNA abundance of functional sperm function proteins and fertility of Holstein bulls. Theriogenology, 78(9): 2007-2019.

Kasimanickam VR, Kasimanickam RK, Kastelic JP, Stevenson JS (2013): Associations of adiponectin and fertility estimates in Holstein bulls. Theriogenology, 79: 766-777.

Kawano M, Kawaji H, Grandjean V, Kiani J, Rassoulzadegan M (2012): Novel small noncoding RNAs in mouse spermatozoa, zygotes, and early embryos. PLoS One, 7: e44542.

Kempisty B, Antosik P, Bukowska D, Jackowska M, Lianeri M, Jaskowski JM, Jagodzinski PP (2008): Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reproduction, Fertility and Development, 20(4): 513-518.

Kimmins S, Kotaja N, Davidson I, Sassone-Corsi P (2004): Testis-specific transcription mechanisms promoting male germ-cell differentiation. Reproduction, 128: 5-12.

Kuhn MT and Hutchinson JL (2008): Prediction of dairy bull fertility from field data: use of multiple services and identification and utilization of factors affecting bull fertility. Journal of Dairy Science, 91(6): 2481-2482.

Kuhn MT, Hutchinson JL, Norman HD (2008): Modeling nuisance variables for prediction of service sire fertility. Journal of Dairy Science, 91(7): 2823-2835.

Lalancette C, Miller D, Li Y, Krawetz SA (2008a): Paternal contributions: new functional insights for spermatozoal RNA. Journal of Cellular Biochemistry, 104(5): 1570-1579.

Lalancette C, Thibault C, Bachand I, Caron N, Bissonnette N (2008b): Transcriptome analysis of bull semen with extreme nonreturn rate: Use of suppression-subtractive hybridization to identify functional markers for fertility. Biology of Reproduction, 78: 618-635.

Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, Krawetz SA (2009): Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. 87:735-748.

Lambard S, Galeraud-Denis I, Martin G, Levy R, Chocat A, Carreau S (2004): Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. Molecular Human Reproduction, 10: 535-541.

Li C and Zhou X (2012): Gene transcripts in spermatozoa: markers of male infertility. Clinical Chimica Acta, 413(13-14): 1035-1038.

Liu WM, Pang RT, Chiu PC, Wong BP, Lao K, Lee KF, Yeung WS (2012): Sperm-born microRNA-34c is required for the first cleavage division in mouse. Proceedings National Academy of Science USA, 109: 490-494.

Miller D, Ostermeier GC, Krawetz SA (2005): The controversy, potential and roles of spermatozoal RNA. Trends in Molecular Medicine, 11(4): 156-163.

Miller D and Ostermeier GC (2006): Towards a better understanding of RNA carriage by ejaculate spermatozoa. Human Reproduction, 12(6): 757-767.

Oliva R (2006): Protamines and male infertility. Human Reproduction, 12: 417-435.

Omran HM, Bakhiet M, Dashti MG (2013): DNA integrity is a critical molecular indicator for the assessment of male infertility. Molecular Medicine and Reproduction, 7(5): 1631-1635.

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA (2002): Spermatozoal RNA profiles of normal fertile men. The Lancet, 360: 772-777.

Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA (2004): Delivering spermatozoan RNA to the oocyte. Nature, 429: 154.

Ostermeier GC, Goodrich RJ, Diamond MP, Dix DJ, Krawetz SA (2005): Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. Fertility and Sterility, 26: 1687-1694.

Pecsok SR, McGilliard ML, Nebel RL (1994): Conception Rates. 2. Economic value of unit differences in percentages of sire conception rates. Journal of Dairy Science, 77: 3016-3021.

Peddinti D, Nanduri B, Kaya A, Feugang JM, Burgess SC, Memili E (2008): Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. BMC Systems Biology, 2:19.

Penagaricano GL, Weigel KA, Zhang Y, Rosa G, Khatib H (2012): Comparitive genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. Journal of Dairy Science, 95: 6122-6129.

Petkov SG, Marks H, Klein T, Garcia RS, Gao Y, Stunnenberg H, Hyttel P (2011): In vitro culture and characterization of putative porcine embryonic germ cels derived from domestic breeds and Yucatan mini pig embryos at days 20-24 of gestation. Stem Cell Research, 6: 226-237.

Petrunkina AM, Waberski D, Gunzel-Apel AR, Topfer-Peterson E (2007): Determinants of sperm quality and fertility in domestic species. Reproduction, 134(1): 3-17.

Phillips NJ, Mcgowan MR, Johnston SD, Mayer DG (2004): Relationship between thirty post-thaw spermatozoal characteristics and the field fertility of 11 high-use Australian dairy AI sires. Animal Reproduction Science, 81(1-2): 47-61.

Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, Rawe VY, Quintana S, Diamond MP, Strader LF, Krawetz SA (2007): Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Human Molecular Genetics, 16(7): 763-773.

Pryce J, Hayes B (2012): A review of how dairy farmers can use and profit from genomic technologies. Animal Reproduction Science, 52: 180-184.

Sendler E, Johnson GD, Mao S, Goodrich RJ, Diamond MP, Hauser R, Krawetz SA (2013): Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Research, 1-14.

Senger PL (2005): Pathways to pregnancy and parturition.

Shankar K, Zhong Y, Kang P, Blackburn ML, Soares MJ, Badger TM, Gomez-Acevedo H (2012): RNA-Seq analysis of the functional compartments within the rate placentation site. Endocrinology, 153(4): 1999-2011.

Steger K, Wilhelm J, Konrad L, Stalf T, Greb R, Deimer T, Kliesch S, Bergmann M, Weidner W (2008): Both protamine-1 to protamine-2 mRNA ratio and Bcl2 mRNA content in testicular spermatids and ejaculated spermatozoa discriminate between fertile and infertile men. Human Reproduction, 23: 11-16.

Swann K, Saunders CM, Rogers NT, Lai FA (2006): PLCz(zeta): A sperm protein that triggers $Ca^{2+}$ oscillations and egg activation in mammals. Seminars in Cell & Developmental Biology, 17: 264-273.

Tang F, Barbacioru C, Nordman E, Bao S, Lee C, Wang X, Tuch BB, Heard E, Lao K, Surani MA (2011): Deterministic and stochastic allele specific gene expression in single mouse blastomeres. PLoS One, 6.

Vassena R, Boue S, Gonzalez-Roca E, Aran B, Auer H, Veiga A, Izpisua BJC (2011): Waves of early transcriptional activiation and pluripotency program initiation during human preimplantation development. Development, 138: 3699-3709.

Wang Z, Gerstein M, Snyder M (2009): RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1): 57-63.

Wei G, Linlin P, Qiang L, YuanYuan Z, CHangQi X, XiaoMin YU, SongNian HU, Jun YU (2013): Transcriptome profiling of the developing postnatal mouse testis using next-generation sequencing. Science China, 56: 1-12.

Werner T (2011): Next generation sequencing allows deeper analysis and understanding of genomes and transcriptomes including aspects to fertility. Reproduction, Fertility and Development, 23(1): 75-80.

**Timeline of Experiments**

1. Sperm RNA Isolation Method #1 (RNeasy; column based approach)
    A. Correlation of sperm RNA amount with sperm morphology and DNA Fragmentation **(Appendix 1)**
        I. Non-amplified RNA
            Although RNA isolation method #1 yielded RNA with acceptable 260/280 ratios (1.85) and no detectable somatic cell contamination, I was unable to consistently amplify known sperm mRNA transcripts. RNA amounts from these experiments were correlated with sperm head morphology and DNA fragmentation measurements taken from bulls from the same batch.

    B. RNA-Seq analysis of amplified spermatozoal mRNA (Isolation method #1; **Appendix 2)**
        I. Amplified mRNA population
            With the potential that transcripts from RNA isolation method (#1) were present at very low levels, sperm RNA was then linearly amplified. Amplified mRNA was devoid of somatic cell contamination and ribosomal RNA. The *PRM1* transcript was amplified successfully therefore this amplified sperm RNA was submitted for RNA-Seq.

        II. RNA-Seq results
            Several thousand transcripts were found to be expressed in the in the amplified sperm RNA including many sperm-specific genes. Though *PRM1* was amplified in the sample via PCR before sequencing, it was not expressed in the transcript profile. There were also inconsistencies when trying to amplify transcripts from the transcript profile in similarly isolated bull sperm RNA. From here I decided troubleshoot the RNA isolation method.

2. Sperm RNA Isolation #2 (TRIzol)
    A. Characterization of bovine spermatozoal transcript profile; Biology of Reproduction manuscript **(Chapter 2)**
        I. TRIzol spermatozoal RNA isolation (RNA isolation method #2)
            TRIzol spermatozoal RNA isolation resulted in lower 260/280 ratios (1.60). After DNAse treatment, samples were ds-cDNA amplified. Spermatozoal RNA sample did not have genomic DNA or somatic cell contamination and transcripts were consistently

amplified via PCR. A sample that contained spermatozoal RNA from 9 bulls was submitted for Illumina RNA-Seq.

II. RNA-Seq via Illumna sequencing

Reads from RNA-Seq were analyzed with Tophat and Cufflinks that found 6,166 expressed transcripts that were further categorized via gene ontology analysis. Nine transcripts were chosen as qPCR candidates to validate expression levels from RNA-Seq data and a high correlation was found between FPKM expression levels (RNA-Seq) and copy number (qPCR).

B. Spermatozoal transcript profiles differ between high and low fertility sires; Journal of Dairy Science in prep **(Chapter 3)**

I. RNA-Seq analysis of higher and lower fertility samples

With the spermatozoal RNA isolation method (#2) and sample processing as Chapter 2, spermatozoal RNA from higher fertility (-1.8 to 3.5 CR; 4 bulls) and lower fertility (-2.9 to -0.4 CR; 4 bulls) was sequenced with RNA-Seq. Cuffdiff analysis was conducted and 36 transcripts were differentially expressed between the two samples while over 100 and 400 transcripts were found to be expressed at higher levels in higher and lower fertility populations respectively when using differential ratio analysis. Once again, qPCR validation of FPKM expression levels was highly correlated. A handful of transcripts were quantified on qPCR across 9 individual sires with varying fertility scores to investigate possible correlations with fertility. Some negative correlations with fertility were found (*COX7C*, *PRM2*, and *TNP1*) while some transcripts showed no correlation with fertility (*PRM1* and *PSMA6*).

## CHAPTER 2

PUBLISHED IN BIOLOGY OF REPRODUCTION (JANUARY 2013)

# Cryopreserved Bovine Spermatozoal Transcript Profile as Revealed by RNA-Seq

Card C[1*], Anderson EJ[1*], Zamberlan S[1], Krieger KE[2], Kaproth M[2], Sartini BL[1,3]

*these authors contributed equally to the writing of this manuscript

[1]Department of Fisheries, Animal and Veterinary Science, University of Rhode Island, Kingston, Rhode Island 02881. [2]Genex Cooperative Inc., Shawano, WI 54166.

[3]Corresponding author: BL Sartini, 179 Center for Biotechnology and Life Sciences, 120 Flagg Road, University of Rhode Island, Kingston, RI 02886, Phone: 401-874-2667, FAX: 401-874-7575, Email: blsartini@uri.edu.

**Short Title: Bovine Spermatozoal Transcript Profile**

**Summary Sentence: The bovine spermatozoal transcript profile contains degraded and full-length mRNAs.**

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

**Abbreviations:** mRNA = messenger RNA; FPKM = Fragments Per per Kilobase of transcript per Million mapped reads; RNA-Seq = ribonucleic acid sequencing; rRNA = ribosomal ribonucleic acid; CR = conception rate; qPCR = quantitative polymerase chain reaction; UTR = untranslated region; MF = Molecular Function; BP = Biological Process; CC = Cellular Component

**Grant Support:** 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.

**Abstract**

Ejaculated bovine spermatozoa retain a pool of RNAs that may have a function in early embryogenesis and be used as predictors of male fertility. The bovine spermatozoal transcript profile remains incomplete because previous studies have relied on hybridization-based techniques, which evaluate a limited pool of transcripts and cannot identify full-length transcripts. The goal of this study was to sequence the complete cryopreserved bovine spermatozoal transcript profile using Illumina RNA-Seq. Spermatozoal RNA was pooled from nine bulls with conception rate (CR) scores ranging from -2.9 to 3.5 and confirmed to exclude genomic DNA and somatic cell mRNA. After selective amplification of polyA$^+$ RNA and high-throughput-sequencing, 6,166 transcripts were identified via alignment to the bovine genome (UMD 3.1/bosTau6). RNA-Seq transcript levels (n=9) were highly correlated with qPCR copy number ($r^2$=0.9747). The bovine spermatozoal transcript profile is a heterogeneous population of degraded and full-length predominantly nuclear-encoded mRNAs. Highly abundant spermatozoal transcripts included *PRM1*, *HMGB4* and mitochondrial-encoded transcripts. Full-length transcripts comprised 66% of the top 368 transcripts (FPKM>100) and amplification of the full-length transcript or 5' and 3' ends was confirmed for selected transcripts. In addition to the identification of transcripts not previously reported in spermatozoa, several known spermatozoal transcripts from various species were also found. Gene ontology analysis of the FPKM>100 spermatozoal transcripts revealed that translation was the most predominant biological process represented. This is the first report of the spermatozoal transcript profile in any species using high-throughput sequencing, supporting the presence of mRNA in spermatozoa for further studies.

**Introduction**

In addition to delivering the paternal genome to the oocyte at fertilization, ejaculated spermatozoa retain a pool of RNAs, containing mRNAs, rRNAs and short non-coding RNAs [1-4]. Spermatozoal antisense RNAs can epigenetically regulate early embryonic development and have a structural role in maintaining histone-bound spermatozoa chromosomal regions [3-6]. Although the complete spermatozoal mRNA profile is not known, spermatozoa contain at least 3,000-7,000 mRNAs with predominantly short fragments, probably indicative of a predominance of degraded RNA [7-9]. Individual spermatozoal transcripts that have been identified include mRNAs for ribosomal proteins, mitochondrial proteins, protamines, and proteins involved in signal transduction and cell proliferation [7-12]. The hypothesized function of the spermatozoal transcripts in transcriptionally-silent spermatozoa is currently unknown although spermatozoa-derived mRNAs, including *PRM1*, *PRM2*, *PSG-1*, *CLU, HLA-E, DBY* and *PLCZ1,* can be detected in embryos post-fertilization suggesting a role for spermatozoal mRNAs in the zygote [13-18]. However, only translation of *PLCZ1* has been demonstrated in embryos and many of these spermatozoal transcripts are rapidly degraded in the embryo rendering them non-functional [15-18]. Some spermatozoal transcripts may be translated in the mitochondria during capacitation [19]. Additionally, the diagnostic potential of the total spermatozoal RNA population as a snapshot of spermatogenic gene expression is emerging [20]. For example, perturbation of the ubiquitin-proteosome pathway during spermatogenesis can be detected in spermatozoal RNA [21]. Individual transcripts are stably regulated within and between individual males making this a promising area for male fertility assay development [22, 23].

The bovine spermatozoal transcript profile remains incomplete because previous studies have relied on hybridization-based techniques, which evaluate a limited pool of transcripts and do not provide information about full-length transcripts [7, 9, 10, 24, 25]. In contrast, RNA-Sequencing (RNA-Seq), based on high-throughput sequencing technology, is revolutionizing our understanding of transcriptomics by enabling sequencing of ~~the~~ complete transcript profiles, including full-length mRNAs and identifying novel splicing junctions and exons [26, 27]. Also unique to this direct sequencing, absolute quantification of a broad range of expression levels across transcripts can be obtained. High-throughput sequencing of the total RNA in human spermatozoa has focused on rRNA and small non-coding RNA populations but the complete mRNA profile has not been reported [2, 4].

We hypothesize that the transcript profile of cryopreserved bovine spermatozoa can be directly sequenced using RNA-Seq. Over 6000 spermatozoal transcripts were sequenced with this approach and a heterogeneous population of degraded and full-length mRNAs was identified. Previously reported spermatozoal transcripts were confirmed while a number of transcripts not previously found in spermatozoa of any species have also been identified including *HMGB4*, *GTSF1*, and *CKS2*. This is the first study to date to utilize RNA-Seq to sequence the spermatozoal mRNA population and report full-length transcripts for any species.

**Materials and Methods**

Spermatozoa Samples

Cryopreserved spermatozoa was obtained from twenty-one Holstein bulls from Genex Cooperative Inc. (Shawano, WI). Semen from 9 bulls (-2.9 to 3.5 CR) was used

for RNA-Seq, qPCR validation, and PCR amplification of the 5' and 3' exons. For analysis of transcript variation among individual bulls, 9 different bulls were used. Sperm RNA from 3 additional bulls was pooled and used for PCR amplification of full-length transcripts. Two straws per bull were thawed in a 37°C water bath for one minute and then washed twice in 4 mL PBS (10 minutes at 600 *x g*). The resulting spermatozoa pellet was subsequently used for RNA isolation.

RNA Isolation

Bull testis RNA was isolated with TRIzol (Sigma-Aldrich; St. Louis, MO). Sperm RNA isolation was conducted using the TRIzol method reported by Das et al., 2010 [8] with slight modifications. Spermatozoa pellets were added to 1 mL TRIzol supplemented with glycogen (15 µg/ml). Samples were then lysed through a 26 gauge needle 20 times and incubated for 30 minutes at room temperature. Chloroform was added to the samples followed by a 10 minute incubation at room temperature. For phase separation, samples were centrifuged at 12000 *x g* for 15 minutes at 4ºC. The top layer (RNA) was removed and added to 500 µL ice cold isopropanol and incubated on ice for 10 minutes followed by centrifugation 12000 *x g* for 10 minutes at 4ºC to precipitate the RNA. RNA pellets were washed with 1 mL 75% ethanol and air dried, followed by resuspension in nuclease free water. RNA samples were treated with DNAse using the RNA Cleanup protocol from the RNeasy Mini Kit (Qiagen, Valencia, CA). RNA concentrations were measured using the NanoDrop UV/Vis Spectrometer (Thermo Scientific; Wilmington, DE) and RNA samples stored at -80º C until used for subsequent analysis.

RT-PCR

For amplification of full-length transcripts, spermatozoal RNA (1 µg) was reverse transcribed using the Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA) with 2.5 µM Oligo-dT primers, 0.5 mM dNTPs, 0.25 X First Strand Buffer, 0.005 mM DTT, 40 U RNaseOut, and 200 U Superscript III. After the incubations at $50^{O}$ C for 45 minutes and $70^{O}$ C for 15 minutes, 2 U RNase H was added followed by an additional incubation at $37^{O}$ C for 20 minutes. For all primer pairs, cDNA was added to a PCR reaction mixture containing 1X reaction buffer, 1.5 mM $MgCl_2$, 10 mM dNTPs, 2.5 µmol forward and reverse primers and 2.5 U Taq polymerase (NEB; Ipswich, MA). To amplify the full-length transcript, forward primers were located in the first exon and reverse primers were located in the last exon (Table 1). PCR 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 and a final extension at 72°C for 10 min. Negative controls containing no template cDNA and no enzyme were run in parallel to ensure gene specific amplification. The PCR products were separated by 2.0% agarose gel electrophoresis, gel purified (Qiagen Gel Extraction kit; Valencia, CA) and both strands sequenced (URI Genomics Center, Kingston, RI). Amplicon sequence identity was confirmed with NCBI BLAST.

Double-Stranded cDNA (ds-cDNA) Synthesis and Amplification

Due to low RNA yields typical of spermatozoal RNA isolations, the spermatozoal RNA was converted to ds-cDNA and amplified for RNA-Seq analysis and qPCR

validation (SMARTer Pico PCR cDNA Synthesis Kit; Clontech, Mountain View, CA). Due to the varying amounts of RNA extracted from each bull, the amount of RNA added for amplification was normalized to the sample with the lowest concentration to ensure equal representation of the nine bulls in the pooled sample. The amplification protocol enriches the full-length mRNA population with a modified oligo(dT) primer. Amplification cycles were optimized to 26 cycles following the protocol instructions to insure amplification the linear phase, maintaining gene representation of the original RNA pool. To verify the quality of ds-cDNA, an aliquot of the sample was run on the Agilent Bioanalyzer 2100 (Agilent Technologies; Santa Clara, CA). Following ds-cDNA conversion and amplification, 5 µg of spermatozoal ds-cDNA was submitted for Illumina sequencing.

The remaining spermatozoal ds-cDNA was used for qPCR validation post-sequencing, validating the lack of genomic DNA and somatic cell RNA in the spermatozoal RNA. A separate pool of ds-cDNA was generated from three additional bulls to amplify the 5' and 3' ends of selected transcripts. Finally, spermatozoal RNA from individual bulls was individually amplified to assess transcript presence among individuals. All PCR reactions were run with spermatozoal ds-cDNA except for the spermatozoal ds-cDNA sample that was spiked with genomic DNA (1 µg) isolated from the bull testis tissue (Lane G in Figure 1; Qiagen DNA Blood and Tissue kit; Valencia, CA). All PCR reactions were conducted with intron spanning primers (Table 1) as described above.

RNA-Seq and Analysis

Paired-end 100 bp reads from spermatozoal ds-cDNA were generated using the Illumina HiSeq 2000 (Otogenetics; Norcross, GA). Sequence analysis was conducted with Galaxy [28-30]. Trimming the adapter AGATCGGAAGAGC removed 14.29% (2,659,330 reads) from file 1 and 1.14% (211,176 reads) from file 2. Adaptor only reads, short sequence reads (15 nt minimum) and reads with unknown N bases were discarded during adapter trimming. Concatamers formed from amplification of the SMARTer II A Oligonucleotide (AAGCAGTGGTATCAACGCAGAGTAA) were found in 41.48% (7,702,931 reads) for file 1 and 47.17% (8,760,365 reads) for file 2 and were removed prior to further analysis. Alignment to the reference genome (UMD 3.1/bosTau6) was conducted using Tophat, which uses Bowtie for alignment [31-32]. A maximum of two mismatches were allowed during alignment. RSeQC was used to generate read and post-alignment summary statistics [33]. Levels of individual transcripts are expressed in Fragments Per Kilobase of exon per Million fragments mapped (FPKM) and were obtained using Cufflinks [30]. Quantification of full-length transcripts was conducted by manually visualizing the read mapping for individual transcripts to the bovine genome (UMD 3.1/bosTau6) in the UCSC Genome Browser (http://genome.ucsc.edu/). Reads were archived in the NCBI SRA055325 (http://www.ncbi.nlm.nih.gov/sra).

qPCR

Quantitative PCR was used to validate RNA-Seq expression levels of the cryopreserved spermatozoal ds-cDNA. Nine transcripts were chosen that represented a range of FPKM values (9.41 to 20,667). A standard curve was generated by diluting

DNA for each transcript into 7 concentrations ranging from $1x10^9$ copies to $1x10^3$ copies/ul. qPCR was performed with spermatozoal ds-cDNA and standard curves using the Brilliant II SYBR Green QPCR Master Mix Kit (Stratagene; Santa Clara, CA). All qPCR samples included negative template controls and were run in duplicate on the Stratagene Mx3005 instrument at the Genome Sequencing Center at the University of Rhode Island. Amplicon sequence identity was confirmed with NCBI BLAST.

Gene Ontology Analysis

Gene ontology analysis was conducted with the DAVID Bioinformatic Database (http://david.abcc.ncifcrf.gov/) using the three Gene Ontology Term categories: Biological Process, Molecular Function, and Cellular Component. Transcripts were analyzed in two different populations: FPKM>0 and FPKM>100.

**Results**

***Bovine spermatozoal RNA purity***

Using the Trizol method, the total amount of RNA isolated from two spermatozoa straws from an individual bull resulted in an average of 31 fg RNA per spermatozoa. Bioanalyzer analysis of the spermatozoa RNA population shows a peak of smaller RNAs and a lack of 18S and 28S ribosomal RNA peaks present in testis RNA (Figure 1A). Genomic DNA was also not detected in the isolated bovine spermatozoal RNA compared to a sample spiked with genomic DNA (Figure 1B). The spermatozoal RNA was free of leukocytes, testicular germ cells and epithelial cells as demonstrated by the lack of *C-KIT*, *CD45*, and *CDH1* amplification respectively (Figure 1C).

***Illumina Sequencing***

High-throughput sequencing of the bovine spermatozoal RNA resulted in 18,570,350 x 2 paired-end 100-bp reads. After removal of concatamers, a total of 2,538,688 reads (14.25%) of the total population mapped to the bovine genome with 79.84% of the aligned reads being uniquely mapped to a single transcript. Reads aligned specifically to coding exons (324,600 reads), 5'UTRs (39,758 reads), 3'UTRs (40,057 reads), and 2,274 reads contained poly(A) sequences. Exon-exon junctions (157,717 reads over 17,285 junctions) were covered and 100,929 of those reads (64.21%) mapped to 9,003 annotated junctions while 56,248 (35.79%) reads mapped to 8,282 novel/partial junctions. All junctions were supported by at least two reads. Also, 144,432 intronic reads were found, several of which may represent novel exons.

***Cryopreserved Bovine Spermatozoal Transcript Profile***

A total of 6,166 transcripts were identified in spermatozoal RNA with a FPKM>0 (Fragments Per Kilobase of transcript per Million reads mapped). The qPCR expression values showed a high correlation with FPKM values ($r^2$ =0.9747; Figure 2). The bovine spermatozoal transcript profile contains predominantly nuclear-encoded mRNAs including 33 mitochondrial-encoded rRNAs and mRNAs representing 0.5% of the spermatozoal transcript profile. Many of these mitochondrial transcripts were highly abundant, with 32 of 33 in the top 100 transcripts ranked by FPKM. The top 10 transcripts based on FPKM, excluding the mitochondrial RNAs, are listed in Table 2.

A heterogeneous population of degraded and full-length transcripts was identified. Degraded transcripts (lacking reads mapping to all exons) were more prevalent below

FPKM = 100. Due to this observation, all transcripts with FPKM>100 (368 transcripts) were analyzed individually for reads mapping to each exon to be considered a full-length transcript. In the FPKM>100 population, 66% of the transcripts had reads aligned to all exons, including amplification of the 5' and 3' exons, potentially indicating the presence of full-length transcripts in the spermatozoal RNA population (Supplementary Table 1). Some of these full-length transcripts also included intronic reads that potentially represent novel exons. Retention of the 5'and 3' exons for *PLCZ1, CRISP2,* and *GSTM3* were validated while many transcripts with FPKM<100 did not retain the 5' exon, including *DDX3Y* (Figure 3A). The presence of full-length transcripts for *GSTM3* and *GTSF1* was confirmed by RT-PCR amplification from the first to last exon (Figure 3B). A preliminary survey of the bovine spermatozoal transcript profile for previously reported spermatozoal RNA candidates identified several transcripts in bovine, human, porcine and mouse (Table 3). These transcripts represented a wide range of FPKM levels, and nine of these transcripts retained the 5' and 3' ends, potentially indicating that these transcripts are also full-length (Table 3).

A number of additional full-length bovine spermatozoal transcripts were identified that have not been previously reported in spermatozoal RNA, including *HMGB4*, *PSMA6*, *GTSF1*, and *CKS2* (Table 4). The presence of select spermatozoal transcripts, *CKS2, EEF1G, EIF1, GTSF1, PRM1*, varied among individual bulls (Figure 4). Spermatozoa RNA from these bulls was not included in the pool submitted for RNA-Seq.

***Gene Ontology Analysis***

For gene ontology analysis, spermatozoal transcripts were analyzed in two different populations: FPKM>0 (n= 6,166) and FPKM>100 (n= 368). Transcripts were classified into the following ontological categories: Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF) and the top ten categories for each are shown in Table 5. In the total spermatozoal transcript population (FPKM>0), 367 BP, 142 CC, and 161 MF categories were found. It is important to note that an individual transcript can be represented in multiple categories. The top BP categories included translation (GO: 0006412; 264 transcripts) and proteolysis (GO: 0051603; 241 transcripts). Because a majority of full-length transcripts were found in the FPKM>100 population, we also analyzed this population separately. Translation remained the most predominant BP represented within this population (55 transcripts). Within the translation category, 38 of the 55 transcripts encoded for ribosomal proteins and the remaining transcripts included eukaryotic translation initiation factors (*EIF1* and *EIF5),* eukaryotic translation elongation factors *(EEF1A1* and *EEF1γ),* polyubiquitin and unknown transcripts. Twenty-four of these ribosomal transcripts were full-length (all exons mapped), as well as *EEF1A1, EEF1γ* and polyubiquitin.

**Discussion**

Here, we report the first cryopreserved bovine spermatozoal transcript profile using RNA-Seq, which includes degraded and full-length nuclear-encoded transcripts and mitochondrial-encoded RNA. The dynamic range of RNA-Seq allows for accurate

identification and quantification of transcripts present at very low and high levels as well as the discovery of more transcripts, novel splicing junctions and novel exons than reported in previous microarray studies [7, 9, 10]. In addition to the identification of transcripts not previously reported in spermatozoal RNA, several known spermatozoal transcripts from a number of different species were also found. Gene ontology analysis of the highly abundant spermatozoal transcripts (FPKM>100) revealed that translation was the most predominant biological process represented. The presence of full-length transcripts in transcriptionally-silent spermatozoa suggests that these transcripts could be translated after spermatogenesis is complete, potentially contributing to capacitation and early embryogenesis [1, 3].

Spermatozoal RNA isolation procedures have been developed to maximize yield and ensure elimination of somatic cell RNA. The total amount of cryopreserved bovine spermatozoal RNA isolated in this study (31 fg RNA per spermatozoa) was comparable to the RNA content previously reported in bovine (10-140 fg), human (12.5 fg), rat (100 fg), porcine (5 fg), and equine (20 fg) spermatozoa [reviewed in 1, 8].

In this study, RNA was isolated from the whole cryopreserved semen straw, after a wash to remove the cryoprotectant, without the removal of non-motile spermatozoa. Using the entire spermatozoa population is representative of the natural transcript variation present across a range of fertility scores for bulls used in artificial insemination and is consistent with the approach used in other studies [12, 21, 24, 34].

The focus of this study was to enrich for and sequence the polyA$^+$ transcripts present in transcriptionally-silent spermatozoa. The mitochondrial-encoded rRNAs and mRNAs sequenced in this population were some of the most abundant transcripts

although these mitochondrial RNAs represented only 0.5% of the total transcripts. Mitochondrial rRNAs and mRNAs have been previously amplified in spermatozoa [10, 19] and the presence of these transcripts is likely due to intact mitochondria present during the RNA isolation procedure and the high mitochondrial activity of spermatozoa. Poly(A-) transcripts and microRNAs were not evaluated in this study but probably present in the total bovine sperm RNA population [4].

Using RNA-Seq, we identified several full-length transcripts in the bovine cryopreserved spermatozoal transcript profile. While some of these transcripts were previously reported in spermatozoa, the presence of full-length transcripts could not be determined from previous microarray studies. The most abundant full-length transcript, *PRM1*, has been reported in spermatozoa from other species as well, including humans and porcine [7, 13, 20, 35]. The high level of *PRM1* is probably due to retention of this transcript in elongating spermatids during the later stages of spermatogenesis. A function for *PRM1* after spermatozoa leave the testis is doubtful as *Prm1* transcripts are rapidly degraded in the mouse embryo [15, 16]. Other transcripts are delivered to the oocyte after fertilization, including the Y chromosome-linked *DBY* and *RPS4Y,* were not identified as full-length transcripts in this study, therefore, a functional role in embryogenesis for these transcripts is also unlikely [17].

Polyubiquitin is also an abundant full-length transcript in bovine spermatozoa. The ubiquitin system has several functions during spermiogenesis and fertilization, including: histone removal, removal of damaged epididymal spermatozoa, and aiding in zona penetration [36, 37]. Disruption of the ubiquitin-proteosome pathway during spermatogenesis is characteristic of teratozoospermic males and can be detected in

human sperm RNA [22]. Spermatozoa-derived ubiquitin RNAs may also have a role in directing the degradation of paternal mitochondrial RNAs, ensuring exclusive maternal mitochondrial DNA inheritance [36]. Further investigation of a role for spermatozoal-derived polyubiqutin mRNA pre- and post-fertilization is warranted.

Previously reported spermatozoal transcripts involved in capacitation and fertilization were also identified as full-length, including: *PLCZ1, CRISP2* and *CLGN1*. *PLCZ1*, a well-characterized activator of the calcium wave after fertilization, is translated in the oocyte and injections of *PLCZ1* RNA into the oocyte are also sufficient for function [18]. *PLCZ1* is present at lower amounts (FPKM= 41.3) in the bovine spermatozoa transcript profile demonstrating that functional transcripts may not be the most abundant transcripts in this population. The presence of full-length *CRISP2* could be indicative of potential translation at fertilization as *CRISP2* is one of the spermatozoal proteins involved in oocyte binding [38]. The *CLGN1* protein is necessary for heterodimerization of fertilization proteins [39, 40]. The presence of spermatozoal mRNA for critical fertilization proteins may be necessary to ensure appropriate function.

A number of previously unreported spermatozoal transcripts are full-length and abundant in the bovine spermatozoal transcript profile including *HMGB4*, *PSMA6*, *GTSF1*, and *CKS2* although a role of transcripts from spermatozoal-derived mRNAs is speculative. *HMGB4* is found at the basal pole of elongating spermatids and is a transcriptional repressor [41]. *PSMA6* is an alpha subunit of proteasomes; inhibition of spermatozoal proteosomes blocks fertilization by preventing spermatozoa penetration of the zona pellucida [42]. *GTSF1* is critical for the suppression of retrotransposons in the male germ cells, as well as causing meiotic arrest in knockout mice [43]. *CKS2* is critical

in early embryonic development, where it controls cell proliferation [44]. In knockout studies of *CKS2* and *CKS1,* embryos arrest development before the morula stage due to cyclin B1 downregulation [44].

A predominant function of the bovine spermatozoal transcripts with FPKM>100 is translation and includes abundant transcripts for ribosomal proteins, polyubiquitin (discussed above), eukaryotic translation initiation factors (*EIF1 and EIF5),* and eukaryotic translation elongation factors *(EEF1A1 and EEF1γ). EIF1A1* is present in human spermatozoa [24] but *EIF1, EEF1γ* and *EIF5* have not been previously reported in any species. The translation elongation factors *EEF1A1* and *EEF1γ* were full-length in this study therefore a role for these transcripts in the early stage embryo is an interesting area for further investigation.

One-third of the transcripts with FPKM>100 were degraded (all exons were not mapped). A predominance of degraded transcripts was also found in the FPKM<100 transcript population although this was not quantified. A degraded RNA population is characteristic of the spermatozoal RNA populations isolated in previous studies [7, 8] and large subset of the spermatozoal mRNAs are probably remnants from gene expression during spermatogenesis and do not have a function. The relatively higher levels of most of the full-length transcripts is probably not due to a 3' end bias, which can occur with RNA-Seq, due to the RNA amplification protocol that selectively amplified full-length transcripts [45]. Although full-length transcripts were identified, the proportion of degraded and full-length transcripts for an individual transcript could not be distinguished and the FPKM values probably represent a sum of the full-length and fragmented exons for each transcript making the levels of intact transcripts probably lower than these

reported values. The presence of degraded mRNAs and full-length mRNAs are not necessarily mutually exclusive events and functional transcripts could be present in a heterogeneous population.

The goal of this present study was to identify a full-range of poly($A^+$) mRNAs present in bovine spermatozoa to identify candidates for further investigation. Spermatozoa RNA from individual bulls with a wide range of fertility scores was pooled for RNA-Seq, and subsequent validation, and biological replicates were not conducted. The presence of select spermatozoal transcripts, *CKS2, EEF1G, EIF1, GTSF1, PRM1*, varied among spermatozoal RNA isolated from a separate population of bulls demonstrating that the spermatozoal transcript profile is probably different for each individual. Additionally, the pool of mRNA from this spermatozoa population contains several previously reported transcripts therefore the likelihood that the identified transcripts are only present in this population of bulls is low but additional transcripts may be identified in other individuals.

The diagnostic potential of the total spermatozoal RNA population (degraded and full-length transcripts) is emerging. Individual transcripts are stably regulated within and between individual males and perturbation of the ubiquitin-proteosome pathway during spermatogenesis could be detected in the spermatozoal RNA [22, 23]. The amount of specific transcripts, including *PRM1, PRM2, CRISP2*, *CCT8*, *PEBP1* and *CD36,* have also been correlated to fertility in humans and bulls [10-12, 20, 46]. These transcripts are full-length in this bovine spermatozoal transcript profile, so prediction of fertility for some of these transcripts may be due to a functional role (for example *CRISP2*) and not just a representation of transcription during spermatogenesis (for example, *PRM1* and

*PRM2*). If the degraded mRNA population is stably regulated, this population can also be used to as a diagnostic tool. Spermatozoal transcript populations also vary with motility, morphology, DNA integrity and seasons [47-51]. The spermatozoal transcript profile reported here was sequenced from a pool of bulls that represent a normal range of fertility scores. While selected transcripts demonstrated individual variance among bulls, further quantitative analysis in a much larger population will better assess the level of individual bull variation and a correlation of transcript levels with fertility scores.

This is the first report of the spermatozoal transcript profile in any species using high-throughput sequencing, supporting the presence of mRNA in spermatozoa. Further studies of the spermatozoal mRNA candidates identified will contribute to our knowledge of the function of spermatozoal mRNA and expand our approaches to assay male fertility.

**References**

[1] Hamatani T. Human spermatozoal RNAs. Fertil Steril 2012; 97: 275-281.

[2] Johnson GD, Sendler E, Lalancette C, Hauser R, Diamond MP, Krawetz SA. Cleavage of rRNA ensures translational cessation in sperm at fertilization. Mol Hum Reprod 2011; 17: 721-726.

[3] Lalancette C, Miller D, Li Y, Krawetz SA. Paternal contributions: new functional insights for spermatozoal RNA. J Cell Biochem 2008; 104: 1570-1579.

[4] Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, Diamond MP. A survey of small RNAs in human sperm. Hum Reprod 2011; 26: 3401-3412.

[5] Jenkins TG and Carrell DT. The sperm epigenome and potential implications for the developing embryo. Reproduction 2012; 143: 727-734.

[6] Liu W, Pang RTK, Chiu PCN, Wong BPC, Lao K, Lee K, Yeung WSB. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. PNAS 2012; 109: 490-494.

[7] Gilbert I, Bissonnette N, Boissonneault G, Vallee M, Robert C. A molecular analysis of the population of mRNA in bovine spermatozoa. Reprod 2007; 133: 1073-1086.

[8] Das PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, Varner DD, Chowdhary BP, Raudsepp T. Total RNA isolation from stallion sperm and testis biopsies. Theriogenology 2010; 74: 1099-1106.

[9] Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA. Spermatozoal RNA profiles of normal fertile men. Lancet 2002; 360: 772-777.

[10] Lalancette C, Thibault C, Bachand I, Caron N, Bissonnette N. Transcriptome analysis of bull semen with extreme nonreturn rate: use of suppression-subtractive hybridization to identify functional markers for fertility. Biol Reprod 2008; 78: 618-635.

[11] Avendano C, Franchi A, Jones E, Oehninger S. Pregnancy-specific B-I-glycoprotein I and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. Hum Reprod 2009; 24: 270-277.

[12] Arangasamy A, Kasimanickam VR, DeJarnette JM, Kasimanickam RK. Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. Theriogenology 2011; 76: 570-577.

[13] Kempisty B, Antosik P, Bukowska D, Jackowska M, Lianeri M, Jaskowski KM, Jagodzinski PP. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reprod Fertil Dev 2008; 20: 513-518.

[14] Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. Delivering spermatozoan RNA to the oocyte. Nature 2004; 429: 154.

[15] Hayashi S, Yang J. Christenson L, Yanagimachi R, Hecht NB. Mouse preimplantation embryos developed from oocytes injected with round spermatids or

spermatozoa have similar but distinct patterns of early messenger RNA expression. Biol Reprod 2003; 69: 1170-1176.

[16] Depa-Martynow M, Kempisty B, Jagodzinski P, Pawelczyk L, Jedrzejczak P. Impact of protamine transcripts and their proteins on the quality and fertilization ability of sperm and the development of preimplantation embryos. Reprod Biol 2012; 12: 57-72.

[17] Yao C, Wang Z, Zhou Y, Xu W, Li Q, Ma D, Wang L, Qiao Z. A study of Y chromosome gene mRNA in human ejaculated spermatozoa. Mol Reprod Dev 2010; 77: 158-166.

[18] Swann K, Saunders CM, Rogers NT, Lai FA. PLCζ (zeta): A sperm protein that triggers $Ca^{2+}$ oscillations and egg activation in mammals. Sem Cell Dev Biol 2006; 17: 264-273.

[19] Gur Y and Breitbart H. Mammalian sperm translate nuclear-encoded protiens by mitochondrial-type ribosomes. Genes Dev 2006; 20: 411-416.

[20] Feugang JM, Rodriguez-Osorio N, Kaya A, Wang H, Page G, Ostermeier GC, Topper EK, Memili E. Transcriptome analysis of bull spermatozoa: implications for male fertility. Reprod Biomed Online 2010; 21: 312-324.

[21] Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, Rawe VY, Quintana S, Diamond MP, Strader LF, Krawetz SA. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. Hum Mol Genet 2007; 16: 763-773.

[22] Ostermeier GC, Goodrich RJ, Diamond MP, Dix DJ, Krawetz SA. Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. Fert Ster 2005; 26: 1687-1694.

[23] Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, Krawetz SA. Identification of human sperm transcripts as candidate markers of male fertility. J Mol Med 2009; 87: 735-748.

[24] Dadoune JP, Pawlak A, Alfonsi MF, Siffroi JP. Identification of transcripts by macroarrays, RT-PCR and in situ hybridization in human ejaculate spermatozoa. Mol Hum Reprod 2005; 11: 133-140.

[25] Zhao Y, Li Q, Yao C, Wang Z, Zhou Y, Wang Y, Liu L, Wang Y, Wang L, Qiao Z. Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. Hum Reprod 2006; 21: 1583-1590.

[26] Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 2009; 10:57-63.

[27] Werner T. Next generation sequencing allows deeper analysis and understanding of genomes and transcriptomes including aspects to fertility. Reprod Fertil Dev 2011; 23(1): 75-80.

[28] Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25; 11(8):R86.

[29] Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Molec Biol 2010; Chapter 19: Unit 19.10.1-21.

[30] Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. Galaxy: a platform for interactive large-scale genome analysis. Genome Res 2005; 15(10):1451-5.

[31] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012; 7(3):562-578.

[32] Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009; 10(3): R25.

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

[34] Wang H, Zhou Z, Xu M, Li J, Xiao J, Xu Z, Sha J. A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. J Mol Med 2004; 317-324.

[35] Hecht N, Cavalcanti MC, Nayudu P, Behr R, Reichenback M, Weidner W, Steger K. Protamine-1 represents a sperm specific gene transcript: a study in Callithrix jacchus and Bos taurus. Andrologia 2011; 43(3):167-173.

[36] Sutovsky P. Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization and sperm quality control: killing three birds with on stone. Microsc Res Tech 2003; 61: 88-102.

[37] Yi YJ, Zimmerman SW, Manandhar G, Odhiambo JF, Kennedy C, Jonakova V, Manaskova-Postlerova P. Ubiquitin-activating enzyme (UBA1) is required for sperm capacitation, acrosomal exocytosis and sperm-egg coat penetration during porcine fertilization. Int J Androl 2012; 35(2): 196–210.

[38] Busso D, Goldweic NM, Hayashi M, Kasahara M, Cuasnicu PS. Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion. Biol Reprod 2007; 76: 701-708.

[39] Ikawa M, Nakanishi T, Yamada S, Wada I, Kominami K, Tanaka H, Nozaki M, Nishimune Y, Okabe M. Calmegin is required for fertilin alpha/beta heterodimerization and sperm fertility. Dev Biol 2001; 240: 254-261.

[40] Yamagata K, Nakanishi T, Ikawa M, Yamaguchi R, Moss SB, Okabe M. Sperm from calmegin-deficient mouse have normal abilities for binding and fusion to the egg plasma membrane. Dev Biol 2002; 250:348-357.

[41] Catena R, Escoffier E, Caron C, Khochbin S, Martianov I, Davidson I. HMGB4, a novel member of the HMGB family, is preferentially expressed in the mouse testis and localizes to the basal pole of elongating spermatids. Biol Reprod 2009; 80: 358-366.

[42] Zimmerman SW, Manandhar G, Yi YJ, Gupta SK, Sutovsky M, Odhiambo JF, Powell MD, Miller DJ, Sutovsky P. Sperm proteasomes degrade sperm receptor on the egg zona pellucida during mammalian fertilization. PLOS one 2011; 6(2): e17256.

[43] Yoshimura T, Toyoda S, Kuramochi-Miyagawa S, Miyazaki T, Miyazaki S, Tashiro F, Yamato E, Nakano T, Miyazaki J. *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 testes. Dev Biol 2009; 335: 216-227.

[44] Martinsson-Ahlzen HS, Liberal V, Grunenfelder B, Chaves SR, Spruck CH, Reed SI. Cyclin-dependent kinase-associated proteins *Cks1* and *Cks2* are essential during early

embryogenesis and for cell cycle progression in somatic cells. Mol Cell Biol 2008; 28: 5698-5709.

[45] Costa V, Angelini C, De Feis I, Ciccodicola A. Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol. 2010; 853916. Epub 2010 June 27.

[46] Steger K, Wilhelm J, Konrad L, Stalf T, Greb R, Deimer T, Kliesch S, Bergmann M, Weidner W. Both protamine-1 to protamine-2 mRNA ratio and Bcl2 mRNA content in testicular spermatids and ejaculated spermatozoa discriminate between fertile and infertile men. Hum Reprod 2008; 23: 11-16.

[47] Aoki VW, Emery BR, Liu L, Carrell DT. Protamine levels vary between individual sperm cells of infertile human males and correlate with viability and DNA integrity. J Androl 2006; 27: 890-898.

[48] Lambard S, Galeraud-Denis I, Martin G, Levy R, Chocat A, Carreau S. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors:relationship to sperm motility and capacitation. Mol Hum Reprod 2004; 10: 535-541.

[49] Roudebush WE, Massey JB, Zhu J. Morphologically normal sperm have significantly greater total-RNA content than abnormal sperm. Int Congr Ser 2004; 193–196.

[50] Yang CC, Lin YS, Hsu CC, Tsai MH, Wu SC, Cheng WTK. Seasonal effect of sperm messenger RNA profile of domestic swine. Anim Reprod Sci 2010; 119:76-84.

[51] Bissonnette N, Levesque-Sergerie JP, Thibault C, Boissonneault G. Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality. Reprod 2009; 138:65-80.

## FIGURES

Figure 1. Purity of bovine cryopreserved spermatozoal RNA was confirmed by lack of somatic cell RNAs and genomic amplification. (A) Bioanalyzer analysis of testis RNA and spermatozoal RNA prior to amplification. (B) Cell-specific transcripts for testicular germ cells (*C-KIT)*, leukocytes (*CD45)* and epithelial cells (*CDH1)* did not amplify in the spermatozoal RNA (Lane S). M= 100 bp DNA marker, T = testis RNA positive control and N = negative control that does not include cDNA template. (C) The spermatozoal RNA (Lane S) does not contain genomic DNA compared to amplification of genomic *EIF1* in spermatozoal cDNA spiked with genomic DNA (Lane G). N = negative control that does include cDNA template.

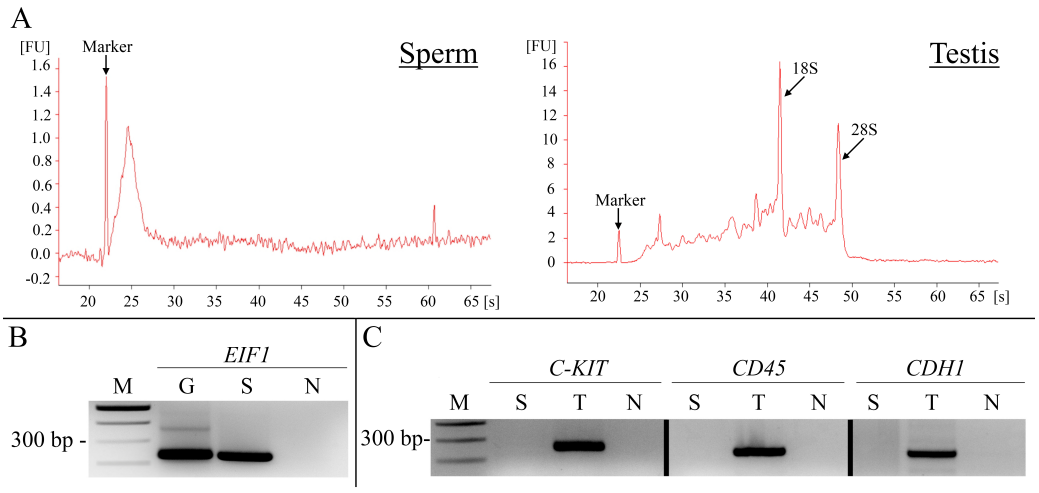


Figure 2. Correlation of qPCR transcript copy number and RNA-Seq FPKM based on nine transcripts. Axes are base 10 log scale.

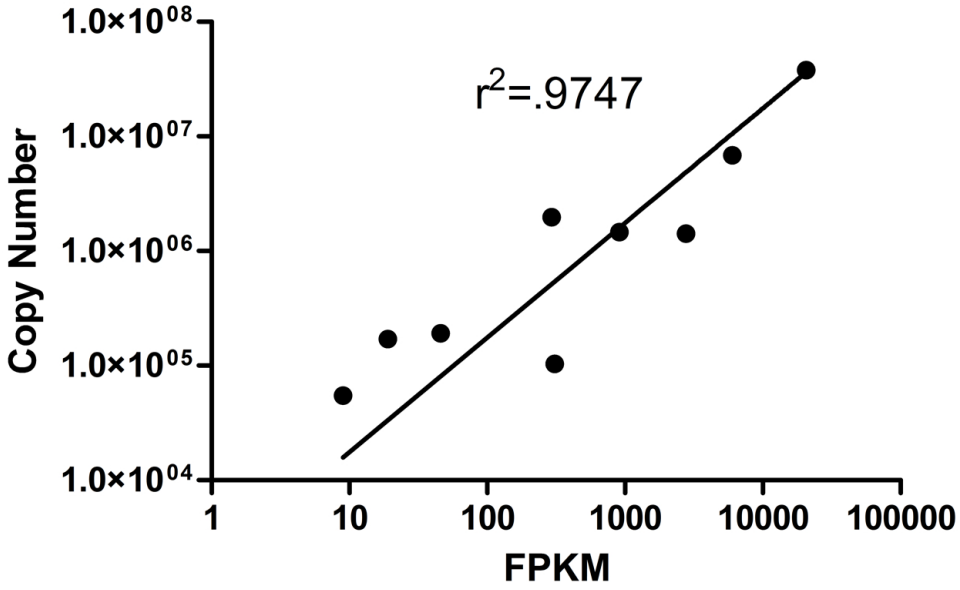


Figure 3. PCR amplification of (A) the 5' and 3' ends of *DDX3Y*, *PLCZ1*, *CRISP2* and *GSTM3* in amplified ds-cDNA. For 5' end primers, all primers begin in the first exon, and for 3' end primers, all primers end in the last exon. All primer sets are intron-spanning. N = negative control that did not include cDNA template and M = 100 bp DNA marker. (B) Transcripts for *GSTM3* and *GTSF1* were PCR amplified using primers within the first and last exons in order to capture full-length transcripts. The cDNA for this section was used from the 3-bull pool created from a Superscript III Reverse Transcription of mRNA (Invitrogen, Carlsbad, CA)

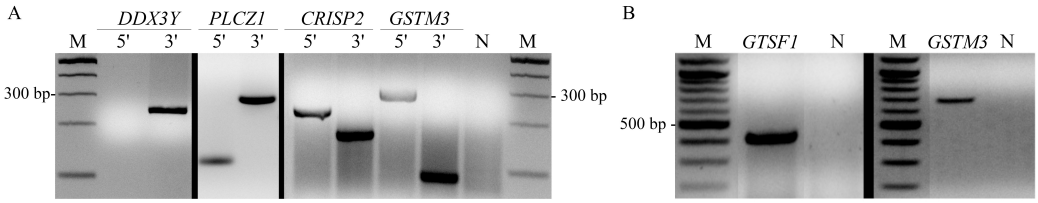
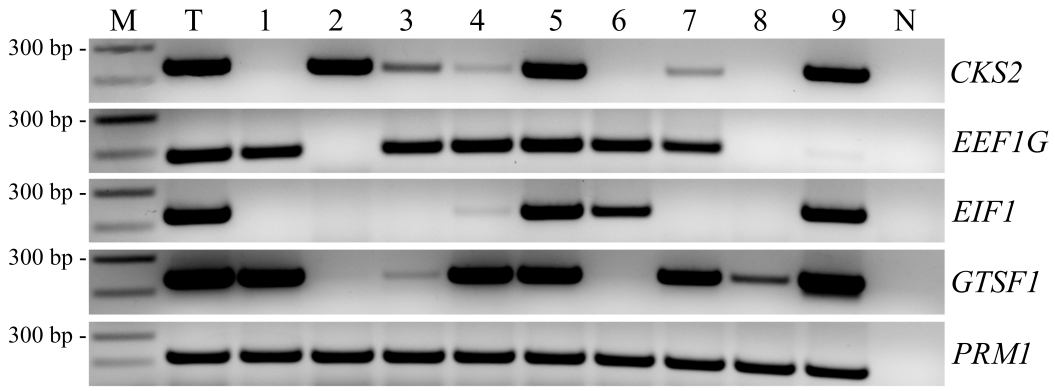


Figure 4. PCR amplification of select transcripts in individual bull sperm amplified cDNA



# TABLES

Table 1. Bovine primer sequences

| Gene Symbol | Genbank ID | | Primer Sequences | | Product Size (bp) |
|---|---|---|---|---|---|
| | | | Forward | Reverse | |
| *BTF3A* | AB098942 | **qPCR** | 5'-GGTGGTTCATAGAACAGCAACAGC -3' | 5'-GGCACCAAGCTGGTTTAAGATGCT -3' | 244 |
| *CD45* | AJ400864 | **PCR** | 5'-TGGACGAAATTGCATCCCTCAGGA -3' | 5'-TGGTCAGGACGTTTACAGCTCACA -3' | 237 |
| *CDH1* | AY508164 | **PCR** | 5'-ACCATGGACTTCTGCCAGAGGAAT -3' | 5'-TGGTCACCTGGTCTTTGTTCTGGT -3' | 244 |
| *CHMP5* | BC103182 | **qPCR** | 5'- TGGCACGGTGGACAGCAGAG -3' | 5' - TGGGCGAGATTGTCCCGCTG -3' | 189 |
| *C-KIT* | AF263827 | **PCR** | 5'-TATAGCACCATTGATGACAGCACA - 3' | 5'-TTATCTCCTCGACAACTTTCCACT -3' | 268 |
| *CKS2* | BC105331 | **Var*** | 5' - GAGTCGAGTCGTTGCCTTCA -3' | 5'-GGACACCAAGTCTCCTCCAC -3' | 248 |
| *CRISP2* | BC109478 | **5' Set** | 5'-CGGCCGCTCTGCAACAGAAG-3' | 5'-GGAAGCAGCACAGCGGTCAGA-3' | 120 |
| | | **3' Set** | 5'-CACCTTGCGGCAGTTGCCCT-3' | 5'-TGCCTTCACACAGACAAGTCGCC -3' | 165 |
| *DDX3Y* | GQ259590 | **5' Set** | 5'-TTGTTTCCGGTAGACCAACCTGTG-3' | 5'-AGCGCCCTTTGCTAGCTGTACT -3' | 220 |
| | | **3' Set** | 5'-GGCCGTTCTAGGAGATTCAGTGG -3' | 5'-CAACTGAATCTGCTTTCCAGCCAAG -3' | 246 |
| *EEF1G* | AB098752 | **Var*** | 5'-ATCCAGTTTCCGCCATGTGT-3' | 5'-GTTGCAACGCTCATCACTGG-3' | 198 |
| *EIF1* | BC103170 | **qPCR; Var*** | 5'-AAGGGTGATGATCTGCTTCCTGCT -3' | 5'-AACTGGCATATGTTCTTGCGCTGG -3' | 235 (379)* |
| *EIF4A* | BC103130 | **qPCR** | 5'-TGCCTTCTGATGTGCTTGAGGTGA -3' | 5'-TGAAGTCTCGGGCATGCATCTTCT -3' | 246 |
| *GSTM3* | BC112491 | **5' Set** | 5'-GCGCTAAGGCACACAGGCGA-3' | 5'-TGCGGGCGATGTAGCGCAAG-3' | 290 |
| | | **3' Set** | 5'-TGTGCCGTTTTGAGGCTTTGGAG-3' | 5'-GGGCCATCTTGTTGTTGACAGGCAT-3' | 90 |
| | | **5' to 3' Exon** | 5'-GCGCTAAGGCACACAGGCGA-3' | 5'-GGGCCATCTTGTTGTTGACAGGCAT-3' | 679 |
| *GTSF1* | BC102713 | **Var*** | 5'-CAGGTTCCTCGGGCTGAAAT-3' | 5'-ACTATGTTGCTTGCAGGGCT-3' | 239 |
| | | **5' to 3' Exon** | 5'-ACAAACTGGCAACTTGTCCCT-3' | 5'-GAACACACTGTAGCGGGAAGA-3' | 427 |
| *HMGB4* | BC109790 | **qPCR** | 5'- AGCTGGTCGGTGGTGCAGGT -3' | 5'- GCAAGCATGTCTTCCGGGC -3' | 167 |
| *PLCZ1* | BC114836 | **5' Set** | 5'- GGTGCCCGGCCAACCAGTTAT -3' | 5'- TGCCGCTTGGCAAGAAAGGG -3' | 138 |
| | | **3' Set** | 5'-GTGGTATCCAGTTGCCTCCCAGT -3' | 5'- GCGGGCTCAAGACTCTCACCC -3' | 319 |
| | | **qPCR** | 5'- CGGGTGGTCGGAATCCCACTCT -3' | 5'-AATTCCCTGGCTGCCAACTTTGT -3' | 194 |
| *PRM1* | BC108207 | **qPCR; Var*** | 5'-AAGAAGATGTCGCAGACGAAGGAG -3' | 5'-ACAGGTGGCATTGTTCGTTAGCAG -3' | 228 |
| *PSMA6* | BC110260 | **qPCR** | 5'-ACAGTGGAAACTGCGATTACATGCC - 3' | 5' - ACAGGCAAGTGGCGTCACGG - 3' | 205 |
| *SEC61G* | BC102186 | **qPCR** | 5'- GCAGACGCGGAGCAGACATCA -3' | 5'-AGCGAATCCTATTGCTGTTGCCA -3' | 155 |

* Primers used for individual bull spermatozoa RNA variation PCRs

** Genomic DNA amplicon which includes 144 bp intron

Table 2. Top 10 bovine spermatozoal transcripts based on FPKM

| Gene Symbol | Gene Name | Accession Number | FPKM |
|---|---|---|---|
| ***PRM1*** | Protamine 1 | BC108207; M14559 | 20667; 12461 |
| ***LOC783058*** | Hypothetical Protein | BC126791 | 10290 |
| ***HMGB4*** | High mobility group box 4 | BC109790 | 6022 |
| ***LOC404073*** | Histone 2B variant PT15 | BC108210; AF315690 | 3048; 2158 |
| *CHMP5* | Chromatin modifying protein 5 | BC103182 | 2778 |
| ***TMSB4X*** | Thymosin beta 4 X-linked | FJ795030 | 2487 |
| ***LOC281370*** | Polyubiquitin | AB099044 | 2426 |
| ***GSTM3*** | Glutathione S-transferase mu 3 | BC112491 | 2374 |
| ***N/A*** | cDNA clone IMAGE:7944277 | BC134702 | 2050 |
| ***KIF5C*** | Kinesin family member 5C | BC151732 | 1862 |

* Bold denotes full-length transcript.

Table 3. Previously reported spermatozoal transcripts identified in the bovine spermatozoal transcript profile

| Transcript | Accession # | FPKM | End Exons Intact 5' | End Exons Intact 3' | Reference | Species | Accession # |
|---|---|---|---|---|---|---|---|
| ***PRM1*** | BC108207 | **20667.2** | Y | Y | Ziyyat, 1999; Gilbert, 2007; Kempisty, 2008; Hecht, 2010; Feugang, 2010 | H; B; P; B, Ma; B | BC108207 |
| *CHMP5* | BC103182 | 2778.08 | N | N | Zhao et al., 2006; Lalancette, 2008 | H | BC103182 |
| ***TNP1*** | X16171 | **1287.96** | **Y** | **Y** | Iguchi, 2006 | M | X16171 |
| *TNP2* | BC109800 | 1206.79 | N | Y | Miller, 2005 | M | BC109800 |
| ***SMCP*** | BC109542 | **938.502** | **Y** | **Y** | Iguchi, 2006; Yang 2009 | M | BC109542 |
| ***CLGN*** | BC103401 | **220.011** | **Y** | **Y** | Kempisty, 2008; Ostermeier, 2004; Wang, 2004 | P | BC103401 |
| *TMBIM6* | BC102469 | 196.512 | N | Y | Gilbert, 2007 | B | BC102469 |
| ***PGK2*** | BC110004 | **173.412** | **Y** | **Y** | Iguchi, 2006 | M | BC110004 |
| *H2AFZ* | BC109743 | 166.742 | N | Y | Gilbert, 2007 | B | BC109743 |
| ***LOC789867 (EF-1, EEF1A1)*** | AF013213 | **133.722** | **Y** | **Y** | Lalancette, 2008; Zhao, 2006 | B | AF013213 |
| ***AKAP4*** | AF100170 | **126.623** | **Y** | **Y** | Gilbert, 2007; Ostermeier, 2004 | B | AF100170 |
| *RPS4Y* | BC133507 | 53.0481 | N | Y | Yao, 2009 | M | BC133507 |
| *PRM2* | BC109783 | 45.5481 | N | Y | Hecht, 2010 | B, Ma | BC109783 |
| *CLU* | BC118223 | 44.045 | N | Y | Gilbert, 2007; Kempisty, 2008 | B; P | BC118223 |
| *ACTG1* | BC102951 | 43.7095 | N | Y | Gilbert, 2007 | B | BC102951 |
| ***PLCZ1*** | AY646356 | **41.3639** | **Y** | **Y** | Hamatani, 2012 | H | AY646356 |
| ***MYCBP*** | BC109848 | **39.731** | **Y** | **Y** | Lambard et al., 2004; Kumar et al. 1993 | H | BC109848 |
| *PEBP1* | BC102389 | 29.3446 | N | Y | Bissonette, 2009; Arangasamy, 2011 | B | BC102389 |
| *SPAG4* | BC109514 | 25.6132 | N | Y | Gilbert, 2007 | B | BC109514 |
| *CCT8* | AF136609 | 17.9915 | N | N | Arangasamy, 2011 | B | AF136609 |
| *DDX3Y* | FJ659845 | 10.9846 | N | N | Sekiguchi, 2004; Yao, 2009 | H; M | FJ659845 |
| *PPIH* | BC120220 | 10.946 | N | N | Gilbert, 2007 | B | BC120220 |
| *STRBP* | BC123453 | 7.26818 | N | N | Gilbert, 2007 | B | BC123453 |
| *FLOT1* | BC104516 | 5.44937 | N | N | Gilbert, 2007 | B | BC104516 |
| *CSN2* | S67277 | 4.76535 | N | Y | Feugang, 2010 | B | S67277 |
| ***CRISP2*** | BC109478 | **4.07274** | **Y** | **Y** | Arangasamy, 2011; Zhao, 2006 | B | BC109478 |
| *EIF2B2* | BC123823 | 2.93951 | N | N | Gilbert, 2007 | B | BC123823 |
| *SPATA20* | BC123689 | 2.16637 | N | N | Gilbert, 2007 | B | BC123689 |

* Bold denotes full-length transcripts

Table 4. Top 10 previously unreported full-length bovine spermatozoal transcripts based on FPKM.

| Gene Symbol | Gene Name | Accession Number | FPKM |
|---|---|---|---|
| ***HMGB4*** | High mobility group box 4 | BC109790 | 6022 |
| ***TMSB4X*** | Thymosin beta 4 X-linked | FJ795030 | 2487 |
| ***LOC281370*** | Polyubiquitin | AB099044 | 2426 |
| ***PSMA6*** | Proteosome subunit, alpha type, 6 | BC110260 | 913 |
| ***GTSF1*** | Gametocyte specific factor 1 | BC102713 | 896 |
| ***ZNF474*** | Zinc finger protein 474 | BC108236 | 817 |
| ***COX7C*** | Cytochrome oxidase subunit 7c | X15725 | 733 |
| ***MLF1*** | Myeloid leukemia factor 1 | BC109859 | 517 |
| ***PFDN5*** | Prefoldin subunit 5 | BC102252 | 405 |
| ***CKS2*** | CDC28 protein kinase regulatory subunit 2 | BC105331 | 352 |

Table 5. Top 10 gene ontology categories for all spermatozoal transcripts FPKM > 0 and for transcripts with FPKM > 100.

| | All Transcripts (% of Transcripts Per Category) | | Transcripts with FPKM > 100 (% of Transcripts Per Category) | |
|---|---|---|---|---|
| Biological Process | Translation | 4.91 | Translation | 14.55 |
| | Proteolysis Involved In Cellular Protein CP | 4.48 | Protein Localization | 6.18 |
| | Cellular Protein CP | 4.39 | Precursor Metabolites And Energy | 5.82 |
| | Modification-Dependent Protein CP | 4.02 | Sexual Reproduction | 5.45 |
| | Modification-Dependent Macromolecule CP | 4.02 | Spermatogenesis | 5.09 |
| | Cellular Macromolecule CP | 4.00 | Male Gamete Generation | 5.09 |
| | Protein CP | 3.57 | Gamete Generation | 5.09 |
| | RNA Processing | 3.52 | Multicellular Organism Reproduction | 5.09 |
| | Macromolecule CP | 3.40 | Reproductive Process | 5.09 |
| | mRNA Processing | 3.14 | Protein Transport | 5.09 |
| Cellular Component | Intracellular NMBO | 11.35 | NMBO | 25.82 |
| | NMBO | 11.35 | Intracellular NMBO | 25.82 |
| | Mitochondrion | 7.98 | Ribonucleoprotein Complex | 16.73 |
| | Membrane-Enclosed Lumen | 6.90 | Ribosome | 14.18 |
| | Intracellular Org Lumen | 6.64 | Mitochondrion | 12.36 |
| | Org Lumen | 6.64 | Org Membrane | 10.18 |
| | Org Membrane | 5.49 | Org Envelope | 8.73 |
| | Cytoskeleton | 5.15 | Envelope | 8.73 |
| | Nuclear Lumen | 4.91 | Mitochondrial Part | 8.36 |
| | Ribonucleoprotein Complex | 4.56 | Mitochondrial Membrane | 8.00 |
| Molecular Function | Ion Bi | 15.68 | Structural Molecule Activity | 13.82 |
| | Cation Bi | 15.51 | Structural Constituent Of Ribosome | 13.09 |
| | Metal Ion Bi | 15.42 | RNA Bi | 4.73 |
| | Nucleotide Bi | 11.90 | Hydrogen Ion TTA | 4.36 |
| | Transition Metal Ion Bi | 10.86 | Monovalent Inorganic Cation TTA | 4.36 |
| | Purine Nucleotide Bi | 9.51 | Inorganic Cation TTA | 4.36 |
| | Purine Ribonucleotide Bi | 9.15 | ATPase Activity | 3.27 |
| | Ribonucleotide Bi | 9.15 | Enzyme Bi | 2.55 |
| | Nucleoside Bi | 7.44 | Protein Domain Specific Bi | 2.18 |
| | Purine Nucleoside Bi | 7.38 | Heme-Copper Terminal Oxidase Activity | 1.82 |

*CP = Catabolic Process, Bi = Binding, Org = Organelle, NMBO = Non-Membrane Bound Organelle, TTA = Transmembrane Transporter Activity

Supplementary Table 1. Full-Length Transcripts for the population of FPKM >100.

| Genbank ID | Official Gene Symbol | Full Name | FPKM | Full Length Transcript | 5' Exon intact | 3' Exon intact |
|---|---|---|---|---|---|---|
| | | **Percent Yes:** | | 65.76 | 70.65 | 88.59 |
| AB098851 | *ORCS10804* | Bos Taurus Mitochondrial Rna, Similar To 16S Rrna | 383371 | Y | Y | Y |
| AB098854 | *ORCS10931* | Bos Taurus Mitochondrial Rna, Similar To 16S Rrna | 301214 | Y | Y | Y |
| AB098841 | *ORCS10096* | Bos Taurus Mitochondrial Rna, Similar To 16S Rrna | 216612 | Y | Y | Y |
| AB098863 | *ORCS11599* | Bos Taurus Mitochondrial Rna, Similar To 16S Rrna | 149030 | Y | Y | Y |
| AB098844 | *ORCS10257* | Bos Taurus Mitochondrial Rna, Similar To 12S Rrna | 145266 | Y | Y | Y |
| AB099138 | *ORCS12829* | Bos Taurus Mitochondrial Rna, Similar To 16S Rrna | 133966 | Y | Y | Y |
| AB098853 | *ORCS10848* | Bos Taurus Mitochondrial Rna, Similar To 12S Rrna | 82000.2 | Y | Y | Y |
| BC108207 | *PRM1* | Bos Taurus Protamine 1 | 20667.2 | Y | Y | Y |
| DQ347622 | *H97* | Bos taurus clone H97 COX1 mRNA | 15042.7 | Y | Y | Y |
| M14559 | *PRM1* | Protamine 1 | 12460.6 | Y | Y | Y |
| DQ347619 | *H31* | Bos taurus clone H31 ND4 mRNA | 11070.1 | Y | Y | Y |
| DQ347618 | *ATP6* | Bos Taurus Clone A14 Atp6 Mrna | 10379.2 | Y | Y | Y |
| BC126791 | *LOC783058* | Bos taurus hypothetical protein LOC783058 | 10289.7 | Y | Y | Y |
| AB098808 | *ORCS12903* | Bos taurus mitochondrial mRNA for similar to ATPase 6 | 10255 | Y | Y | Y |
| DQ347621 | *H63* | Bos taurus clone H63 COX2 mRNA | 9735.03 | Y | Y | Y |
| AB099097 | *ORCS11619* | Bos Taurus Mitochondrial Rna, Similar To D-Loop | 8300.97 | Y | Y | Y |
| AB098776 | *ORCS12073* | Bos Taurus Mitochondrial Mrna For Similar To Cytochrome Oxidase III | 7772.1 | Y | Y | Y |
| AB098777 | *ORCS12084* | Bos Taurus Mitochondrial Mrna For Similar To Cytochrome Oxidase III | 6720.22 | Y | Y | Y |
| BC109790 | *Hmgb4* | High-Mobility Group Box 4 | 6021.96 | Y | Y | Y |
| AB099131 | *ORCS11856* | Bos taurus mitochondrial RNA, similar to 12S rRNA | 5795.61 | Y | Y | Y |
| DQ347627 | *H40* | Bos taurus clone H40 COX2 mRNA | 5408.65 | Y | Y | Y |
| AB099077 | *ORCS13694* | Bos taurus mitochondrial mRNA for similar to cytochrome oxidase I | 5297.64 | Y | Y | Y |
| AB098902 | *ORCS10210* | Bos taurus mRNA for similar to cytochrome oxidase I | 5046 | Y | Y | Y |
| AB099009 | *ORCS12081* | Bos taurus mRNA for similar to cytochrome b | 4319.09 | Y | Y | Y |
| AB098967 | *ORCS11394* | Bos Taurus Mrna For Similar To Cytochrome B | 3394.89 | Y | Y | Y |
| BC126791 | *MGC148328* | Bos taurus hypothetical protein LOC783058, mRNA | 3198.14 | Y | Y | Y |
| FJ976184 | *ND5* | Bos taurus NADH dehydrogenase subunit 5 (ND5) mRNA | 3091.53 | Y | Y | Y |
| BC108210 | *LOC404073* | Histone H2B Variant Pt15 | 3047.5 | Y | Y | Y |
| AB098941 | *ORCS10715* | Bos taurus mRNA for similar to cytochrome b | 2918.85 | Y | Y | Y |
| BC103182 | *Chmp5* | Chromatin Modifying | 2778.08 | N | N | N |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Protein 5 | | | | |
| AB098789 | *ORCS12473* | Bos Taurus Mitochondrial Mrna For Similar To Cytochrome Oxidase III | 2521.74 | Y | Y | Y |
| FJ795030 | *LOC785455* | Thymosin Beta 4, X-Linked | 2486.99 | Y | Y | Y |
| AB099044 | *LOC281370* | Polyubiquitin | 2425.89 | Y | Y | Y |
| BC112491 | *GSTM3* | Glutathione S-Transferase Mu 3 (Brain) | 2373.84 | Y | Y | Y |
| AF315690 | *LOC404073* | Histone H2B Variant Pt15 | 2158.22 | Y | Y | Y |
| AB098774 | *ORCS11961* | Bos taurus mitochondrial mRNA for similar to cytochrome oxidase III | 2133.68 | Y | Y | Y |
| AB099096 | *ORCS11109* | Bos taurus mitochondrial RNA, similar to D-loop | 2091.43 | Y | Y | Y |
| BC134702 | *IMAGE:7944277* | Bos taurus cDNA clone IMAGE:7944277 | 2050.08 | Y | Y | Y |
| AB098801 | *ORCS12731* | Bos taurus mitochondrial mRNA for similar to cytochrome oxidase III | 2029.43 | Y | Y | Y |
| BC151732 | *KIF5C* | Kinesin Family Member 5C | 1862.47 | Y | Y | Y |
| BC123382 | *LOC777592* | Hypothetical Protein Loc777592 | 1846.72 | N | N | Y |
| BC126793 | *IMAGE:8056303* | Bos taurus cDNA clone IMAGE:8056303 | 1808.5 | N | N | Y |
| AB098980 | *ORCS11606* | Bos taurus mitochondrial mRNA for similar to NADH dehydrogenase subunit 1 | 1586.19 | Y | Y | Y |
| AB098969 | *ORCS11414* | Bos taurus mRNA for similar to NADH dehydrogenase subunit 1 | 1563.82 | Y | Y | Y |
| BC114001 | *LOC281370* | Polyubiquitin | 1520.12 | Y | Y | Y |
| BC111648 | *MGC137055* | Hypothetical Protein Mgc137055 | 1466.25 | Y | Y | Y |
| AB098767 | *ORCS11606* | Bos taurus mitochondrial mRNA, similar to protein 1 | 1310.8 | Y | Y | Y |
| X16171 | *tnp1* | Transition Protein 1 | 1287.96 | Y | Y | Y |
| BC109730 | *C13H20orf79* | Chromosome 20 Open Reading Frame 79 Ortholog | 1237.94 | Y | Y | Y |
| BC142065 | *IMAGE:8037824* | Bos taurus cDNA clone IMAGE:8037824 | 1224.68 | Y | Y | Y |
| BC109800 | *LOC781496* | Similar To Tnp2 Protein; Transition Protein 2 (During Histone To Protamine Replacement) | 1206.79 | N | N | Y |
| BC126791 | *IMAGE:30957795* | Bos taurus hypothetical protein LOC783058, mRNA | 1137.99 | Y | Y | Y |
| AB098750 | *LOC614114* | Cytochrome C Oxidase Subunit Vib Pseudogene | 1099.5 | N | N | Y |
| BC111151 | *IMAGE:8052434* | Bos taurus cDNA clone IMAGE:8052434 | 993.402 | N | N | Y |
| K00243 | *tRNA-Leu* | Bovine Mitochondrial Leu-Trna-Tag | 972.836 | Y | Y | Y |
| AY796023 | *Smcp* | Sperm Mitochondria-Associated Cysteine-Rich Protein | 938.502 | Y | Y | Y |
| BC109478 | *IMAGE:8048928* | Bos taurus cysteine-rich secretory protein 2, mRNA | 933.969 | Y | Y | Y |
| BC103421 | *Spa17* | Sperm Autoantigenic Protein 17 | 927.374 | Y | Y | Y |
| BC110260 | *Psma6* | Proteasome (Prosome, Macropain) Subunit, Alpha Type, 6 | 913.21 | Y | Y | Y |
| BC102663 | *C12orf54* | Chromosome 12 Open Reading Frame 54 Ortholog | 897.064 | Y | Y | Y |
| BC102713 | *GTSF1* | Gametocyte Specific Factor 1 | 896.368 | Y | Y | Y |
| BC102609 | *C3H1orf182* | Chromosome 1 Open Reading Frame 182 | 887.309 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Ortholog | | | | |
| BC102599 | *GTSF1L* | Gametocyte Specific Factor 1-Like | 861.791 | Y | Y | Y |
| BC108236 | *ZNF474* | Zinc Finger Protein 474 | 816.725 | Y | Y | Y |
| BC102973 | *LOC539855* | Histone H3-Like | 805.895 | Y | Y | Y |
| AB099083 | *LOC281370* | Ubiquitin C; Polyubiquitin; Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1 | 797.874 | Y | Y | Y |
| DQ347600 | *A24* | Bos Taurus Clone H1 Atpase Na+/K+ Transporting Beta 3 Polypeptide-Like Mrna | 786.943 | Y | Y | Y |
| BC126792 | *LOC784495* | Hypothetical Protein Loc784495 | 741.663 | Y | Y | Y |
| X15725 | *Cox7c* | Cytochrome C Oxidase Subunit Viic | 732.771 | Y | Y | Y |
| DQ347636 | *COX7A2* | Cytochrome C Oxidase Subunit Viia Polypeptide 2 (Liver) | 719.358 | Y | Y | Y |
| BC109926 | *IQCF5* | Iq Motif Containing F5 | 707.481 | Y | Y | Y |
| AB098957 | *LOC281370* | Polyubiquitin | 677.922 | Y | Y | Y |
| BC102598 | *tnp1* | Transition Protein 1 (During Histone To Protamine Replacement) | 676.993 | Y | Y | Y |
| BC103105 | *CISD1* | Cdgsh Iron Sulfur Domain 1 | 656.043 | N | N | Y |
| BC114790 | *IMAGE:8063641* | Bos taurus cDNA clone IMAGE:8063641 | 619.729 | N | N | Y |
| BC109542 | *Smcp* | Sperm Mitochondria-Associated Cysteine-Rich Protein | 604.724 | Y | Y | Y |
| BC148014 | *rpl23* | Ribosomal Protein L23 | 599.605 | Y | Y | Y |
| BC102582 | *MP68* | 6.8 Kda Mitochondrial Proteolipid | 584.314 | N | N | Y |
| Z86042 | *LEO1* | Leo1, Paf1/Rna Polymerase Ii Complex Component, Homolog (S. Cerevisiae) | 562.891 | Y | Y | Y |
| BC110036 | *Clph* | Chromosome 4 Open Reading Frame 35 Ortholog | 545.537 | Y | Y | Y |
| DQ347576 | *SLC25A5* | Solute Carrier Family 25 (Mitochondrial Carrier; Adenine Nucleotide Translocator), Member 5 | 542.739 | Y | Y | Y |
| M62428 | *LOC281370* | Ubiquitin C; Polyubiquitin; Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1 | 538.61 | N | N | Y |
| K00194 | *tRNA-Glu* | Bovine Mitochondrial Glu-Trna-Uuc | 533.929 | Y | Y | Y |
| BC111614 | *LOC768323* | Hypothetical Protein Loc768323 | 522.375 | Y | Y | Y |
| BC109859 | *MLF1* | Myeloid Leukemia Factor 1 | 516.634 | Y | Y | Y |
| AY911357 | *rpl31* | Similar To Ribosomal Protein L31; Ribosomal Protein L31 | 507.005 | Y | Y | Y |
| AY260742 | *LIS1* | Bos taurus platelet activating factor acetylhydrolase 45 kDa subunit brain isoform (LIS1) mRNA | 503.684 | Y | Y | Y |
| J03604 | *GLUL* | Glutamate-Ammonia Ligase (Glutamine Synthetase) | 500.939 | Y | Y | Y |
| BC102702 | *LOC782520* | Ribosomal Protein S29 | 497.777 | Y | Y | Y |
| DQ347636 | *COX7A2* | Cytochrome C Oxidase Subunit Viia Polypeptide 2 (Liver) | 496.514 | Y | Y | Y |
| BC109927 | *MORN2* | Morn Repeat Containing 2 | 472.825 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| BC105360 | *spata6* | Spermatogenesis Associated 6 | 453.406 | Y | Y | Y |
| AF294616 | *TMSB10* | Thymosin Beta 10 | 451.9 | N | N | Y |
| BC102650 | *MGC128040* | Hypothetical Protein Mgc128040 | 442.179 | Y | Y | Y |
| BC149673 | *MGC152346* | Uncharacterized Protein Loc285141 Homolog | 440.761 | Y | Y | Y |
| BC126781 | *TXNDC8* | Thioredoxin Domain Containing 8 (Spermatozoa) | 433.154 | Y | Y | Y |
| S79980 | *RPL37* | ribosomal protein L37 | 432.006 | Y | Y | Y |
| AB099079 | *LOC789867* | Eukaryotic Translation Elongation Factor 1 Alpha 1 | 431.873 | Y | Y | Y |
| BC102748 | *rpl32* | Ribosomal Protein L32 | 419.736 | Y | Y | Y |
| BC111614 | *LOC768323* | Hypothetical Protein Loc768323 | 414.631 | Y | Y | Y |
| BC102252 | *PFDN5* | Prefoldin Subunit 5 | 404.632 | Y | Y | Y |
| BC102044 | *RPL37A* | Ribosomal Protein L37A | 403.288 | N | Y | N |
| BC109951 | *CAPZA3* | Capping Protein (Actin Filament) Muscle Z-Line, Alpha 3 | 402.076 | Y | Y | Y |
| BC102248 | *LOC281370* | Polyubiquitin | 401.415 | Y | Y | Y |
| BC120080 | *CALM* | Calmodulin-Like | 395.156 | Y | Y | Y |
| BC142077 | *IMAGE:8050622* | Bos taurus cDNA clone IMAGE:8050622 | 393.899 | N | N | Y |
| AY186585 | *GLUL* | Glutamate-Ammonia Ligase (Glutamine Synthetase) | 392.537 | Y | Y | Y |
| DQ347578 | *A17* | Bos taurus clone A17 actin cytoplasmic 2 mRNA | 387.604 | Y | Y | Y |
| AJ297742 | *GABARAP* | Gaba(A) Receptor-Associated Protein | 384.967 | Y | Y | Y |
| BC142060 | *DNAJB7* | Dnaj (Hsp40) Homolog, Subfamily B, Member 7 | 366.633 | Y | Y | Y |
| BC108144 | *BANF2* | Barrier To Autointegration Factor 2 | 359.621 | Y | Y | Y |
| BC114198 | *IMAGE:8055902* | Bos taurus cDNA clone IMAGE:8055902 | 357.25 | Y | Y | Y |
| BC105331 | *CKS2* | Cdc28 Protein Kinase Regulatory Subunit 2 | 351.893 | Y | Y | Y |
| BC114201 | *IMAGE:8056539* | Bos taurus cDNA clone IMAGE:8056539 | 348.138 | Y | Y | Y |
| BC149889 | *DCUN1D1* | Dcn1, Defective In Cullin Neddylation 1, Domain Containing 1 (S. Cerevisiae) | 342.726 | Y | Y | Y |
| AF109198 | *CLIC4* | Chloride Intracellular Channel 4 | 338.487 | Y | Y | Y |
| BC126766 | *FAM24A* | Similar To Protein Fam24A | 335.134 | Y | Y | Y |
| BC110256 | *Fam71d* | Family With Sequence Similarity 71, Member D Ortholog | 335.075 | Y | Y | Y |
| BC109624 | *cetn1* | Centrin, Ef-Hand Protein, 1 | 332.853 | N | N | Y |
| BC102682 | *SERF2* | Small Edrk-Rich Factor 2 | 330.258 | N | N | Y |
| BC102249 | *rps11* | Ribosomal Protein S11 | 327.944 | Y | Y | Y |
| BC148018 | *rps17* | Ribosomal Protein S17 | 327.412 | Y | Y | Y |
| BC109989 | *C13H20ORF71* | Chromosome 20 Open Reading Frame 71 Ortholog | 326.017 | Y | Y | Y |
| BC102437 | *atox1* | Atx1 Antioxidant Protein 1 Homolog (Yeast) | 320.276 | Y | Y | Y |
| DQ347614 | *LOC784052* | 40S Ribosomal Protein S26-2-Like | 319.419 | Y | Y | Y |
| BC109725 | *SAA4* | Serum Amyloid A4, Constitutive | 319.23 | Y | Y | Y |
| U19802 | *btg1* | B-Cell Translocation Gene 1, Anti-Proliferative | 316.825 | N | N | Y |
| BC108179 | *RPL38* | Ribosomal Protein L38 | 316.575 | Y | Y | Y |
| BC111617 | *Tmco2* | Transmembrane And Coiled-Coil Domains 2 | 313.784 | N | N | Y |
| BC114194 | *IMAGE:8063913* | Bos taurus cDNA clone | 311.9 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | IMAGE:8063913 | | | | |
| BC103057 | *UQCRB* | Ubiquinol-Cytochrome C Reductase Binding Protein | 309.82 | Y | Y | Y |
| BC102186 | *sec61g* | Sec61 Gamma Subunit | 309.446 | Y | Y | Y |
| AB098960 | *ORCS11043* | Bos taurus mRNA for similar to poly(A)-binding protein 1 | 307.612 | Y | Y | Y |
| EU036210 | *BBD120* | Bos taurus beta-defensin 120 mRNA | 306.483 | N | Y | N |
| BC108218 | *C29H11orf10* | Chromosome 11 Open Reading Frame 10 Ortholog | 297.346 | N | N | N |
| BC103170 | *LOC781102* | Eukaryotic Translation Initiation Factor 1 | 292.75 | N | N | Y |
| BC142260 | *taf10* | Taf10 Rna Polymerase Ii, Tata Box Binding Protein (Tbp)-Associated Factor, 30Kda | 289.583 | N | N | Y |
| BC102743 | *Tmco5a* | Transmembrane And Coiled-Coil Domains 5A | 287.894 | N | N | Y |
| BC108230 | *SERF1A* | Small Edrk-Rich Factor 1B (Centromeric) | 277.569 | N | N | Y |
| AF058700 | *LOC281370* | Ubiquitin C; Polyubiquitin; Ubiquitin A-52 Residue Ribosomal Protein Fusion Product 1 | 277.039 | Y | Y | Y |
| BC109684 | *LOC540268* | Hypothetical Loc540268 | 272.881 | Y | Y | Y |
| BC102675 | *DCUN1D1* | Dcn1, Defective In Cullin Neddylation 1, Domain Containing 1 (S. Cerevisiae) | 269.749 | Y | Y | Y |
| M19217 | *Atp5j* | Atp Synthase, H+ Transporting, Mitochondrial F0 Complex, Subunit F6 | 269.601 | Y | Y | Y |
| AY911383 | *LOC786337* | Ribosomal Protein S24 | 267.359 | N | N | Y |
| BC102168 | *LOC781607* | Ribosomal Protein L36A | 262.035 | N | N | Y |
| BC103196 | *IMAGE:7986614* | Bos taurus transcription elongation factor B (SIII), polypeptide 2 | 261.222 | N | N | Y |
| BC110154 | *MS4A13* | Membrane-Spanning 4-Domains, Subfamily A, Member 13 | 258.624 | Y | Y | Y |
| BC151426 | *LOC786258* | Ran, Member Ras Oncogene Family | 258.31 | Y | Y | Y |
| X15112 | *LOC614114* | Cytochrome C Oxidase Subunit Vib Pseudogene | 257.306 | N | N | N |
| BC109719 | *SPINK2* | Serine Peptidase Inhibitor, Kazal Type 2 (Acrosin-Trypsin Inhibitor) | 248.263 | Y | Y | Y |
| BT030506 | *UBE2N* | Ubiquitin-Conjugating Enzyme E2N (Ubc13 Homolog, Yeast) | 245.596 | Y | Y | Y |
| EU036209 | *BBD119* | Bos taurus beta-defensin 119 mRNA | 244.782 | Y | Y | Y |
| Z46789 | *CYLC2* | Cylicin, Basic Protein Of Sperm Head Cytoskeleton 2 | 243.523 | Y | Y | Y |
| DQ347568 | *LOC781571* | Histidine Triad Nucleotide Binding Protein 1; Similar To Histidine Triad Nucleotide-Binding Protein 1 | 242.297 | Y | Y | Y |
| BC102631 | *LOC617040* | Similar To Hcg23722 | 241.208 | N | Y | N |
| AY911363 | *LOC507141* | Ce5 Protein-Like | 241.165 | Y | Y | Y |
| BC108150 | *Selk* | Selenoprotein K | 238.417 | Y | Y | Y |
| BC102957 | *GPX4* | Glutathione Peroxidase 4 (Phospholipid Hydroperoxidase) | 238.117 | N | N | Y |
| BC149307 | *LOC100125949* | Similar To Iq Domain-Containing Protein F1 | 235.802 | Y | Y | Y |
| AB099097 | *ORCS11619* | Bos Taurus Mitochondrial | 235.006 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Rna, Similar To D-Loop | | | | |
| BC110123 | *C16H1orf49* | Chromosome 1 Open Reading Frame 49 Ortholog | 234.333 | Y | Y | Y |
| BC105361 | *Ldhc* | Lactate Dehydrogenase C | 230.203 | Y | Y | Y |
| BC123583 | *AP2B1* | Adaptor-Related Protein Complex 2, Beta 1 Subunit | 228.35 | N | Y | Y |
| AY911358 | *LOC781565* | Ribosomal Protein S6 | 227.299 | Y | Y | Y |
| DQ347613 | *rps8* | Ribosomal Protein S8 | 222.367 | Y | Y | Y |
| AB098827 | *LOC781379* | Dynein, Light Chain, Lc8-Type 1 | 222.35 | Y | Y | Y |
| DQ347611 | *rps11* | Ribosomal Protein S11 | 221.749 | Y | Y | Y |
| Y10372 | *CAPZB* | Capping Protein (Actin Filament) Muscle Z-Line, Beta | 220.387 | Y | Y | Y |
| BC103401 | *clgn* | Calmegin | 220.011 | Y | Y | Y |
| BC114181 | *DBI* | Diazepam Binding Inhibitor (Gaba Receptor Modulator, Acyl-Coenzyme A Binding Protein) | 219.957 | Y | Y | Y |
| M19962 | *COX5B* | Cytochrome C Oxidase Subunit Vb | 218.769 | Y | Y | Y |
| X16978 | *LOC782270* | Similar To Atp Synthase Subunit Epsilon, Mitochondrial | 217.565 | Y | Y | Y |
| BC108217 | *Dynlrb2* | Dynein, Light Chain, Roadblock-Type 2 | 215.832 | N | Y | N |
| BC102491 | *LOC281370* | Polyubiquitin | 215.342 | N | Y | N |
| BC126796 | *C23H6orf129* | Chromosome 6 Open Reading Frame 129 Ortholog | 211.668 | N | Y | N |
| BC148017 | *IMAGE:7946562* | Bos taurus ribosomal protein L37, mRNA | 208.936 | Y | Y | Y |
| BC103060 | *GABARAP* | Gaba(A) Receptor-Associated Protein | 203.266 | N | N | Y |
| BC126795 | *DEFB123* | Defensin, Beta 123 | 201.937 | N | Y | N |
| BC102751 | *SPATA19* | Spermatogenesis Associated 19 | 201.143 | Y | Y | Y |
| BC108162 | *SEC62* | Sec62 Homolog (S. Cerevisiae) | 200.111 | N | N | Y |
| BC108191 | *C29H11orf67* | Chromosome 11 Open Reading Frame 67 Ortholog | 199.775 | Y | Y | Y |
| AY835842 | *H2AFZ* | Bos taurus histone H2A mRNA | 197.567 | N | N | N |
| BC102469 | *tmbim6* | Transmembrane Bax Inhibitor Motif Containing 6 | 196.512 | N | N | Y |
| BC102286 | *GNB2L1* | Guanine Nucleotide Binding Protein (G Protein), Beta Polypeptide 2-Like 1 | 196.51 | Y | Y | Y |
| BC120462 | *tspan5* | Tetraspanin 5 | 195.749 | Y | Y | Y |
| BC108233 | *polr2i* | Polymerase (Rna) Ii (Dna Directed) Polypeptide I, 14.5Kda | 195.592 | Y | Y | Y |
| BC103314 | *LOC784243* | Ribosomal Protein L34; Similar To Ribosomal Protein L34 | 195.156 | Y | Y | Y |
| BC102445 | *RpL30* | Ribosomal Protein L30 | 195.066 | N | Y | N |
| BC114016 | *Ccdc54* | Coiled-Coil Domain Containing 54 | 194.483 | Y | Y | Y |
| BC102549 | *Ropn1* | Ropporin, Rhophilin Associated Protein 1 | 193.878 | N | N | Y |
| BC109557 | *meig1* | Meiosis Expressed Gene 1 Homolog (Mouse) | 193.741 | N | N | Y |
| BC111660 | *LOC526524* | Fk506 Binding Protein 1A, 12Kda; Fk506 Binding Protein 1A, 12Kda-Like | 193.185 | Y | Y | Y |
| BC118480 | *S100G* | S100 Calcium Binding Protein G | 192.026 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| BC118372 | *SRPK2* | Sfrs Protein Kinase 2 | 190.687 | N | N | Y |
| BC109867 | *DDX25* | Dead (Asp-Glu-Ala-Asp) Box Polypeptide 25 | 188.791 | N | N | Y |
| BC108151 | *Rangrf* | Ran Guanine Nucleotide Release Factor | 188.571 | N | N | Y |
| BC116060 | *capns1* | Calpain, Small Subunit 1 | 187.956 | N | N | Y |
| AF520959 | *Fau* | Finkel-Biskis-Reilly Murine Sarcoma Virus (Fbr-Musv) Ubiquitously Expressed; Similar To Ubiquitin-Like/S30 Ribosomal Fusion Protein | 187.816 | N | N | N |
| BC111147 | *LOC786899* | Similar To Gtpase Activating Protein Testicular Gap1; Hypothetical Loc786899; Hypothetical Protein Mgc134093 | 187.06 | Y | Y | Y |
| AB099017 | *LOC789997* | Similar To 40S Ribosomal Protein S3A; Similar To Ribosomal Protein S3A; Ribosomal Protein S3A; Similar To Ribosomal Protein S3A | 186.803 | Y | Y | Y |
| BC110030 | *BCAP29* | B-Cell Receptor-Associated Protein 29 | 186.465 | Y | Y | Y |
| BC142080 | *LOC100271685* | Membrane-Spanning 4-Domains, Subfamily A-Like Similar To Mcg10725; | 186.038 | Y | Y | Y |
| AY911354 | *LOC785691* | Ribosomal Protein S25; Similar To Ribosomal Protein S25 | 185.298 | Y | Y | Y |
| BC109670 | *MRPL42* | Mitochondrial Ribosomal Protein L42 | 184.466 | Y | Y | Y |
| BC111147 | *LOC786899* | Similar To Gtpase Activating Protein Testicular Gap1; Hypothetical Loc786899; Hypothetical Protein Mgc134093 | 182.392 | Y | Y | Y |
| BC102669 | *Ppp1r2* | Protein Phosphatase 1, Regulatory (Inhibitor) Subunit 2 | 181.173 | N | N | Y |
| AY911347 | *RpL35A* | Ribosomal Protein L35A | 177.346 | Y | Y | Y |
| BC102382 | *YWHAZ* | Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein, Zeta Polypeptide | 175.813 | N | N | Y |
| BC102877 | *snrpd2* | Small Nuclear Ribonucleoprotein D2 Polypeptide 16.5Kda | 173.477 | Y | Y | Y |
| BC110004 | *PGK2* | Phosphoglycerate Kinase 2 | 173.412 | Y | Y | Y |
| AB098832 | *LOC789997* | Similar To 40S Ribosomal Protein S3A; Similar To Ribosomal Protein S3A; Ribosomal Protein S3A; Similar To Ribosomal Protein S3A | 173.237 | Y | Y | Y |
| GU817014 | *YWHAZ* | Bos Taurus Tyrosine-3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta Polypeptide | 172.597 | Y | Y | Y |
| GU817014 | *YWHAZ* | Bos Taurus Tyrosine-3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta Polypeptide | 172.597 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| BC126782 | *LOC100126817* | Hypothetical Protein Loc100126817 | 172.491 | Y | Y | Y |
| DQ347605 | *LOC782668* | Ribosomal Protein L6 | 171.318 | Y | Y | Y |
| BC105179 | *rpl35* | Ribosomal Protein L35 | 170.04 | N | N | Y |
| BC111663 | *LYRM7* | Lyrm7 Homolog (Mouse) | 168.54 | Y | Y | Y |
| BC102194 | *EIF5* | Eukaryotic Translation Initiation Factor 5 | 168.115 | N | N | Y |
| BC109743 | *H2AFZ* | H2A Histone Family, Member Z | 166.742 | N | N | Y |
| BC118158 | *IMAGE:8211381* | Bos taurus ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1, 3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2, mRNA | 166.262 | Y | Y | Y |
| BT025435 | *C14orf153* | Hypothetical Protein Loc617441 | 165.939 | N | N | Y |
| BC108222 | *IMAGE:8043996* | Bos taurus cDNA clone IMAGE:8043996 | 165.44 | Y | Y | Y |
| DQ677839 | *C13H20ORF71* | Chromosome 20 Open Reading Frame 71 Ortholog | 165.204 | Y | Y | Y |
| BC108247 | *SLIRP* | Sra Stem-Loop-Interacting Rna-Binding Protein | 162.812 | Y | Y | Y |
| AB099059 | *rps3* | Ribosomal Protein S3 | 162.118 | Y | Y | Y |
| BC102455 | *LOC786431* | Atp Synthase, H+ Transporting, Mitochondrial F0 Complex, Subunit G | 160.879 | Y | Y | Y |
| AB098994 | *LOC784528* | Atpase, H+ Transporting, Lysosomal 34Kda, V1 Subunit D | 160.731 | Y | Y | Y |
| BC102175 | *C26H10orf84* | Chromosome 10 Open Reading Frame 84 Ortholog | 160.09 | Y | Y | Y |
| BC109561 | *Rpl10l* | Ribosomal Protein L10-Like | 159.931 | N | N | N |
| BC109732 | *IMAGE:8059175* | Bos taurus cDNA clone IMAGE:8059175 | 159.464 | Y | Y | Y |
| BC111654 | *RpL35A* | Ribosomal Protein L35A | 158.971 | Y | Y | Y |
| BC102313 | *Rpl27* | Similar To Ribosomal Protein L27; Ribosomal Protein L27 | 158.242 | N | N | N |
| BC105143 | *LOC789244* | Lysophospholipase I; Similar To Lysophospholipase I | 157.322 | N | Y | N |
| DQ347607 | *LOC509829* | Ribosomal Protein L10; Ribosomal Protein L10 Pseudogene; Similar To Ribosomal Protein L10 | 157.248 | N | N | N |
| BC102970 | *hsbp1* | Heat Shock Factor Binding Protein 1 | 156.888 | Y | Y | Y |
| BC102292 | *NDUFS4* | Nadh Dehydrogenase (Ubiquinone) Fe-S Protein 4, 18Kda (Nadh-Coenzyme Q Reductase) | 156.337 | Y | Y | Y |
| BC103431 | *ELP2P* | Endozepine-Like Peptide 2 Pseudogene | 156.076 | N | N | Y |
| BC120463 | *MGC151969* | Uncharacterized Protein Ensp00000334415 Homolog | 156.043 | N | N | Y |
| AY911366 | *rps11* | Ribosomal Protein S11 | 155.631 | Y | Y | Y |
| S70447 | *GI:7579921* | F1Fo-ATP synthase complex Fo membrane domain f subunit | 155.371 | Y | Y | Y |
| BC109581 | *DYDC1* | Dpy30 Domain Containing 1 | 154.473 | Y | Y | Y |
| BC111293 | *LOC780805* | Hypothetical Protein Loc780805 | 152.434 | Y | Y | Y |
| BC146140 | *Dydc2* | Dpy30 Domain Containing 2 | 152.001 | N | N | Y |
| DQ347605 | *LOC782668* | Ribosomal Protein L6 | 151.466 | Y | Y | Y |
| AB098890 | *ORCS10052* | Bos Taurus Mrna For | 151.332 | N | N | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | Similar To Beta 2-Microglobulin | | | | |
| X64836 | *NDUFB9* | Nadh Dehydrogenase (Ubiquinone) 1 Beta Subcomplex, 9, 22Kda | 149.949 | Y | Y | Y |
| BC102593 | *MORF4L1* | Similar To Morf-Related Gene 15; Mortality Factor 4 Like 1 | 149.417 | N | N | Y |
| BC109924 | *Tspan6* | Tetraspanin 6 | 147.491 | N | Y | Y |
| BC103363 | *KPNA2* | Karyopherin Alpha 2 (Rag Cohort 1, Importin Alpha 1) | 146.212 | N | N | Y |
| DQ347612 | *rps12* | Ribosomal Protein S12 | 144.865 | N | N | Y |
| BC103260 | *CA2* | Y Box Binding Protein 1 | 144.827 | N | N | N |
| BC109577 | *Cetn4* | Centrin 4 | 144.395 | Y | Y | Y |
| BC108180 | *rps21* | Ribosomal Protein S21 | 142.837 | N | Y | N |
| BC111609 | *Iqcf2* | Iq Motif Containing F2 | 142.552 | N | N | Y |
| BC102890 | *Aif1* | Allograft Inflammatory Factor 1 | 141.932 | N | Y | N |
| BC109726 | *C3H1orf189* | Chromosome 1 Open Reading Frame 189 Ortholog | 141.298 | Y | Y | Y |
| DQ347592 | *LOC781370* | Ferritin, Heavy Polypeptide 1; Similar To Ferritin Heavy Chain; Similar To Ferritin, Heavy Polypeptide 1 | 140.117 | Y | Y | Y |
| BC103298 | *CCT2* | Chaperonin Containing Tcp1, Subunit 2 (Beta) | 140.009 | Y | Y | Y |
| AF265669 | *RPGRIP1* | Retinitis Pigmentosa Gtpase Regulator Interacting Protein 1 | 139.66 | N | N | Y |
| AF164025 | *RNASE6* | Ribonuclease, Rnase A Family, K6 | 138.861 | Y | Y | Y |
| BC102535 | *TPPP2* | Tubulin Polymerization-Promoting Protein Family Member 2 | 138.018 | N | N | Y |
| BC102655 | *LRRC67* | Leucine Rich Repeat Containing 67 | 137.872 | Y | Y | Y |
| BC146224 | *QTRTD1* | Queuine Trna-Ribosyltransferase Domain Containing 1 | 137.523 | N | N | N |
| BC148911 | *TRDN* | Triadin | 137.457 | N | N | N |
| BC111170 | *C10H15orf23* | Chromosome 15 Open Reading Frame 23 Ortholog | 137.17 | N | N | Y |
| BC111202 | *ilf2* | Interleukin Enhancer Binding Factor 2, 45Kda | 136.727 | N | N | Y |
| BC108198 | *PRM3* | Bos taurus protamine 3, mRNA | 136.117 | N | N | N |
| BC102492 | *LOC616936* | Male-Enhanced Antigen 1 | 135.939 | N | N | N |
| BC102230 | *Rnf181* | Ring Finger Protein 181 | 135.888 | N | N | N |
| BC102391 | *PSMC2* | Proteasome (Prosome, Macropain) 26S Subunit, Atpase, 2 | 135.115 | N | Y | N\ |
| BC103021 | *LOC785297* | Ferritin, Light Polypeptide | 134.371 | Y | Y | Y |
| BC110226 | *C20orf111* | Hypothetical Protein Loc510457 | 134.345 | N | N | Y |
| AY911377 | *LOC785455* | Similar To Thymosin, Beta 4; Thymosin Beta 4, X-Linked | 134.314 | N | N | Y |
| BC111643 | *IMAGE:8018076* | Bos taurus cDNA clone IMAGE:8018076 | 134.23 | N | N | Y |
| BC140514 | *CSDE1* | Cold Shock Domain Containing E1, Rna-Binding | 134.098 | Y | Y | Y |
| BC102325 | *ARL4A* | Adp-Ribosylation Factor-Like 4A | 133.928 | N | N | N |
| AF013213 | *LOC789867* | Eukaryotic Translation Elongation Factor 1 Alpha 1 | 133.722 | Y | Y | Y |
| BC119912 | *C22H3ORF19* | Chromosome 3 Open Reading Frame 19 Ortholog | 132.945 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| BC114202 | *SON* | Son Dna Binding Protein | 132.028 | Y | Y | Y |
| BC114038 | *LOC540061* | Hypothetical Loc540061 | 131.713 | Y | Y | Y |
| BC109696 | *Image:8061225* | Bos Taurus Cdna Clone Image:8061225 | 131.694 | N | N | Y |
| DQ347583 | *myl6* | Myosin, Light Chain 6, Alkali, Smooth Muscle And Non-Muscle | 131.028 | Y | Y | Y |
| BC112612 | *Dnajc5b* | Dnaj (Hsp40) Homolog, Subfamily C, Member 5 Beta | 130.665 | Y | Y | Y |
| X64897 | *Ndufa4* | Nadh Dehydrogenase (Ubiquinone) 1 Alpha Subcomplex, 4, 9Kda | 130.65 | N | N | Y |
| BC109715 | *MGC:134272* | Bos taurus cDNA clone MGC:134272 | 130.387 | N | N | N |
| X64898 | *LOC781609* | Nadh Dehydrogenase (Ubiquinone) 1 Beta Subcomplex, 4, 15Kda | 130.066 | Y | Y | Y |
| AY911370 | *LOC786773* | Ribosomal Protein L26 | 129.524 | Y | Y | Y |
| BC114805 | *SPATA3* | Spermatogenesis Associated 3 | 129.437 | N | N | Y |
| BC102311 | *FILIP1L* | Filamin A Interacting Protein 1-Like | 129.251 | N | N | Y |
| AY911320 | *Cox7c* | Cytochrome C Oxidase Subunit Viic | 128.463 | Y | Y | Y |
| BT021019 | *Naca* | Nascent Polypeptide-Associated Complex Alpha Subunit; Similar To Nascent-Polypeptide-Associated Complex Alpha Polypeptide | 127.487 | Y | Y | Y |
| AF100170 | *AKAP4* | A Kinase (Prka) Anchor Protein 4 | 126.623 | Y | Y | Y |
| BC120019 | *MLLT11* | Myeloid/Lymphoid Or Mixed-Lineage Leukemia (Trithorax Homolog, Drosophila); Translocated To, 11 | 126.453 | N | Y | N |
| BC122782 | *LOC781500* | Hypothetical Protein Loc781500 | 126.357 | N | N | Y |
| BT030513 | *Rpn2* | Ribophorin Ii | 126.355 | Y | Y | Y |
| AB099075 | *LOC784061* | Similar To 60S Ribosomal Protein L21; Similar To Ribosomal Protein L21; Ribosomal Protein L21 | 125.979 | Y | Y | Y |
| BC120104 | *C1H3orf38* | Chromosome 3 Open Reading Frame 38 Ortholog | 125.625 | Y | Y | Y |
| BC140633 | *IMAGE:8190785* | Bos taurus platelet-activating factor acetylhydrolase, isoform Ib, alpha subunit 45kDa, mRNA | 125.591 | Y | Y | Y |
| BC105172 | *STON1-GTF2A1L* | Ston1-Gtf2A1L Readthrough Transcript | 125.133 | Y | Y | Y |
| BC102090 | *rps3* | Ribosomal Protein S3 | 124.907 | N | N | N |
| AB099047 | *LOC531679* | Ribosomal Protein 17-Like | 123.815 | N | Y | N |
| AB434936 | *TERF2* | Telomeric Repeat Binding Factor 2 | 123.541 | Y | Y | Y |
| BC108202 | *ube2b* | Ubiquitin-Conjugating Enzyme E2B (Rad6 Homolog) | 123.39 | N | N | Y |
| BC109731 | *C16H1orf100* | Chromosome 1 Open Reading Frame 100 Ortholog | 123.238 | Y | Y | Y |
| BC111270 | *srp54* | Signal Recognition Particle 54Kda | 123.224 | Y | Y | Y |
| BC126821 | *Upf2* | Upf2 Regulator Of Nonsense Transcripts Homolog (Yeast) | 122.278 | N | N | N |

| | | | | | | |
|---|---|---|---|---|---|---|
| AF144764 | *timp2* | Timp Metallopeptidase Inhibitor 2 | 121.825 | N | N | N |
| BC102670 | *MGC127695* | Hypothetical Protein Mgc127695 | 121.633 | Y | Y | Y |
| BC148013 | *RPL14* | Bos Taurus Ribosomal Protein L14 | 120.553 | Y | Y | Y |
| BC112887 | *IMAGE:8009582* | Bos taurus ribosomal protein S27 (metallopanstimulin 1), mRNA | 120.098 | N | N | N |
| BC110187 | *fhl5* | Four And A Half Lim Domains 5 | 119.147 | Y | Y | Y |
| BC116058 | *KCMF1* | Potassium Channel Modulatory Factor 1 | 118.853 | N | N | Y |
| DQ347593 | *LOC781370* | Ferritin, Heavy Polypeptide 1; Similar To Ferritin Heavy Chain; Similar To Ferritin, Heavy Polypeptide 1 | 117.121 | Y | Y | Y |
| AB098931 | *rps8* | Ribosomal Protein S8 | 116.998 | N | N | Y |
| BC109560 | *LOC784487* | Ribosomal Protein L7; Similar To Ribosomal Protein L7; Similar To 60S Ribosomal Protein L7 | 116.859 | N | Y | N |
| BC109745 | *LOC528549* | Similar To Dnaj (Hsp40) Homolog, Subfamily B, Member 3 | 116.751 | N | N | N |
| BC150005 | *LSM2* | Lsm2 Homolog, U6 Small Nuclear Rna Associated (S. Cerevisiae) | 116.734 | N | N | Y |
| BC111147 | *LOC786899* | Similar To Gtpase Activating Protein Testicular Gap1; Hypothetical Loc786899; Hypothetical Protein Mgc134093 | 116.682 | Y | Y | Y |
| BC103454 | *mrps36* | Mitochondrial Ribosomal Protein S36 | 116.104 | Y | Y | Y |
| AB098752 | *LOC782525* | Eukaryotic Translation Elongation Factor 1 Gamma; Similar To Eukaryotic Translation Elongation Factor 1 Gamma | 116.059 | Y | Y | Y |
| BC110254 | *TES* | Testis Derived Transcript (3 Lim Domains); Similar To Testis Derived Transcript | 115.614 | N | N | N |
| BC111209 | *pdhA2* | Pyruvate Dehydrogenase (Lipoamide) Alpha 2 | 114.458 | N | N | Y |
| BC109677 | *FXR1* | Fragile X Mental Retardation, Autosomal Homolog 1 | 114.322 | Y | Y | Y |
| BC116167 | *FAIM2* | Fas Apoptotic Inhibitory Molecule 2 | 112.984 | N | N | Y |
| BC112616 | *Trim59* | Hypothetical Loc540154 | 112.836 | Y | Y | Y |
| BC102601 | *ropn1l* | Ropporin 1-Like | 112.278 | N | N | Y |
| BC105363 | *YBX1* | Y Box Binding Protein 1 | 112.262 | N | N | Y |
| BC146060 | *THAP7* | Tubulin, Alpha 1A; Tubulin, Alpha 1B; Similar To Alpha-Tubulin I; Thap Domain Containing 7 | 112.181 | N | N | N |
| AF541971 | *DDX4* | Dead (Asp-Glu-Ala-Asp) Box Polypeptide 4 | 111.38 | Y | Y | Y |
| BC114188 | *LOC507141* | Ce5 Protein-Like | 110.65 | Y | Y | Y |
| BC108246 | *MGC133632* | Hypothetical Protein Loc614279 | 110.389 | N | N | Y |
| BC110170 | *CSNK2B* | Casein Kinase 2, Beta Polypeptide | 110.244 | N | N | Y |
| BC109563 | *TRYX3* | Trypsin X3 | 110.204 | Y | Y | Y |
| BC102081 | *DAD1* | Defender Against Cell Death 1 | 110.078 | Y | Y | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| X55389 | *F1-ATPase* | mRNA for F1-ATPase gamma-subunit | 110.066 | N | N | Y |
| BC109599 | *ADORA3* | Adenosine A3 Receptor | 109.351 | Y | Y | Y |
| BC102328 | *SNRPB2* | Small Nuclear Ribonucleoprotein Polypeptide B" | 109.087 | N | N | Y |
| BC109851 | *Asb17* | Ankyrin Repeat And Socs Box-Containing 17 | 108.906 | Y | Y | Y |
| BC108243 | *NDUFA5* | Nadh Dehydrogenase (Ubiquinone) 1 Alpha Subcomplex, 5, 13Kda | 108.605 | Y | Y | Y |
| AY911358 | *LOC781565* | Ribosomal Protein S6 | 108.563 | Y | Y | Y |
| BC110237 | *Mlec* | Malectin | 108.544 | N | N | Y |
| BC110212 | *LOC786673* | Atp Synthase, H+ Transporting, Mitochondrial F0 Complex, Subunit B1 | 108.505 | Y | Y | Y |
| BC109625 | *Gkap1* | G Kinase Anchoring Protein 1 | 108.462 | Y | Y | Y |
| BC140615 | *ADAM3A* | Adam Metallopeptidase Domain 3A (Cyritestin 1) | 108.146 | Y | Y | Y |
| BC102873 | *Fau* | Finkel-Biskis-Reilly Murine Sarcoma Virus (Fbr-Musv) Ubiquitously Expressed; Similar To Ubiquitin-Like/S30 Ribosomal Fusion Protein | 108.13 | N | N | N |
| BC102656 | *IMAGE:30956887* | Bos taurus pituitary tumor-transforming 1, mRNA | 107.717 | Y | Y | Y |
| AB373012 | *CYP1B1* | Cytochrome P450, Family 1, Subfamily B, Polypeptide 1 | 107.159 | Y | Y | Y |
| BC102135 | *BZW1* | Basic Leucine Zipper And W2 Domains 1 | 105.11 | Y | Y | Y |
| BT030749 | *LOC506261* | Similar To 14-3-3 Protein Theta (14-3-3 Protein Tau) (14-3-3 Protein T-Cell) (Protein Hs1) | 105.009 | N | N | Y |
| BC111628 | *IMAGE:8019171* | Bos taurus cDNA clone IMAGE:8019171 | 104.983 | Y | Y | Y |
| BC109721 | *INSL6* | Bos taurus insulin-like 6, | 104.965 | N | N | Y |
| BC126695 | *KLHL10* | Kelch-Like 10 (Drosophila) | 104.86 | N | N | Y |
| BC112511 | *VTI1B* | Vesicle Transport Through Interaction With T-Snares Homolog 1B | 103.334 | N | N | Y |
| BC102885 | *Paip2* | Poly(A) Binding Protein Interacting Protein 2 | 103.312 | N | N | N |
| AB098765 | *FTH1* | mRNA for similar to ferritin H subunit | 103.092 | N | N | Y |
| BC108215 | *RTF1* | Rtf1, Paf1/RNA polymerase II complex component | 102.78 | N | N | Y |
| BC111179 | *PSMG2* | Proteasome (prosome, macropain) assembly chaperone 2 | 102.639 | N | Y | N |
| BC133582 | *cnot1* | Ccr4-Not Transcription Complex, Subunit 1 | 102.582 | Y | Y | Y |
| BC109747 | *Hemgn* | Hemogen | 102.453 | Y | Y | Y |
| AB098753 | *LOC781609* | Similar To B15 Subunit Of The NADH | 102.178 | Y | Y | Y |
| BC102499 | *naa38* | Lsm8 Homolog, U6 Small Nuclear Rna Associated (S. Cerevisiae) | 100.954 | N | N | Y |
| BC109698 | *FUNDC2* | FUN14 domain containing 2 | 100.902 | N | N | Y |
| HQ423186 | *BBD126* | Bos Taurus Beta-Defensin 126 Mrna | 100.859 | Y | Y | Y |
| AF307320 | *RPS28* | ribosomal protein S28-like protein mRNA | 100.799 | N | N | N |
| BC109495 | *WDR61* | WD repeat domain 61 | 100.679 | N | N | Y |
| BC102453 | *STMN1* | Stathmin 1/oncoprotein 18 | 100.645 | N | N | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| BC126794 | *LYZL1* | Lysozyme-Like 2 | 100.291 | Y | Y | Y |
| BC112727 | *CCDC91* | Coiled-coil domain containing 91 | 100.177 | N | N | Y |

## Chapter 3

PREPARED FOR SUBMISSION TO JOURNAL OF DAIRY SCIENCE

# Spermatozoal transcript profiles differ between lower and higher fertility sires

C. J. Card,* E. J. Anderson,* K. E. Krieger,** M. Kaproth,** and B. L. Sartini*

*Department of Fisheries, Animal and Veterinary Science, University of Rhode Island, Kingston, Rhode Island 02881
**Genex Cooperative Inc., Shawano, WI 54166.

[1]Corresponding author: BL Sartini, 179 Center for Biotechnology and Life Sciences, 120 Flagg Road, University of Rhode Island, Kingston, RI 02886, Phone: 401-874-2667, FAX: 401-874-7575, Email: blsartini@uri.edu

**Short Title: Spermatozoal bovine transcript profiles differ with fertility**

**Grant Support:** 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.

**Abstract**

Spermatozoal mRNA transcripts can reflect spermatogenic gene expression and therefore may have potential as accurate markers for sire fertility. The goal of this study was compare the transcript profiles of lower fertility (Conception Rate scores (CR) -2.9 to -0.4) and higher fertility (CR 1.8 to 3.5) sire spermatozoa using Ribonucleic Acid Sequencing (RNA-Seq). A total of 5,366 transcripts and 3,227 transcripts were identified in the lower and higher fertility populations respectively. Common transcripts between the two populations were identified (2,422 transcripts) while 2,944 transcripts were unique to the lower fertility populations and 805 transcripts were unique to the higher fertility population. Gene ontology analysis for transcripts common between the populations revealed the most represented Biological Processes (BP) were translation, protein localization, and proteolysis while categories differed for transcripts unique to each population. Differential expression of transcripts between the two populations identified 36 transcripts that were uniquely expressed between the lower and higher fertility populations (Cuffdiff analysis) and 432 transcripts with elevated expression in the lower fertility sires compared to 141 transcripts with elevated expression levels in the higher fertility sires (Ratio analysis). Candidate fertility spermatozoal transcripts were identified and four transcripts, *GTSF1, PRM2, TNP1* and *COX7C* have a negative correlation with sire fertility. A number of spermatozoal transcripts varied across individual sires. 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**

Male subfertility is a contributing factor to the reproductive efficiency of livestock production. The range of dairy sire fertility exceeds the variation in fertility measured in females suggesting that male fertility can be a critical factor in both low and high herd fertility. Accurate sire ranking by fertility is difficult with current fertility assays (Flowers, 2013). Conventional methods of evaluating the fertilizing potential of bull semen typically measure compensable traits such as motility, morphology, and concentration, that can be compensated for in artificial insemination (AI) by increasing sperm numbers. However, these evaluations are inadequate for assessing the variation in fertility typically found among individuals (Foote, 2003; Kastelic, 2013). Conventional semen analysis can often detect sub-fertile animals but is unsuccessful in detecting differences in fertility amongst individuals with average and above average fertility (Gadea et al, 2004; Kastelic and Thundathil, 2008). Inclusion of uncompensable sperm traits, traits that impact successful fertilization and early embryonic development, such as DNA fragmentation, in multi-parameter sire fertility assays have improved the accuracy of sire fertility rankings (Omran et al, 2013). Although fertility assays have been improved, *in vivo* fertility scores have not been mirrored by *in vitro* assays to date. It is likely that a number of traits need to be evaluated concurrently to get an accurate assessment of sire fertility (Omran et al, 2013). More precise fertility tests incorporating additional sperm traits are necessary to help ensure that the most fertile males are used for AI to improve herd pregnancy rates.

In addition to the paternal genome, transcriptionally- and translationally-silent spermatozoa also carry a subset of RNA that contain coding and non-coding RNAs as

well as rRNAs (Hamatani, 2012; Card and Anderson et al., 2013; Sendler et al, 2013). Estimates from various species have identified approximately 6,000 mRNAs (nuclear and mitochondrial) and ribosomal RNAs in the spermatozoal transcript profile (Ostermeier et al., 2002; Gilbert et al., 2007; Das et al., 2010; Card and Anderson et al., 2013; Das et al., 2013; Sendler et al., 2013). Although most transcripts appear to be degraded, full-length sperm mRNAs with the potential to be translated into functional protein are also present (Ostermeier et al., 2005; Card and Anderson et al., 2013). These full-length mRNAs are interesting targets for further study as they are potentially critical for early embryonic development.

Due to the lack of transcription in spermatozoa, the transcripts present represent spermatogenic gene expression including past efficiencies and therefore may be useful as a "fingerprint" of past spermatogenic events (Ostermeier et al., 2002; Lambard et al., 2004; Ostermeier et al., 2005; Platts et al., 2007; Lalancette et al., 2009). Spermatozoal transcripts associated with the ubiquitin-proteosome pathway, acrosome formation, and proper tail formation have different expression levels between normal semen and teratozoospermic samples (Platts et al., 2007). Spermatid and spermatocyte specific transcripts are also absent in teratozoospermatozic samples thus indicating errors in the later stages of spermatogenesis (Platts et al., 2007). The spermatozoal RNA population can reveal the success or failure of past spermatogenic events therefore errors in spermatogenesis could be detected through transcript profiling of spermatozoa, aiding in fertility assessment.

Previously, differences in bovine spermatozoal transcripts between sires with different fertility have been reported (Gilbert et al., 2007; Lalancette et al., 2008;

Feugang et al., 2010). Differentially expressed transcripts have been identified when utilizing hybridization-based microarray analysis to compare spermatozoal transcripts from higher and lower fertility bulls (Gilbert et al., 2007; Lalancette et al., 2008; Feugang et al., 2010). Additionally, individual spermatozoal transcripts have been individually correlated with sire fertility. A positive correlation with fertility was found with *AK1, IB5, DOPPEL, NGF, TIMP2, PLCZ1, CRISP2,* and *PEBP1* while *CCT8* and *PRM2* have a negative correlation with fertility (Arangasamy et al., 2011; Lalancette et al., 2008; Kasimanickam et al., 2012). Due to the limitations of the microarray approach that lead to the identification these individual transcripts, the comparison of the spermatozoa transcript profiles among sires of different fertility is incomplete.

The objective of this study was to sequence and compare the complete spermatozoal transcript profiles of lower and higher fertility sires with RNA-Seq. RNA-Seq has several advantages over hybridization-based methods, including microarrays, namely the ability to accurately assess absolute transcript levels, full-length transcripts, and novel splicing/exon discovery (Wang et al., 2009; Werner, 2011). It is hypothesized that the composition of spermatozoal transcript profiles will differ with fertility and select transcripts will show correlations with fertility score. This study will provide a greater understanding of the potential utility of 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 bulls (-2.9 to -0.4 CR) was used to create the lower fertility amplified cDNA pool and four separate bulls (1.8 to 3.5 CR) were used for the higher fertility pooled sample submitted for RNA-Seq and subsequent validation by qPCR. For variation of selected transcripts in individual sires, spermatozoal RNA, from 9 additional bulls with varying CR scores (-6.7, -5.5, -3.5, -3.4, -2.7, 0.1, 0.8, 2.2, 3.6), was used.

RNA Isolation

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

Double-Stranded cDNA (ds-cDNA) Synthesis and Amplification

Equal amounts of spermatozoal RNA from individual sires (to ensure equal representation of all sires in the pool) was combined to generate a lower fertility pool and a higher fertility pool that were converted to ds-cDNA and linearly amplified for RNA-Seq and qPCR validation (SMARTer Pico PCR cDNA Synthesis Kit; Clontech,

Mountain View, CA). During cDNA amplification, cycle optimization was performed to maintain amplification in the linear phase. For lower and higher fertility pooled ds-cDNA samples, 5 µg of each sample was submitted for RNA-Seq. Additional ds-cDNA from each pool was used in qPCR validation analysis described below.

RNA-Seq and Analysis

Lower and higher fertility spermatozoal ds-cDNA 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) via Tophat was performed using Galaxy (Giardine et al., 2005; Blankenberg et al., 2010; Langmead et al., 2009; 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 61,059 included removed due to being adapter only reads, reads < 15 nucleotides, and 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. This removed 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 their 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 SRA#XXXXX (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.

Differential Expression Analysis

Differential expression analysis was performed using Cuffdiff (Trapnell et al, 2012) with a false discovery rate of 5% and a significance value of $p < 0.05$. Alternate differential expression analysis (Ratio analysis) examined lower fertility:higher fertility FPKM ratios and vice versa. The parameters that were required for a transcript to be considered differentially expressed between the populations include FPKM of $\geq 5$ in both populations and a minimum 2:1 ratio between FPKM of the two populations.

Polymerase Chain Reaction (PCR)

Select fertility candidate transcripts were analyzed via PCR on the lower and higher fertility pools and 9 individual bulls with variable CR scores to determine individual variation in transcript presence. PCR cycling conditions are outlined in (Card and Anderson et al, 2013) and PCR primers can be found in Table 1. Along with

spermatozoal RNA samples, a testis RNA positive control and a no template negative control were included in each experiment.

qPCR

qPCR was performed to validate RNA-Seq expression levels (FPKM) and transcripts identified by Ratio analysis, as well as to compare transcript levels among individual sires for fertility correlation analysis. Seven transcripts with FPKM ranges from (12 - 42,275 for higher fertility; 58 - 28,826 for lower fertility) were run for validation qPCRs. For FPKM validation qPCR analysis, standard curves were made for all transcripts by serial dilutions of purified cDNA into 8 standard concentrations from $1x10^9$ to $1x10^2$ copies/ul. qPCR reactions were run 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. All correlations were determined using linear regression analysis using GraphPad Prism 6.0.

**Results**

***RNA-Seq***

Sequencing of the higher fertility spermatozoal RNA resulted in 11,832,830 x 2 paired-end 100-bp reads while lower fertility spermatozoal RNA sequenced 11,434,525 x 2 paired-end 100-bp reads. Read and genome mapping statistics can be found in Table 2.

### *High and Low Fertility Transcript Profiles*

A total of 5,366 transcripts were expressed in lower fertility sires while 3,227 transcripts were expressed in higher fertility sires (FPKM > 0). Common transcripts between the lower and higher fertility populations were identified (2,422 transcripts) while 2,944 transcripts were unique to the lower fertility population and 805 transcripts were unique to the higher fertility population (Figure 1). The qPCR expression levels showed a high correlation with FPKM values for both high fertility ($r^2$ = .9646) and low fertility ($r^2$ = .9485) populations (Figure 2). The top 10 transcripts expressed, unique to high and low fertility populations excluding mitochondrial transcripts can be found in Tables 3 and 4.

### *Gene Ontology Analysis*

Gene ontology analysis was carried out for transcripts common between the two fertility groups as well as transcripts unique to each group (Table 5). It is important to note that an individual transcript can be classified into multiple ontological categories during this analysis. When examining transcripts common to the lower and higher fertility populations (2,422 transcripts), 224 BP, 105 CC, and 84 MF categories were represented. The top three BP categories represented were translation, protein localization, and proteolysis. The top categories differed for transcripts unique to the lower and higher fertility transcript profiles. Transcripts unique to the lower fertility population (2,930 transcripts), there were 284 BP, 92 CC, and 101 MF categories represented. The top three BP categories represented were transcription, proteolysis, and

phosphate metabolic process. Transcripts unique to the higher fertility population (798 transcripts) had a total of 148 BP, 74 CC, and 47 MF categories represented. The top three BP represented in this population were oxidation reduction, intracellular signaling cascade, and protein localization.

***Differential Expression Analysis***

Two analysis methods were conducted to identify transcripts differentially expressed between the lower and higher fertility spermatozoal RNA populations that would be potential fertility candidates. Using Cuffdiff, 36 transcripts were differentially expressed between fertility groups (Table 6). Five transcripts were chosen (*DDX4*, *DDX20*, *PCNA*, *G2E3*, *PSMA1*; all present in low fertility only) and were validated via PCR (Figure 3). The Ratio analysis revealed 432 transcripts with greater expression in the lower fertility population and 142 transcripts with greater expression in the higher fertility population (Figure 4, Appendix 1 and Appendix 2). Eight transcripts were chosen from this analysis and five of the eight transcripts were validated by showing a fold change consistent with FPKM ratios.

***Individual sire variation of spermatozoal mRNA presence***

To determine if spermatozoal transcript presence varied among sires, candidate fertility transcripts (n=27) were run on 9 individual bulls across a range of CR scores -6.7 to 3.6 (Table 7). These transcripts were identified from differential expression analysis (Cuffdiff and Ratio analysis) and previously published spermatozoal transcripts. Only one transcript, *PRM1*, was present in all 9 individuals. Two transcripts, *PRM2* and

*TNP1*, were amplified in 8 of the 9 bulls. Eleven transcripts were present in 5 or more bulls while the other 16 transcripts were present in less than half of the bulls. The bull with the most transcripts present was the lowest fertility bull that had 17 of the 27 transcripts. Transcripts identified from Cuffdiff analysis (n=3; *APOABP1*, *G2E3* and *PCNA*) were only present in 1-2 bulls. One transcript identified by ratio analysis, *COX7C* was present in 7 of the 9 bulls tested while *ATP5J* and *GPX4* were only present in 1 and 2 bulls respectively. Three transcripts, *AKAP4*, *APOABP1*, and *BRP* were not amplified in any bull spermatozoa.

***Spermatozoal RNAs identified from previous publications***

The spermatozoal transcript profiles for both fertility populations were mined for the presence of transcripts previously reported to be associated with bull fertility (Arangasamy et al, 2007; Feugang et al, 2010; Kasimanickam et al, 2012; Kasimanickam et al, 2013). Only a few of the transcripts examined were identified in the transcript profiles reported here including *PRM1* and *CRISP2* (Table 8). Additionally, the transcripts for spermatozoa fertility proteins were also identified in the transcript profiles; many transcripts for these proteins, including *PSMA6* and *ROPN1*, were found in both populations with the lower fertility population typically having greater expression FPKM values (Table 9).

***Individual spermatozoal transcripts and sire fertility***

To determine if amount of spermatozoal transcripts is correlated with sire fertility, 12 transcripts (identified from Cuffdiff analysis, Ratio analysis and previous

publications) were amplified in an independent population of 9 individual bull spermatozoal ds-cDNA samples with fertility scores ranging from CR -6.7 to 3.6 (Table 10). Four out of eleven transcripts, *PRM2*, *TNP1*, *COX7C*, and *GTSF1* have a negative correlation with fertility. The other 7 transcripts, *PRM1*, *PSMA1*, *PSMA6*, *ACRV1*, *SPAG4*, *ATP5J*, and *GPX4* showed no significant correlations with fertility.

**Discussion**

Overall, high-throughput sequencing of spermatozoal transcript profiles for lower and higher fertility bulls resulted in transcript profiles with several common transcripts (2,422 transcripts) between the two populations. In previous studies, over 1,700 transcripts have been shown to be common in spermatozoal RNA populations amongst 23 fertile human subjects while 96 and 37 transcripts were common between high and medium quality semen samples and medium and poor quality semen samples respectively (Lalancette et al., 2009; Ostermeier et al., 2005). When comparing 9 pooled individual ejaculates (3,281 transcripts) and one individual (2,780 transcripts), all but four transcripts from the individual were found in the pooled sample (Ostermeier et al., 2002). Transcripts that are present across individuals and populations are considered stably regulated, though their expression levels may vary.

Though there are a number of stably regulated transcripts between the two populations reported here, there remain a number of differences between fertility groups as well, including a number of transcripts unique to each population, differences in

expression levels, and variability in prominent gene ontology categories represented by each population. These differences between fertility groups are consistent with previous studies in subfertile males using microarrays (Lalancette et al, 2008; Lalancette et al, 2009; Feugang et al, 2010). The lower fertility population had a much larger transcript profile (greater than 2,000 more transcripts) than the higher fertility populations. The greater number of transcripts in the lower fertility population could be accounted for by inefficiencies in translation during spermatogenesis leading to missing essential proteins for further development of critical mRNAs/proteins needed for proper fertilization or early embryo development (Jodar et al, 2013).

In transcripts unique to higher fertility sires there was a prevalence of transcription and RNA processing as well as cell proliferation based transcripts. The highly represented transcripts unique to the lower fertility population were transcripts associated with phosphate metabolic processes and a group of categories involving protein production/processing. A focus on transcription and RNA processing unique to higher fertility population could have an additive effect to common transcripts in making sure that an excess of essential RNAs are produced in early stages of spermatogenesis. These RNAs might be necessary to protein production as sperm develop or required in the egg at fertilization or during early embryonic development. Having a much larger focus on protein processing in the low fertility population could be indicative of greater inefficiency in protein formation in spermatozoa as many transcripts appear to be leftover that may have been essential for protein formation. Though transcripts are involved in translation and related processes, some overarching error in translational machinery could

be inhibiting these from being translated and usable to aid further protein formation/processing.

Two methods of differential transcript expression were validated and identified approximately 600 transcripts that differed in expression between the two fertility populations. Further validation of transcripts identified with the Cuffdiff analysis found that most transcripts were only expressed in a few bulls and of the transcripts analyzed, none were correlated with fertility. This is likely due to the fact that Cuffdiff analysis for this dataset only reported transcripts unique to each population, thus only represented by a maximum of 4 out of 8 bulls that were in the RNA-Seq samples. To determine if a difference in transcript abundance was present with transcripts expressed in both the lower and higher fertility populations, we also compared expression levels by Ratio analysis. Transcripts chosen from this analysis were tested in 9 individual sires and some transcripts were only expressed in a few bulls (*SPAG4*, *ATP5J* and *GPX4*) while others (*PSMA6*, *COX7C* and *PRM1*) were expressed in most, if not all, sires analyzed. Both analysis methods identified further fertility transcript candidates and more can be extracted from the datasets for presence and quantification analysis.

Significant variation of transcript presence among sires was prevalent with several transcripts only being expressed in a few individuals. While some transcripts were conserved, individual variation in transcript presence exists. Many spermatozoal transcripts are found as stably regulated, in terms of being commonly expressed, across individuals while there remains a lot of individual variation in terms of expression quantity as well as some transcripts that are not expressed in all individuals (Ostermeier et al, 2005; Lalancette et al, 2009). Of the 27 transcripts tested for presence by PCR

analysis, only one transcript (*PRM1*) was expressed in all individuals. It has been shown that transcript expression level differences can be indicative in fertility but it is evident with these results that there is potential to compare individuals for transcript presence for those transcripts that are not consistent across the population. There is a great deal of interest in stably regulated transcripts, such as *PRM1* in this study, because any transcripts that are present in all individuals that show a correlation with fertility are candidates for a potential fertility assay. To be able to detect subtle differences in fertility amongst sires, one must be able to test expression levels for a transcript that is present in all individuals. *PRM1* did not show a correlation with fertility when compared across 9 individuals but there is the potential for other stably regulated transcripts in our differential expression analysis that could be further explored for possible correlations with fertility. Though transcripts present in only some bulls can offer insight into fertility, a stably regulated transcript that shows a very high correlation with fertility score would be optimal for ranking individuals across an entire population.

Other than *PRM1*, ten more transcripts were chosen from differential expression analysis methods and previous literature for qPCR analysis to examine possible correlations with fertility (Table 10). Four of these genes, *TNP1*, *GTSF1*, *PRM2*, and *COX7C* have a negative correlation with fertility. Other transcripts tested, *PSMA1*, *PSMA6*, *ACRV1*, *SPAG4*, *ATP5J*, and *GPX4* did not have significant linear correlations with sire fertility. Many of these transcripts that did not show correlations with fertility were not present, or present at very low levels, in some of the individuals. Cytochrome oxidase subunit 7c (COX7C) is one of many subunits involved in the terminal step in the electron transport chain and has been found in the inner mitochondrial membrane, aiding

in the synthesis of ATP (Seelan et al, 1997). We hypothesize that the correlation with fertility may be an indication of sperm mitochondrial function; in turn potentially have an effect on motility. Transition protein 1 (TNP1) is one of two transition proteins that replace the histones that DNA is bound to during spermatogenesis before protamines replace the transition proteins to compact the DNA much tighter than originally via histones (Oliva, 2006). Correlation of TNP1 with fertility may indicate inefficient translation, and therefore reduced DNA compaction, during spermatogenesis. Gametocyte-specific factor 1 (GTSF1) has been found in murine testes and knockout studies to be essential for spermatogenesis as a retrotransposon suppressor (Yoshimura et al., 2009).

A number of transcripts have shown positive correlations of the mRNA levels and fertility in previous studies including *AK1*, *IB5*, *DOPPEL*, *NGF*, *TIMP2*, *PLCZ1*, *CRISP2*, *PEBP1, CD36, CENPA, ADIPO, ADIPOR1,* and *ADIPOR2* (Feugang et al., 2010; Arangasamy et al., 2011; Kasimanickam et al., 2012; Kasimanickam et al., 2013). Select transcripts including *CCT8* and *PRM2* have negative correlations with fertility in bulls in previous publications (Lalancette et al., 2008; Arangasamy et al., 2011). In this study, none of the previously identified differentially expressed transcripts were identified with the Cuffdiff analysis. Three previously identified transcripts (*PLCZ1*, *CRISP2*, *PEBP1*) were identified by the Ratio analysis. However, it is important to keep in mind factors that differed among studies including different isolation/amplification methods, a different population of bulls, and differences in fertility ranges amongst the populations. As we have seen in this study, variation in transcript presence or amount is prevalent between individual bulls, which could lead to differences between our data and

previously published data as individuals studied as well as methods utilized differ between studies. PLCZ1 is crucial in initiation of the $Ca^{2+}$ waves in the oocytes at fertilization that triggers the resumption of meiosis, initiating early embryonic development. This transcript is particularly interesting because it is spermatozoa-specific but has been shown to be transferred to the oocyte at fertilization and even has been found to be translated post-fertilization (Swann et al., 2006). Our data suggests no significant correlation with *PLCZ1* and fertility but this is a candidate that shows some promise moving forward in looking at fertilizing potential of spermatozoa. Both CRISP2 and PEBP1 are associated with spermatozoa capacitation and could contribute to the spermatozoal RNA population predicting sperm function (Arangasamy et al, 2010).

A number of other spermatozoa mRNAs have been explored for having potential links to fertility in other species. Though a potential link to fertility has been hypothesized in humans, these transcripts were not expressed in sire fertility spermatozoal transcript profiles. Both PSG1 and HLA-E were found to have mRNA levels significantly higher in fertile men than infertile men (Avendano et al., 2009). Proteins for these genes were not found in spermatozoa, however the transcripts were found to be present in the zygote more than 24 hours after fertilization. PSG1 is hypothesized to aid in protection from the maternal immune system while HLA-E helps avoid death from NK cells. Both of these transcripts could be key for early embryonic development and maintenance of pregnancy in humans.

Although the Cuffdiff and Ratio analysis did not identify *PRM1* and *PRM2* as differentially expressed between the two fertility populations, we investigated these transcripts further because other studies have found that protamine mRNA and protein

levels had positive correlations with spermatozoa concentration, motility, fertilization potential, and embryo quality (Kempisty et al., 2007; Depa-Martynow et al., 2012). During spermatogenesis, the protamines replace the histones and compact DNA tighter. This in turn, protects the DNA from any damage as well as halting transcription. Improper compaction may leave DNA open to damage, which can cause various defects including the inability to fertilize. Infertile males have increased levels of protamine mRNA and decreased protamine protein levels as well as abnormal protamine 1/protamine 2 ratios (Aoki et al., 2006a; Aoki et al., 2006b). This result offers the possibility that improper translation of protamines leads to spermatozoa that do not offer the ability to fertilize due to proper compaction of the DNA.

In this study, the *PRM1* levels were increased in the higher fertility population (FPKM = 42276) versus the lower fertility population (FPKM = 28826). The same trend was evident with *PRM2* with the higher fertility (FPKM = 98) have slightly higher levels than the lower fertility population (FPKM = 78). Though neither of these appeared in either differential expression analysis method, these trends agree with most previously published data in that protamine transcript levels are higher in individuals of higher fertility. Both protamines were further examined utilizing qPCR on 9 individuals with a range of fertility scores via linear correlation analysis. *PRM1* showed no correlation ($r^2$ = 0.36; p = 0.08) to CR score while *PRM2* showed a negative correlation with CR score ($r^2$ = 0.60; p = 0.01). *PRM2* has previously been shown to have higher expression levels in lower fertility bulls (Lalancette et al, 2008), which is in agreement with these results. This is an interesting candidate to show a connection with fertility, as PRM2 protein is not found in bulls. It appears that protamines have the potential to show the ability to

predict fertility based on this and previous studies, but further exploration is needed, especially in *PRM2* levels.

Some proteomic studies have been performed to evaluate protein differences amongst bulls of varying fertility. Higher fertility animals have higher protein expression levels for proteins associated with spermatogenesis, cell motility, energy metabolism, and cell communication (Peddinti et al., 2008). A number of proteins are expressed at higher levels in high fertility bulls including ATPSB, ENO1, ASPP2, GPX4, AHSG, AKI, and PEBP1. Some proteins, on the other hand, have lower expression levels in high fertility, including CCT5, CCT8, ELSPBP1, PSMA6, BSP1, VDAC1, UQCRC2, and ROPPORIN-1 (D'Amours et al., 2010; Park et al., 2012). In this study, a number of the transcripts for these proteins were present in the lower and higher fertility populations (Table 8). All of the transcripts that were present in both populations were at higher levels in the lower fertility population. Those transcripts that are at increased levels in the lower fertility population in this study and have been shown to be increased protein levels in high fertility animals, such as GPX4, could be an example of a transcript that undergoes incomplete or insufficient translation in lower fertility animals. Coupling RNA-Seq data with proteomic data can hopefully provide more insight into silencing or translation of transcripts in sires with varying fertility ratings.

Spermatozoa RNA has been proposed as a potential diagnostic tool for fertility due to its representation of proper spermatogenesis but it is also hypothesized to play a role in early embryonic development. Spermatozoal transcripts can be delivered to the oocyte at fertilization and persist for hours in the zygote (Ostermeier et al., 2004). *PLCZ1* is the only transcript thus far that has proven to be translated post-fertilization in

the zygote (Swann et al., 2006). This shows how there is the potential for full-length transcripts that are delivered by spermatozoa to have function post-fertilization. A number of full-length transcripts have been found in spermatozoa showing the potential for some of these to have early embryonic function if they are to be translated into protein post-fertilization (Ostermeier et al., 2005; Card and Anderson et al., 2013).

This study is the first to directly sequence and compare the complete spermatozoal transcript profiles of sires with different fertility. Using this approach, similarities and differences in transcript presence were identified and a correlation of individual transcripts with fertility was found. This data offers insight into a number of potential candidates for further investigation as potential markers for measuring fertility. Although fertility assays have been improved, *in vivo* fertility scores have not been mirrored by *in vitro* assays to date. It is likely that a number of traits need to be evaluated concurrently to get an accurate assessment of sire fertility (Omran et al, 2013). Inclusion of spermatozoal RNA abundance in multi-parameter sire fertility assays may improve accuracy of sire fertility rankings to ensure that the most fertile sires are used for AI to improve herd pregnancy rates.

**References**

Aoki ,V. W., L. Liu, P. Jones, H. H. Hatasaka, M. Gibson, C. M. Peterson, and D. T. Carrell. 2006a. Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. Fertil. Steril. 86(5):1408-1415.

Aoki, V. W., L. Liu, and D. T. Carrell. 2006b. A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. Mol. Hum. Reprod. 12:41-50.

Arangasamy, A., V. R.Kasimanickam, J. M. DeJarnette, and R. K. Kasimanickam. 2011. Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. Theriogenology. 76:570-577.

Avendano, C., A. Franchi, E. Jones, and S. Oehninger. 2009. Pregnancy-specific B-I-glycoprotein I and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. Hum. Reprod. 24:270-277.

Blankenberg , D., G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor. 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Molec. Biol. Chapter 19: Unit 19.10.1-21.

Card, C. J., E. J. Anderson, S. Zamberlan, K. E. Krieger, M. Kaproth, and B. L. Sartini. 2013. Cryopreserved bovine spermatozoal transcript profile as revealed by high-throughput ribonucleic acid sequencing. Biol. Reprod. In Press.

D'Amours, O., G. Frenette, M. Fortier, P. Leclerc, and R. Sullivan. 2010. Proteomic comparison of detergent-extracted sperm proteins from bulls with different fertility indexes. Reproduction. 139: 545-556.

Das, P. J., N. Paria, A. Gustafson-Seabury, M. Vishnoi, S. P. Chaki, C. C. Love, D. D. Varner, B. P. Chowdhary, and T. Raudsepp. 2010. Total RNA isolation from stallion sperm and testis biopsies. Theriogenology. 74:1099-1106.

Das, P. J., F. McCarthy, M. Vishnoi, N. Paria, C. Gresham, G. Li, P. Kachroo, A. K. Sudderth, S. Teague, C. C. Love, D. D. Varner, B. P. Chowdhary, and T. Raudsepp. 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., B. Kempisty, P. Jagodzinski, L. Pawelczyk, and P. Jedrzejczak. 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., N. Rodriguez-Osorio, A. Kaya, H. Wang, G. Page, G. C. Ostermeier, E. K. Topper, and E. Memili. 2010. Transcriptome analysis of bull spermatozoa: implications for male fertility. Reprod. Biomed. Online. 21:312-324.

W. Flowers. 2013. Sperm characterisics that limit success for fertilization. J. Anim. Sci. Online.

Foote, R. H. 2003. Fertility estimation: a review of past experience and future prospects. Anim. Reprod. Sci. 75:119-139.

Gadea, J., E. Selles, and M.A. Marco. 2004. The predictive value of porcine seminal parameters on fertility outcome under commercial conditions. Reproduction in Domestic Animals. 39: 303-308.

Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko. 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15(10):1451-1455.

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

Goecks, J., A. Nekrutenko, J. Taylor, 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.

Hamatani, T. 2012. Human spermatozoal RNAs. Fertil. Steril. 97:275-281.

Jodar, M., S. Kalko, J. Castillo, J. L. Ballesca, and R. Oliva. 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Hum. Reprod. 27(5):1431-1438.

Kasimanickam, V., R. Kasimanickam, A. Arangasamy, A. Saberivand, J. S. Stevenson, and J. P. Kastelic. 2012. Association between mRNA abundance of functional sperm function proteins and fertility of Holstein bulls. Theriogenology, 78(9): 2007-2019.

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

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

Kempisty, B., M. Depa-Martynow, M. Lianeri, P. Jedrzejczak, A. Darul-Wasowicz, and P. P. Jagodzinski. 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. Folia Histochemica Et Cytobiologica. 45:109-113.

Kempisty, B., P. Antosik, D. Bukowska, M. Jackowska, M. Lianeri, K. M. Jaskowski, and P. P. Jagodzinski. 2008. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reprod. Fertil. Dev. 20:513-518.

Lalancette, C., C. Thibault, I. Bachand, N. Caron, and N. Bissonnette. 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., A. E. Platts, G. D. Johnson, B. R. Emery, D. T. Carrell, and S. A. Krawetz. 2009. Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. 87:735-748.

Lambard, S., I. Galeraud-Denis, G. Martin, R. Levy, A. Chocat, and S. Carreau. 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. Molecular Hum. Reprod. 10:535-541.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3):R25.

R. Oliva. 2006. Protamines and male infertility. Human Reproduction. 12: 417-435.

Omran, H. M., M. Bakhiet, and M. G. Dashti. 2013. DNA integrity is a critical molecular indicator for the assessment of male infertility. Molecular Medicine and Reproduction. 7(5): 1631-1635.

Ostermeier, G. C., D. J. Dix, D. Miller, P. Khatri, and S. A. Krawetz. 2002. Spermatozoal RNA profiles of normal fertile men. The Lancet. 360:772-777.

Ostermeier, G. C., D. Miller, J. D. Huntriss, M. P. Diamond, and S. A. Krawetz. 2004. Delivering spermatozoan RNA to the oocyte. Nature 429:154.

Ostermeier, G. C., R. J. Goodrich, M. P. Diamond, D. J. Dix, and S. A. Krawetz. 2005. Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. Fertil. Steril. 26:1687-1694.

Park, Y. J., W. S. Kwon, S. A. Oh, and M. G. Pang. 2012. Fertility-related proteomic profiling bull spermatozoa separated by percoll. Proteome Research. 11:4162-4168.

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

Seelan, R. S. and L.I. Grossman. 1997. Structural organization and promoter analysis of the bovine cytochrome c oxidase subunit VIIc gene a functional role for YY1. J. Biol. Chem. 272(15): 10175-10181.

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

Swann, K., C. M. Saunders, N. T. Rogers, and F. A. Lai FA. 2006. PLCζ (zeta): A sperm protein that triggers $Ca^{2+}$ oscillations and egg activation in mammals. Cell Dev. Biol. 17:264-273.

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

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

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.

Werner, T. 2011. Next generation sequencing allows deeper analysis and understanding of genomes and transcriptomes including aspects to fertility. Reprod. Fertil. Dev. 23(1):75-80.

Yoshimura, T., Toyoda, S., Kuramochi-Miyagaw, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T., and J. Miyazaki. 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(1): 216-227.

**Figures**

Figure 1. Common and unique transcripts in lower (A) and higher (B) fertility populations.

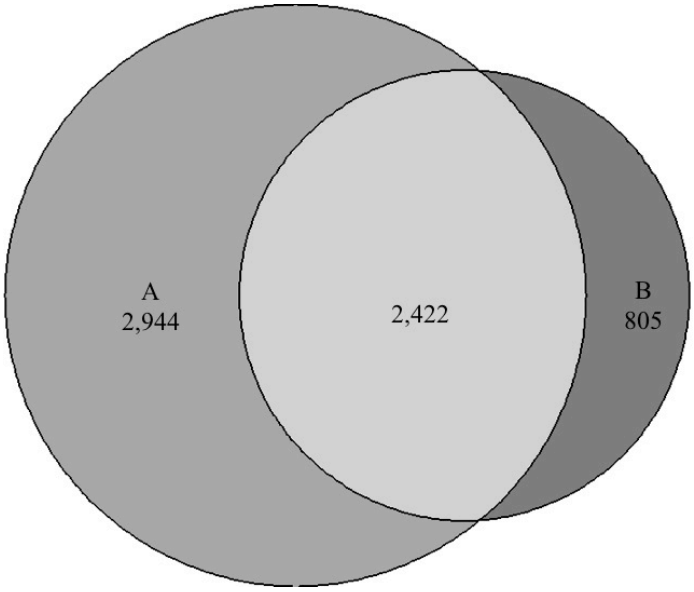


Figure 2. Correlation of qPCR transcript copy number and RNA-Seq FPKM for lower and higher fertility populations. Axes are base 10 log scale.

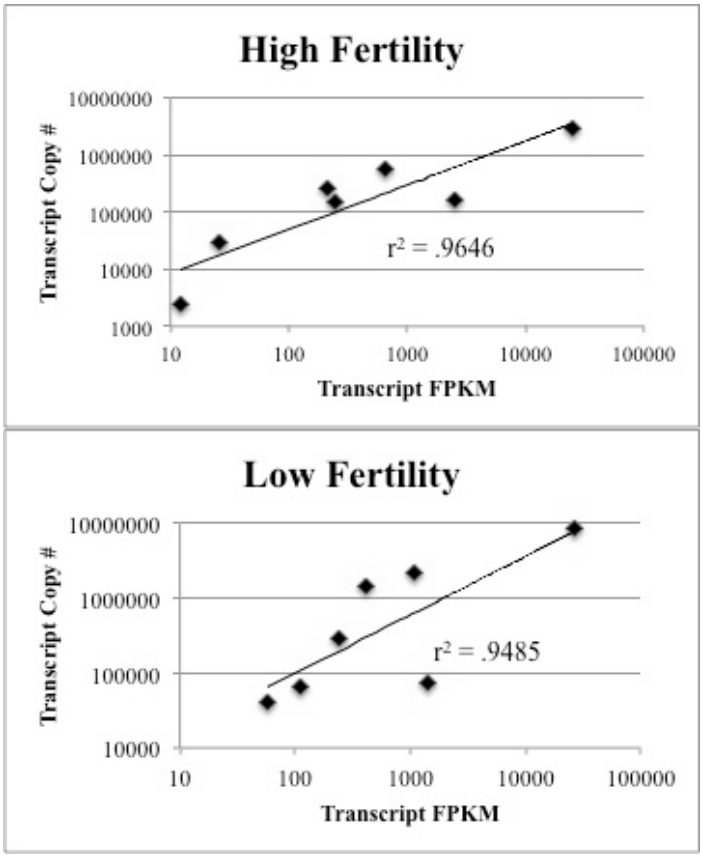


Figure 3. Differentially expressed transcripts (Cuffdiff analysis) in T = testis, H = higher fertility ds-cDNA, L = lower fertility ds-cDNA, N = no template control.

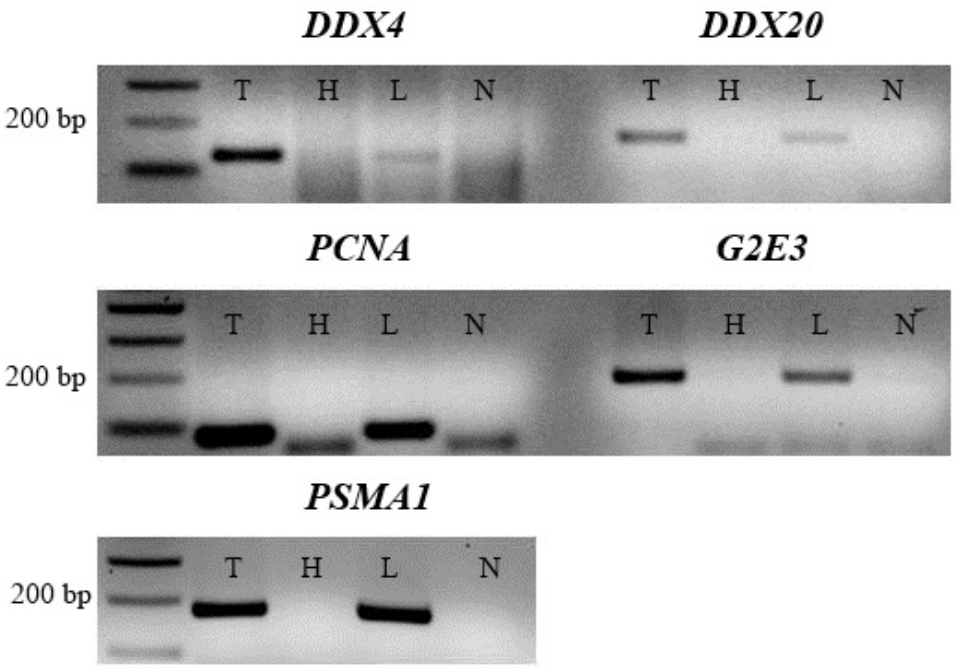
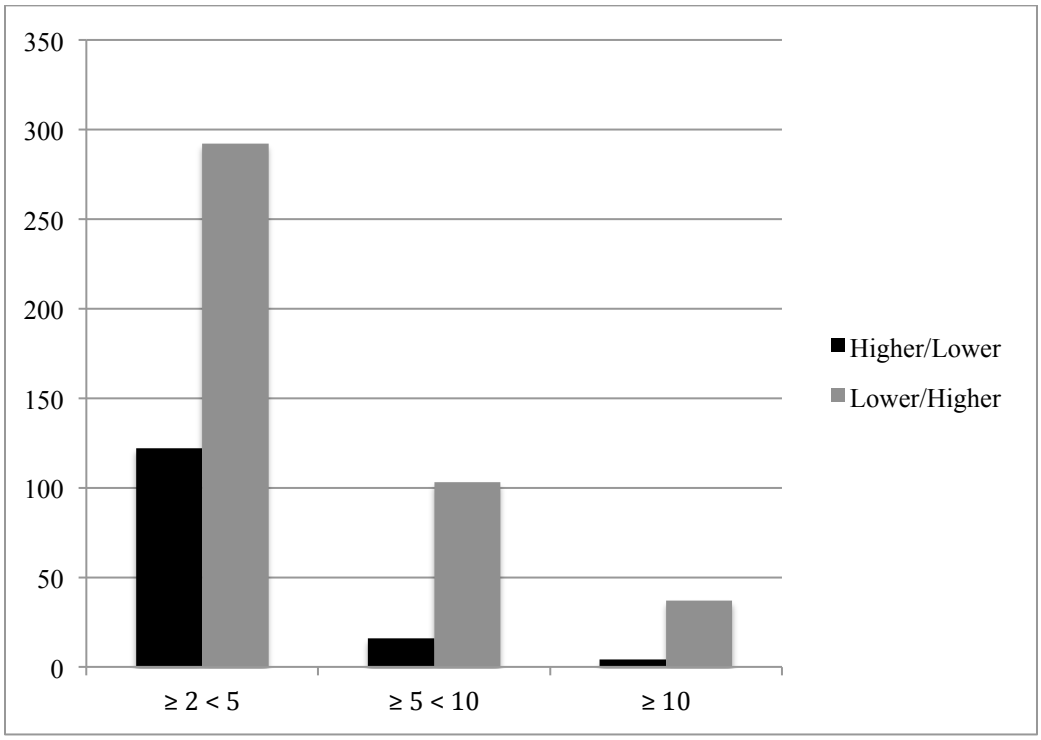


Figure 4. Differentially expressed transcripts (Ratio analysis) between higher and lower fertility populations.



Tables

**Table 1. Bovine primer sequences**

| Gene Symbol | Genbank ID | Primer Sequences | | | |
|---|---|---|---|---|---|
| | | | Forward | Reverse | Product Size (bp) |
| *ACRV1* | BC109966 | **qPCR** | 5'- TGAACAGGCCGTAGGTGAAC -3' | 5'- CAGCTTAATGGGTGGGCCTGA -3' | 174 |
| *AKAP4* | AF100170 | **Ind** | 5'- AGGGGTCAGTGTGCCTTTTC -3' | 5'- TCCAGACGTAGGCTCTGAGG -3' | 234 |
| *APOABP1* | AY528250 | **Ind** | 5'- CCCTGACACGGAATGTGTCT -3' | 5'- AGAGAGGCACTAAGTCTTTATTGA -3' | 80 |
| *ATPIF1* | NM_175816 | **Ind; Ratio** | 5'- CTTTGGCTCGGAATCGGGAG -3' | 5'- GCCAGCTGTTCTTTAGCACG -3' | 132 |
| *ATP5J* | NM_174717 | **Ind; qPCR; Ratio** | 5'- CTTCCTGTCCGGATCACCAT -3' | 5'- TCCTCCAGATGTCTGTCGCT -3' | 195 |
| *BRP* | NM_001012682 | **Ind** | 5'- CCAGGCTTTAGGCATCACCA -3' | 5'- GGCGCCTACTTTGTCTCCTGT -3' | 94 |
| *CCT8* | NM_001033609 | **Ind** | 5'- TGGATCAGATCATCATGGCAA -3' | 5'- AGGGTGCAAGATACAAACACAAAA -3' | 202 |
| *CETN2* | NM_001038515 | **Ind; Ratio** | 5'- GTCCGGGATGGCCTCTAACT -3' | 5'- TCCTGTTTCTGCTCTTCGGT -3' | 102 |
| *CLGN* | BC103401 | **Ind** | 5'- ATCCCTGATGCTTCTGCTGTC -3' | 5'- ATTCACCACACCCAATCCGA -3' | 243 |
| *CKS2* | BC105331 | **Ind** | 5'- CCTCGCTTCGCTCTAGTCAG -3' | 5'- AGGATGTGTGGTTCTGGTTCA -3' | 245 |
| *CMYC* | BC109848 | **Ind** | 5'- CCACCTAGGAGGAGAAACG -3' | 5'- ACAGTTAACGTGTGATAGGTGAAT -3' | 168 |
| *COX7A2* | DQ347636 | **Ind** | 5'- AACTGGCTGTGGCTTCGTTT -3' | 5'- TGCTTTATTGGTGGCAGCTAA -3' | 204 |
| *COX7C* | NM_175831 | **Ind; qPCR; Ratio** | 5'- GAGGAGGGTCCAGGGAAGAATA -3' | 5'- TTTCAGCAGTTGGTGTCTTACT -3' | 129 |
| *CRISP2* | BC109478 | **Ind; qPCR** | 5'- CGGCCGCTCTGCAACAGAAG -3' | 5'- GGAAGCAGCACAGCGGTCAGA -3' | 120 |
| *DDX4* | AF541971 | **Cuff** | 5'- TCGCATGCTGGATATGGGTT -3' | 5'- TCCCCAGCCAACCTTTGAAT -3' | 135 |
| *DDX20* | BC151337 | **Cuff** | 5'- GCTGGCCGTTTTGGAACTTT -3' | 5'- GCTTCAACCTCCACATCCCA -3' | 173 |

**Table 1. Bovine primer sequences (continued)**

| | | | | | |
|---|---|---|---|---|---|
| *DNAJB7* | BC142060 | **Ind** | 5'- GCACTTAAATGGCACCCAGA -3' | 5'- TGGCTTACGGAATGTGAAGC -3' | 204 |
| *EEF1G* | AB098752 | **Ind; Val** | 5'- ATCCAGTTTCCGCCATGTGT -3' | 5'- GTTGCAACGCTCATCACTGG -3' | 198 |
| *EIF1* | BC103170 | **Ind; Val** | 5'- AAGGGTGATGATCTGCTTCCTGCT -3' | 5'- AACTGGCATATGTTCTTGCGCTGG -3' | 235 |
| *G2E3* | NM_001038671 | **Cuff; Ind** | 5'- TCACTGGCAATTTTGCGTCA -3' | 5'- TGCACCTGCAAACAGTCTCT -3' | 187 |
| *GPX4* | NM_174770 | **Ind; qPCR; Ratio** | 5'- CAATGTGGCCTCGCAATGAG -3' | 5'- CATTACTCCCTGGCTCCTGC -3' | 140 |
| *GSTM3* | BC112491 | **Ind** | 5'- TGTGCCGTTTTGAGGCTTTGGAG -3' | 5'- GGGCCATCTTGTTGTTGACAGGCAT -3' | 112 |
| *GTSF1* | BC102713 | **Ind; qPCR; Ratio** | 5'- CAGGTTCCTCGGGCTGAAAT -3' | 5'- ACTATGTTGCTTGCAGGGCT -3' | 239 |
| *HEMGN* | BC109747 | **Ind** | 5'- CAGCAAGAAATGGCTGTGCC -3' | 5'- CAGGTGCATAGGCTTCAGGT -3' | 247 |
| *HMGB4* | BC109790 | **Val** | 5'- AGCTGGTCGGTGGTGCAGGT -3' | 5'- GCAAGCATGTCTTCCGGGC -3' | 167 |
| *PABPC1* | NM_174568 | **Ratio** | 5'- ACATGGATGATGAGCGCCTT -3' | 5'- TCTCATCCACAGCTTTTTGTGC -3' | 150 |
| *PCNA* | BC103068 | **Cuff; Ind** | 5'- GAACCTCACCAGCATGTCCA -3' | 5'- ACGTGTCCGCGTTATCTTCA -3' | 86 |
| *PLCZ1* | BC114836 | **Val** | 5'- CGGGTGGTCGGAATCCCACTCT -3' | 5'- AATTCCCTGGCTGCCAACTTTGT -3' | 194 |
| *PRM1* | BC108207 | **Ind; qPCR; Val; Ratio** | 5'- AAGAAGATGTCGCAGACGAAGGAG -3' | 5'- ACAGGTGGCATTGTTCGTTAGCAG -3' | 228 |
| *PRM2* | NM_174157 | **Ind; qPCR** | 5'- ATGCGGAAGGCAGCTCTAAG -3' | 5'- CTCAAGATCTCGTGGGCTCC -3' | 109 |
| *PSMA1* | BC102216 | **Cuff; qPCR** | 5'- TGGTTAAGCATGGTCTGCGT -3' | 5'- GGTTGAGTAGGCTGTGCCTT -3' | 187 |
| *PSMA6* | BC110260 | **Ind; qPCR** | 5'- ACAGTGGAAACTGCGATTACATGCC -3' | 5'- ACAGGCAAGTGGCGTCACGG -3' | 205 |
| *SEC61G* | BC102186 | **Val** | 5'- GCAGACGCGGAGCAGACATCA -3' | 5'- AGCGAATCCTATTGCTGTTGCCA -3' | 155 |
| *SPAG4* | NM_001076507 | **qPCR; Ratio** | 5'- CCACGGTTATCTTGGAGCCAG -3' | 5'- GATGCTGCAGAGTGATGTCG -3' | 123 |
| *SPATA6* | NM_001046371 | **Ind; Ratio** | 5'- GATGAGATCCATGACCGGGTA -3' | 5'- TAAGAGGCTGCCCTGTTGGA -3' | 230 |
| *TNP1* | NM_174199 | **Ind; qPCR** | 5'- AAACGCTGTGACGATGCCAA -3' | 5'- CTTGGCAGTCCCCTTCTGTT -3' | 118 |

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

**Table 2. RNA-Seq read and mapping statistics**

| | Total Reads | Reads Mapped | Reads Uniquely Mapped | Reads in Coding Exons | Reads in 5' UTRs | Reads in 3' UTRs | Transcripts Expressed |
|---|---|---|---|---|---|---|---|
| **Higher Fertility** | 10,454,393 | 1,021,536 (9.77%) | 537,102 (52.57%) | 56,736 | 9,732 | 8,320 | 3,227 |
| **Lower Fertility** | 10,721,121 | 1,336,844 (12.47%) | 940,778 (70.37%) | 177,359 | 24,227 | 24,776 | 5,336 |

**Table 3. Top 10 transcripts unique to higher fertility population based on FPKM**

| Transcript | Gene Name | Accession Number | FPKM |
|---|---|---|---|
| *MIR708* | microRNA 708 | n/a | 200.937 |
| *C4H7orf55* | chromosome 4 open reading fram, human C7orf55 | NM_001076997 | 122.222 |
| *VSNL1* | visinin-like 1 | NM_174490 | 106.011 |
| *AKR1B1* | aldo-keto reductase family 1, member B1 (aldose reductase) | NM_001012519 | 95 |
| *APOA1BP* | apolipoprotein A-I binding protein | NM_205796 | 91 |
| *SQRDL* | sulfide quinone reductase-like (yeast) | NM_001040511 | 91.0495 |
| *CD28* | CD28 molecule | NM_181004 | 87 |
| *MPI* | mannose phosphate isomerase | NM_001035284 | 65 |
| *RNGTT* | RNA guanylyltransferase and 5'-phosphatase | NM_001046085 | 64 |
| *FCF1* | similar to rRNA-processing protein FCF1 homolog; FCF1 small subunit (SSU) processome component homolog (S. cerevisiae) | NM_001037452 | 63 |

**Table 4. Top 10 transcripts unique to lower fertility population based on FPKM**

| Transcript | Gene Name | Accession Number | FPKM |
|---|---|---|---|
| *RN5-8S1* | 5.8S ribosomal RNA | NR_036643 | 348.918 |
| *VDAC3* | voltage-dependent anion channel 3 | NM_174729 | 300.26 |
| *DEFB122A* | beta-defensin 122a | NM_001102339 | 247.702 |
| *ATP6V1D* | ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D; similar to Vacuolar proton pump subunit D (V-ATPase subunit D) (V-ATPase 28 kDa accessory protein) | NM_001075141 | 228.678 |
| *LOC526524* | FK506 binding protein 1A, 12kDa; FK506 binding protein 1A, 12kDa-like | NM_001035456 | 203.048 |
| *UQCR10* | ubiquinol-cytochrome c reductase complex 7.2 kDa protein | NM_001113723 | 195.179 |
| *PSMA1* | proteasome (prosome, macropain) subunit, alpha type, 1 | NM_001035310 | 171.487 |
| *INSL6* | insulin-like 6 | NM_001077521 | 162.748 |
| *C10H14orf166* | chromosome 14 open reading frame 166 ortholog | NM_001035280 | 161.771 |
| *METAP2* | methionyl aminopeptidase 2 | NM_001040493 | 157.701 |

**Table 5. Top 10 Biological Processes (BP) for common transcripts and unique transcripts when comparing lower and higher fertility populations via gene ontology (GO) analysis**

| | Biological Processes | % of Transcripts |
|---|---|---|
| **Unique to Higher Fertility** | regulation of transcription | 9.23 |
| | regulation of transcription, DNA-dependent | 6.53 |
| | oxidation reduction | 4.41 |
| | macromolecule catabolic process | 3.55 |
| | RNA processing | 3.27 |
| | protein catabolic process | 2.98 |
| | cellular macromolecule catabolic process | 2.98 |
| | positive regulation of macromolecule metabolic process | 2.84 |
| | negative regulation of macromolecule metabolic process | 2.56 |
| | regulation of cell proliferation | 2.56 |
| **Unique to Lower Fertility** | phosphate metabolic process | 4.92 |
| | phosphate metabolic process | 4.92 |
| | phosphorus metabolic process | 4.92 |
| | transcription | 4.41 |
| | proteolysis | 4.41 |
| | phosphorylation | 4.09 |
| | oxidation reduction | 3.97 |
| | protein localization | 3.86 |
| | protein transport | 3.71 |
| | establishment of protein localization | 3.71 |
| | intracellular signaling cascade | 3.54 |
| **Common Transcripts** | regulation of transcription | 8.04 |
| | protein localization | 5.35 |
| | proteolysis | 5.05 |
| | protein transport | 4.86 |
| | establishment of protein localization | 4.86 |
| | translation | 4.61 |
| | transcription | 4.51 |
| | macromolecule catabolic process | 3.92 |
| | protein catabolic process | 3.78 |
| | cellular macromolecule catabolic process | 3.78 |

**Table 6. Differentially expressed transcripts between high and low fertility populations via cuffdiff analysis**

| Accession Number | Gene Symbol* | Gene Name |
|---|---|---|
| NM_001035310 | PSMA1 | proteasome (prosome, macropain) subunit, alpha type, 1 |
| NM_001035280 | C10H14orf166 | chromosome 14 open reading frame 166 ortholog |
| NM_001015639 | PSMC2 | proteasome (prosome, macropain) 26S subunit, ATPase, 2 |
| NM_001101989 | TMEM144 | transmembrane protein 144 |
| NM_001040493 | METAP2 | methionyl aminopeptidase 2 |
| NM_001035284 | **MPI** | mannose phosphate isomerase |
| NM_001034494 | PCNA | proliferating cell nuclear antigen |
| NM_174490 | **VSNL1** | visinin-like 1 |
| NM_001040511 | **SQRDL** | sulfide quinone reductase-like (yeast) |
| NM_001075141 | ATP6V1D | ATPase, H+ transporting, lysosomal 34kDa, V1 subunit D |
| NM_001079628 | TMEM217 | transmembrane protein 217 |
| NM_001075926 | ACRV1 | acrosomal vesicle protein 1 |
| NM_001038671 | G2E3 | G2/M-phase specific E3 ubiquitin ligase |
| NM_001012519 | **AKR1B1** | aldo-keto reductase family 1, member B1 (aldose reductase) |
| NM_001077068 | TMBIM1B | hypothetical protein MGC134563; similar to CG3814 CG3814-PA |
| NM_001046085 | **RNGTT** | RNA guanylyltransferase and 5'-phosphatase |
| NM_001046143 | CDKN2AIP | CDKN2A interacting protein |
| NM_001075396 | PIAS1 | protein inhibitor of activated STAT, 1 |
| NM_001075836 | RIOK3 | RIO kinase 3 (yeast) |
| NM_205796 | **APOA1BP** | apolipoprotein A-I binding protein |
| NM_001015595 | **GFRA2** | GDNF family receptor alpha 2 |
| NM_001075527 | CRBN | cereblon |
| NM_001083728 | **MAPK10** | mitogen-activated protein kinase 10; mitogen-activated protein kinase 8 |
| NM_001076077 | **PHTF1** | putative homeodomain transcription factor 1 |
| NM_001007819 | DDX4 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 4 |
| NM_001081584 | ATF2 | activating transcription factor 2 |
| NM_001103234 | **ACAP1** | ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 |
| NM_001099062 | FBXO3 | F-box protein 3 |
| NM_001102118 | DDX20 | DEAD (Asp-Glu-Ala-Asp) box polypeptide 20 |
| AJ439530 | **CACNA2D1** | calcium channel, voltage-dependent, alpha 2/delta subunit 1 |
| AF086808 | ADAM2 | ADAM metallopeptidase domain 2 |
| XM_870523 | **SDC3** | syndecan 3 |
| NM_001206525 | CCDC41 | coiled-coil domain containing 41 |
| NM_001114192 | HSPA4 | heat shock 70kDa protein 4 |
| NM_174729 | VDAC3 | voltage-dependent anion channel 3 |
| NM_001242584 | IGBP1 | immunoglobulin (CD79A) binding protein 1 |

*** bold** = present in higher fertility only; not bold = present in lower fertility only

**Table 7. Transcript presence (x) in individual bulls from lower (1) to higher (9) fertility.**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| ***AKAP4*** | | | | | | | | | |
| ***APOABP1*** | | | | | | | | | |
| ***ATPIF1*** | | | | | x | | | | x |
| ***ATP5J*** | | | x | | | | | | |
| ***BRP*** | | | | | | | | | |
| ***CCT8*** | x | | | | | | | | |
| ***CETN2*** | | | x | x | x | | x | | |
| ***CLGN*** | | x | | | | | x | | |
| ***CKS2*** | | x | x | x | x | | x | | x |
| ***CMYC*** | | | | x | x | x | | x | |
| ***COX7A2*** | x | | | x | | | x | x | |
| ***COX7C*** | x | x | | x | | x | x | x | x |
| ***CRISP2*** | x | x | | x | | x | | x | |
| ***DNAJB7*** | | x | | x | x | | | | |
| ***EEF1G*** | x | x | | x | x | x | x | | x |
| ***EIF1*** | | | | x | x | x | | | x |
| ***G2E3*** | x | x | | x | | x | x | x | x |
| ***GPX4*** | x | | | | x | | | | |
| ***GSTM3*** | x | | x | x | x | x | x | x | |
| ***GTSF1*** | x | | x | x | x | | x | x | x |
| ***HEMGN*** | | | | | | | x | | |
| ***PCNA*** | x | | | | | | | | x |
| ***PRM1*** | x | x | x | x | x | x | x | x | x |
| ***PRM2*** | x | x | x | x | x | x | x | x | |
| ***PSMA6*** | x | x | | x | x | x | | | x |
| ***SPATA6*** | x | | | x | | | | x | x |
| ***TNP1*** | x | x | x | x | x | x | x | x | |

**Table 8. Higher and lower fertility expression levels of previously published bovine spermatozoal transcripts**

| | Higher Fert. FPKM | Lower Fert. FPKM | Publication |
|---|---|---|---|
| ***CCT8*** | - | 18 | Arangasamy et al. 2011 |
| ***CRISP2*** | 661 | 1366 | " |
| ***PEBP1*** | 5 | 41 | " |
| ***TSSK6*** | - | - | Feugang et al. 2010 |
| ***ADAM5P*** | - | - | " |
| ***CD36*** | - | - | Kasimanickam et al. 2012 |
| ***CENPA*** | - | - | " |
| ***AKI*** | - | - | " |
| ***IB5*** | - | - | " |
| ***DOPPEL*** | - | - | " |
| ***NGF*** | 11 | 2 | " |
| ***TIMP2*** | 4 | 14 | " |
| ***LDC1*** | - | - | " |
| ***SNRPN*** | 0.0001 | 3 | " |
| ***ODF2*** | 50 | 35 | " |
| ***PLCZ1*** | 27 | 59 | " |
| ***ADIPOR1*** | 4 | 3 | Kasimanickam et al. 2013 |
| ***ADIPOR2*** | - | 0.66 | " |
| ***ADIPO*** | - | - | " |

**Table 9. Transcript expression levels in lower and higher fertility populations for spermatozoal proteins previously identified as having correlations with fertility. - = not present**

| | Higher Fert. FPKM | Lower Fert. FPKM | Publication |
|---|---|---|---|
| ***ATPSB*** | - | - | Park et al. 2012 |
| ***ENO1*** | 6 | 15 | " |
| ***ASPP2*** | - | - | " |
| ***GPX4*** | 18 | 138 | " |
| ***AHSG*** | - | - | " |
| ***AKI*** | 10 | 20 | D'Amours et al. 2010 |
| ***PEBP1*** | 5 | 41 | " |
| ***CCT5*** | 3 | 9 | " |
| ***CCT8*** | - | 18 | " |
| ***ELSPBP1*** | - | - | " |
| ***PSMA6*** | 537 | 1296 | " |
| ***BSP1*** | 11 | 67 | " |
| ***VDAC1*** | - | 0.62 | Park et al. 2012 |
| ***UQCRC2*** | 7 | 18 | " |
| ***ROPN1*** | 85 | 409 | " |

**Table 10. Linear correlation analysis for expression levels (qPCR) of 12 transcripts across 9 individual bulls of varying fertility**

| | *GTSF1* | *PRM2* | *TNP1* | *PSMA1* | *PSMA6* | *ACRV1* | *SPAG4* | *COX7C* | *ATP5J* | *GPX4* | *PRM1* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $r^2$ | 0.8432 | 0.6019 | 0.497 | 0.04819 | 0.1217 | 0.0962 | 0.1045 | 0.7776 | 0.0005 | 0.04733 | 0.3673 |
| **p value** | 0.0005 | 0.014 | 0.0339 | 0.5704 | 0.3574 | 0.5496 | 0.3961 | 0.0017 | 0.9643 | 0.5739 | 0.0836 |

Note: A positive correlation indicates that higher fertility sires have lower copy #s for a given transcript

**Appendix 1. Transcripts with > 2:1 ratios in lower:higher fertility populations**

| Official Gene Symbol | Gene Long Name | Higher FPKM | Lower FPKM | Lower/ Higher Ratio |
|---|---|---|---|---|
| RPL30 | #N/A | 8.06 | 420.66 | 52.22 |
| CETN1 | centrin, EF-hand protein, 1 | 9.31 | 433.41 | 46.57 |
| MRPS33 | #N/A | 6.45 | 209.46 | 32.49 |
| PTTG1 | protein tyrosine phosphatase type IVA, member 1 | 6.62 | 194.87 | 29.43 |
| C28H10orf53 | UPF0728 protein C10orf53 homolog | 12.68 | 333.48 | 26.30 |
| LOC781895 | hypothetical protein LOC768323 | 39.07 | 980.19 | 25.09 |
| FANCD2OS | #N/A | 7.00 | 135.75 | 19.39 |
| HSFY2 | #N/A | 13.56 | 259.74 | 19.16 |
| HSFY2 | #N/A | 13.56 | 259.74 | 19.16 |
| HSFY2 | #N/A | 13.56 | 259.74 | 19.16 |
| HSFY2 | #N/A | 13.56 | 259.74 | 19.16 |
| COPS5 | similar to COP9 signalosome subunit 5; COP9 constitutive photomorphogenic homolog subunit 5 (Arabidopsis) | 8.57 | 160.08 | 18.67 |
| DYNLL1 | #N/A | 12.17 | 222.01 | 18.24 |
| TOMM7 | translocase of outer mitochondrial membrane 5 homolog (yeast) | 6.41 | 116.32 | 18.15 |
| NDUFA7 | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa | 5.80 | 93.32 | 16.09 |
| H2AFZ | H2A histone family, member Z | 11.16 | 178.18 | 15.97 |
| CDO1 | cysteine dioxygenase, type I | 8.56 | 134.59 | 15.72 |
| HINT1 | #N/A | 6.31 | 98.05 | 15.55 |
| RPS18 | ribosomal protein S17 | 14.02 | 216.93 | 15.47 |
| SRR | signal recognition particle 72kDa | 8.17 | 124.32 | 15.23 |
| MGC148328 | hypothetical protein MGC137055 | 1625.53 | 23928.30 | 14.72 |
| MGC148328 | #N/A | 1625.53 | 23928.30 | 14.72 |
| RPL34 | ribosomal protein L32 | 14.58 | 191.73 | 13.15 |
| ERH | similar to enhancer of rudimentary homolog; enhancer of rudimentary homolog (Drosophila) | 7.77 | 99.55 | 12.80 |
| IFT57 | intraflagellar transport 57 homolog (Chlamydomonas) | 5.67 | 65.99 | 11.64 |
| RPL14 | ribosomal protein L13a | 21.60 | 250.16 | 11.58 |
| SUMO2 | sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 | 8.71 | 95.97 | 11.02 |
| SUMO2 | SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae) | 8.71 | 95.97 | 11.02 |
| CD59 | CD59 molecule, complement regulatory protein | 12.22 | 134.27 | 10.99 |
| MRFAP1 | #N/A | 13.90 | 152.68 | 10.99 |
| CEP57 | centrosomal protein 57kDa | 6.12 | 66.63 | 10.89 |
| ATP5B | ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide | 8.10 | 84.88 | 10.48 |
| NDUFA5 | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa | 9.05 | 93.86 | 10.37 |
| PGK2 | FGF receptor activating protein 1 | 35.68 | 369.93 | 10.37 |
| PABPC1 | proliferation-associated 2G4, 38kDa | 22.22 | 228.22 | 10.27 |
| CCDC23 | coiled-coil domain containing 23 | 8.50 | 85.90 | 10.11 |
| FAIM2 | Fas apoptotic inhibitory molecule 2 | 8.33 | 83.50 | 10.03 |
| KLHL10 | #N/A | 14.12 | 140.42 | 9.94 |
| LOC788142 | #N/A | 7.26 | 70.41 | 9.70 |
| ASB17 | ankyrin repeat and SOCS box-containing 17 | 21.52 | 205.35 | 9.54 |

| | | | | |
|---|---|---|---|---|
| UBE2K | ubiquitin-conjugating enzyme E2, J1 (UBC6 homolog, yeast) | 7.65 | 72.61 | 9.49 |
| C8H9orf135 | #N/A | 12.17 | 115.21 | 9.47 |
| CHPT1 | choline phosphotransferase 1 | 6.65 | 63.00 | 9.47 |
| CA6 | carbonic anhydrase VI | 5.65 | 52.50 | 9.28 |
| PSENEN | #N/A | 8.19 | 76.04 | 9.28 |
| RPS14 | ribosomal protein S12 | 8.73 | 81.05 | 9.28 |
| COX7C | cytochrome c oxidase subunit VIIc | 52.63 | 484.73 | 9.21 |
| GHITM | #N/A | 6.74 | 61.96 | 9.20 |
| SLIRP | solute carrier family 30 (zinc transporter), member 3 | 53.19 | 482.84 | 9.08 |
| UCHL3 | ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase) | 9.85 | 88.44 | 8.97 |
| C1H21orf59 | chromosome 21 open reading frame 59 ortholog | 14.51 | 128.88 | 8.88 |
| CLIC4 | chloride intracellular channel 4 | 14.66 | 129.83 | 8.86 |
| RPS4Y1 | ribosomal protein S4, Y-linked 2; similar to ribosomal protein S4, X-linked X; ribosomal protein S4, X-linked | 11.98 | 105.73 | 8.82 |
| ATP5J | ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6 | 50.28 | 441.92 | 8.79 |
| PPP1R2 | protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform | 47.53 | 414.36 | 8.72 |
| PLK1S1 | phospholipase D1, phosphatidylcholine-specific | 5.28 | 45.53 | 8.62 |
| ATP5O | ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit | 20.64 | 175.16 | 8.49 |
| TRIM59 | tripartite motif-containing 36 | 13.99 | 116.44 | 8.32 |
| RPS24 | ribosomal protein S23; similar to ribosomal protein S23 | 54.06 | 449.04 | 8.31 |
| RPL10L | #N/A | 27.93 | 229.88 | 8.23 |
| PHYHIPL | phosphoglycolate phosphatase | 18.77 | 153.54 | 8.18 |
| GNAS | GNAS complex locus | 6.58 | 53.73 | 8.17 |
| NUPL2 | nucleobindin 1 | 13.63 | 110.49 | 8.11 |
| C21H15orf63 | #N/A | 6.28 | 50.53 | 8.05 |
| LZTFL1 | lysozyme-like 2 | 6.47 | 52.02 | 8.05 |
| LZTFL1 | leucine zipper transcription factor-like 1 | 6.47 | 52.02 | 8.05 |
| PEBP1 | #N/A | 5.28 | 40.88 | 7.74 |
| SAMD13 | serum amyloid A4, constitutive | 27.50 | 212.74 | 7.74 |
| RPS19 | ribosomal protein S18 | 10.32 | 79.29 | 7.68 |
| CSDE1 | cold shock domain containing E1, RNA-binding | 22.61 | 173.34 | 7.67 |
| LOC783012 | hypothetical protein LOC781895 | 121.35 | 926.74 | 7.64 |
| NUPR1L | nucleoporin like 2 | 13.20 | 100.80 | 7.63 |
| SPCS1 | spermatogenesis associated, serine-rich 2 | 5.25 | 39.83 | 7.58 |
| GPX4 | glutathione peroxidase 4 (phospholipid hydroperoxidase) | 18.29 | 138.68 | 7.58 |
| LOC768323 | testis-specific histone 2a-like | 194.47 | 1472.79 | 7.57 |
| LOC768323 | hypothetical protein LOC768323 | 194.47 | 1472.79 | 7.57 |
| TXNDC17 | thioredoxin | 7.71 | 57.29 | 7.43 |
| PSMA4 | presenilin enhancer 2 homolog (C. elegans) | 6.85 | 50.87 | 7.43 |
| ATPIF1 | ATPase inhibitory factor 1 | 6.31 | 46.83 | 7.43 |
| SELK | Sec61 gamma subunit | 36.11 | 268.22 | 7.43 |
| MS4A14 | membrane-spanning 4-domains, subfamily A, member 13 | 8.49 | 62.32 | 7.34 |
| RPS23 | ribosomal protein S21 | 11.43 | 82.94 | 7.26 |
| PCBP1 | poly (ADP-ribose) polymerase family, member 11 | 5.62 | 40.56 | 7.22 |
| ADAM32 | ADAM metallopeptidase domain 32 | 17.83 | 128.51 | 7.21 |
| SOD1 | sine oculis binding protein homolog (Drosophila) | 9.69 | 68.96 | 7.12 |

| | | | | |
|---|---|---|---|---|
| C17H5orf52 | chromosome 5 open reading frame 52 ortholog | 10.92 | 76.04 | 6.96 |
| KHDRBS3 | potassium channel modulatory factor 1 | 9.23 | 64.28 | 6.96 |
| SH3GLB1 | SH3 domain binding glutamic acid-rich protein like 3 | 8.10 | 55.98 | 6.91 |
| C7H5orf48 | #N/A | 56.24 | 385.08 | 6.85 |
| NEDD8 | NADH dehydrogenase (ubiquinone) Fe-S protein 5, 15kDa (NADH-coenzyme Q reductase) | 10.63 | 72.36 | 6.81 |
| STK17A | signal transducer and activator of transcription 1, 91kDa; signal transducer and activator of transcription 4 | 13.08 | 88.03 | 6.73 |
| ARL3 | ADP-ribosylation factor-like 3 | 18.51 | 122.73 | 6.63 |
| RNF181 | ring finger protein 138 | 12.51 | 82.60 | 6.60 |
| PSMB3 | proteasome (prosome, macropain) subunit, alpha type, 6 | 8.70 | 57.43 | 6.60 |
| DYDC1 | DPY30 domain containing 1 | 44.45 | 292.98 | 6.59 |
| FTH1 | #N/A | 13.19 | 85.70 | 6.50 |
| STT3B | STT3, subunit of the oligosaccharyltransferase complex, homolog A (S. cerevisiae) | 13.94 | 90.19 | 6.47 |
| RNF138 | similar to ring finger protein 103 | 15.15 | 97.93 | 6.46 |
| HMGN3 | high mobility group nucleosomal binding domain 3 | 5.19 | 32.14 | 6.19 |
| OXCT2 | OTU domain, ubiquitin aldehyde binding 2 | 6.54 | 40.46 | 6.19 |
| DCUN1D1 | DCN1, defective in cullin neddylation 1, domain containing 1 (S. cerevisiae) | 165.17 | 1022.28 | 6.19 |
| HSBP1 | heat shock factor binding protein 1 | 52.82 | 324.18 | 6.14 |
| UBC | #N/A | 9.04 | 55.19 | 6.10 |
| MGC134066 | hypothetical protein MGC133647 | 5.33 | 32.18 | 6.03 |
| SERP2 | stress-associated endoplasmic reticulum protein 1 | 20.26 | 121.81 | 6.01 |
| BSP1 | seminal vesicle secretory protein 109 | 11.27 | 67.42 | 5.98 |
| USP50 | ubiquitin specific peptidase 44 | 18.49 | 109.18 | 5.90 |
| C8H9orf24 | #N/A | 6.82 | 40.10 | 5.88 |
| MEA1 | #N/A | 30.55 | 179.29 | 5.87 |
| GLUL | glutamate-ammonia ligase (glutamine synthetase) | 11.52 | 66.94 | 5.81 |
| RPS12 | ribosomal protein S11 | 42.29 | 243.05 | 5.75 |
| ARRDC4 | arrestin domain containing 4 | 20.37 | 117.05 | 5.75 |
| PCYT1A | protein-L-isoaspartate (D-aspartate) O-methyltransferase | 12.30 | 70.59 | 5.74 |
| AKAP4 | A kinase (PRKA) anchor protein 4 | 31.55 | 179.45 | 5.69 |
| TEX26 | testis expressed 101 | 11.31 | 64.16 | 5.67 |
| C10H15orf23 | chromosome 15 open reading frame 23 ortholog | 50.81 | 284.82 | 5.61 |
| AGPAT3 | 1-acylglycerol-3-phosphate O-acyltransferase 3 | 6.49 | 36.13 | 5.57 |
| ARGLU1 | arginine and glutamate rich 1 | 40.61 | 226.22 | 5.57 |
| ARGLU1 | arginine and glutamate rich 1 | 40.61 | 226.22 | 5.57 |
| CKS2 | CDC28 protein kinase regulatory subunit 2 | 86.19 | 477.80 | 5.54 |
| GLRX2 | glutaredoxin 2 | 11.83 | 65.56 | 5.54 |
| CDC5L | CDC5 cell division cycle 5-like (S. pombe) | 5.43 | 29.85 | 5.50 |
| RPS6 | ribosomal protein S4, Y-linked 1 | 92.63 | 506.01 | 5.46 |
| RPS6 | #N/A | 92.63 | 506.01 | 5.46 |
| LDHC | lysosomal protein transmembrane 4 alpha | 62.50 | 341.40 | 5.46 |
| MS4A5 | MS4A13 protein | 23.81 | 128.92 | 5.42 |
| ODF1 | OCIA domain containing 1 | 16.65 | 87.01 | 5.23 |
| NPC2 | #N/A | 9.09 | 46.88 | 5.16 |
| C21H14orf142 | Uncharacterized protein C14orf142 homolog | 7.87 | 40.57 | 5.16 |

| | | | | |
|---|---|---|---|---|
| SHCBP1L | SH3-domain GRB2-like endophilin B1 | 19.92 | 102.37 | 5.14 |
| PSMG1 | proteasome (prosome, macropain) inhibitor subunit 1 (PI31) | 7.69 | 39.24 | 5.11 |
| CSDA | cold shock domain protein A | 22.72 | 115.67 | 5.09 |
| CCNB1 | similar to Cyclin B1; cyclin B1 | 6.27 | 31.80 | 5.08 |
| SRP72 | signal recognition particle 54kDa | 17.49 | 88.69 | 5.07 |
| COX7A2 | cytochrome c oxidase subunit VIIa polypeptide 2 (liver) | 46.39 | 235.08 | 5.07 |
| CAPZA1 | capping protein (actin filament) muscle Z-line, alpha 1 | 8.33 | 42.17 | 5.06 |
| CLGN | calmegin | 82.69 | 417.94 | 5.05 |
| C18H16orf87 | #N/A | 17.27 | 87.05 | 5.04 |
| RPS11 | ribosomal protein S10; similar to 40S ribosomal protein S10 | 177.02 | 886.91 | 5.01 |
| BCL2L14 | Bcl2-like 14 (apoptosis facilitator) | 27.05 | 135.40 | 5.01 |
| SNHG4 | #N/A | 6.88 | 34.06 | 4.95 |
| RPL18 | #N/A | 7.37 | 36.51 | 4.95 |
| KRTCAP2 | karyopherin alpha 4 (importin alpha 3) | 6.51 | 32.24 | 4.95 |
| KIF27 | #N/A | 9.88 | 48.75 | 4.94 |
| EEF1A1 | #N/A | 59.09 | 290.46 | 4.92 |
| TFAM | #N/A | 18.14 | 88.90 | 4.90 |
| PAFAH1B1 | poly(A) binding protein, cytoplasmic 1 | 13.95 | 68.36 | 4.90 |
| TMBIM4 | #N/A | 14.07 | 68.86 | 4.89 |
| HIST1H2BN | #N/A | 6.69 | 32.44 | 4.85 |
| PPP1R42 | protein phosphatase 1, regulatory (inhibitor) subunit 2 | 39.21 | 189.94 | 4.84 |
| ROPN1 | reactive oxygen species modulator 1 | 84.95 | 408.68 | 4.81 |
| SS18L2 | #N/A | 6.65 | 31.90 | 4.80 |
| SS18L2 | #N/A | 6.65 | 31.90 | 4.80 |
| COX6B1 | #N/A | 17.02 | 80.78 | 4.75 |
| EIF4G2 | eukaryotic translation initiation factor 4 gamma, 2 | 11.49 | 54.34 | 4.73 |
| MMADHC | myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11 | 15.80 | 74.69 | 4.73 |
| FXR1 | fragile X mental retardation, autosomal homolog 1 | 46.08 | 217.73 | 4.72 |
| RPL4 | ribosomal protein L39 | 13.38 | 62.92 | 4.70 |
| SRSF2 | #N/A | 5.49 | 25.83 | 4.70 |
| MESDC2 | maternal embryonic leucine zipper kinase | 5.53 | 25.99 | 4.70 |
| BLZF1 | basic leucine zipper nuclear factor 1 | 16.20 | 75.93 | 4.69 |
| VDAC2 | vasohibin 2 | 16.24 | 75.77 | 4.67 |
| RPL5 | similar to ribosomal protein L4; ribosomal protein L4 | 7.67 | 35.49 | 4.63 |
| CYLC1 | cylicin, basic protein of sperm head cytoskeleton 1 | 33.50 | 150.68 | 4.50 |
| FRG1 | #N/A | 16.15 | 72.48 | 4.49 |
| FAM81B | family with sequence similarity 81, member B | 42.17 | 187.90 | 4.46 |
| PICALM | phytanoyl-CoA 2-hydroxylase interacting protein-like | 15.76 | 69.65 | 4.42 |
| RPS26 | #N/A | 15.98 | 70.29 | 4.40 |
| FKBP2 | #N/A | 5.40 | 23.41 | 4.33 |
| AMZ2 | archaelysin family metallopeptidase 2 | 11.71 | 50.73 | 4.33 |
| EID3 | EP300 interacting inhibitor of differentiation 3 | 6.21 | 26.89 | 4.33 |
| CHCHD7 | coiled-coil-helix-coiled-coil-helix domain containing 7 | 10.27 | 44.51 | 4.33 |
| CETN4 | centrin 4 | 77.47 | 329.77 | 4.26 |
| IGFBP7 | insulin-like growth factor binding protein 7 | 5.16 | 21.29 | 4.13 |

| | | | | |
|---|---|---|---|---|
| STMN1 | serine/threonine kinase 17a | 12.21 | 50.37 | 4.13 |
| MORF4L1 | meiosis-specific nuclear structural 1 | 52.56 | 216.69 | 4.12 |
| NDUFB4 | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa | 81.78 | 336.24 | 4.11 |
| SBDS | similar to Cytokine induced protein 29 kDa; SAP domain containing ribonucleoprotein | 11.24 | 46.09 | 4.10 |
| DYDC2 | DPY30 domain containing 2 | 19.33 | 78.96 | 4.08 |
| DAD1 | defender against cell death 1 | 99.92 | 406.94 | 4.07 |
| MFF | methyltransferase like 13 | 44.85 | 181.10 | 4.04 |
| GML | #N/A | 5.78 | 23.25 | 4.02 |
| GGNBP2 | gametogenetin binding protein 2 | 19.12 | 76.33 | 3.99 |
| RPGRIP1 | RAR-related orphan receptor A | 20.51 | 80.71 | 3.94 |
| COX7B2 | #N/A | 58.19 | 228.08 | 3.92 |
| ADORA3 | adenosine A3 receptor | 35.26 | 137.83 | 3.91 |
| C16H1orf100 | chromosome 1 open reading frame 100 ortholog | 61.96 | 241.06 | 3.89 |
| YBX2 | Y box binding protein 1 | 34.08 | 132.03 | 3.87 |
| ADAM3A | ADAM metallopeptidase domain 3A (cyritestin 1) | 36.65 | 141.01 | 3.85 |
| CCT2 | chaperonin containing TCP1, subunit 2 (beta) | 18.39 | 70.57 | 3.84 |
| NDUFB3 | #N/A | 35.05 | 133.77 | 3.82 |
| RPS27A | #N/A | 223.62 | 852.29 | 3.81 |
| SKP1 | signal-regulatory protein delta; similar to signal-regulatory protein delta | 39.11 | 149.05 | 3.81 |
| C23H6orf106 | #N/A | 13.58 | 51.64 | 3.80 |
| HMGB1 | high-mobility group box 1-like 1; high-mobility group box 1; similar to Hmgb1 protein | 12.31 | 46.65 | 3.79 |
| RPL17 | #N/A | 18.74 | 70.74 | 3.78 |
| GPI | glucose phosphate isomerase | 13.89 | 52.13 | 3.75 |
| SNRPG | small nuclear ribonucleoprotein polypeptide F | 29.55 | 110.44 | 3.74 |
| NDUFB1 | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, assembly factor 3 | 22.98 | 85.34 | 3.71 |
| RPL39 | ribosomal protein L38 | 19.44 | 72.19 | 3.71 |
| CAPZB | capping protein (actin filament) muscle Z-line, beta | 30.03 | 111.53 | 3.71 |
| UBXN6 | UBX domain protein 4 | 5.51 | 20.45 | 3.71 |
| NDUFS5 | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa | 11.39 | 42.30 | 3.71 |
| C3H1orf189 | chromosome 1 open reading frame 189 ortholog | 70.91 | 261.32 | 3.69 |
| BCAP29 | B-cell receptor-associated protein 29 | 62.95 | 231.40 | 3.68 |
| GNG5 | #N/A | 18.11 | 64.44 | 3.56 |
| FAM71E1 | #N/A | 11.19 | 39.82 | 3.56 |
| ATP1B3 | ATPase, Na+/K+ transporting, beta 3 polypeptide | 56.01 | 198.96 | 3.55 |
| RPL38 | ribosomal protein L37a | 180.85 | 639.59 | 3.54 |
| SPATA9 | #N/A | 29.98 | 105.73 | 3.53 |
| HSP90B1 | tumor rejection antigen (gp96) 1 | 19.68 | 69.27 | 3.52 |
| ROPN1L | ropporin, rhophilin associated protein 1 | 14.79 | 51.87 | 3.51 |
| NDUFA6 | NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa | 65.71 | 229.21 | 3.49 |
| PSMB7 | proteasome (prosome, macropain) subunit, beta type, 3 | 16.95 | 59.01 | 3.48 |
| RNF38 | ring finger protein 216 | 19.27 | 66.93 | 3.47 |
| TOMM70A | translocase of outer mitochondrial membrane 7 homolog (yeast) | 18.52 | 64.30 | 3.47 |
| PSMG2 | proteasome (prosome, macropain) assembly chaperone 1 | 28.63 | 98.73 | 3.45 |
| SHFM1 | #N/A | 41.97 | 144.73 | 3.45 |

| | | | | |
|---|---|---|---|---|
| C15H11orf71 | chromosome 11 open reading frame 71 ortholog | 17.83 | 61.47 | 3.45 |
| TMBIM6 | transmembrane BAX inhibitor motif containing 4 | 26.85 | 92.08 | 3.43 |
| PPWD1 | protein phosphatase 2, regulatory subunit B', gamma isoform | 17.56 | 60.04 | 3.42 |
| EIF3K | eukaryotic translation initiation factor 3, subunit K | 5.59 | 19.01 | 3.40 |
| CSRP3 | cysteine and glycine-rich protein 3 (cardiac LIM protein) | 9.07 | 30.86 | 3.40 |
| ATP5E | #N/A | 50.85 | 173.08 | 3.40 |
| TMSB10 | transmembrane protein 60 | 265.87 | 897.20 | 3.37 |
| USP15 | ubiquinol-cytochrome c reductase hinge protein | 7.42 | 25.05 | 3.37 |
| ELK4 | ELK4, ETS-domain protein (SRF accessory protein 1) | 5.58 | 18.65 | 3.34 |
| ETF1 | eukaryotic translation termination factor 1 | 6.14 | 20.41 | 3.33 |
| TAF9 | TAF10 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 30kDa | 28.71 | 95.44 | 3.32 |
| DNAJB3 | #N/A | 67.47 | 224.17 | 3.32 |
| C4H7orf10 | #N/A | 14.02 | 46.57 | 3.32 |
| PFN1 | prefoldin subunit 5 | 7.76 | 25.63 | 3.30 |
| EIF5 | eukaryotic translation initiation factor 5 | 36.95 | 121.20 | 3.28 |
| MELK | meiosis expressed gene 1 homolog (mouse) | 7.79 | 25.42 | 3.26 |
| ROMO1 | RNA (guanine-7-) methyltransferase | 15.80 | 51.35 | 3.25 |
| RPL10A | #N/A | 19.30 | 62.21 | 3.22 |
| STAT4 | signal transducing adaptor molecule (SH3 domain and ITAM motif) 2 | 22.73 | 73.23 | 3.22 |
| NBN | N-acyl phosphatidylethanolamine phospholipase D | 6.88 | 22.13 | 3.22 |
| YPEL2 | YME1-like 1 (S. cerevisiae) | 11.53 | 37.11 | 3.22 |
| SLC30A3 | #N/A | 8.94 | 28.78 | 3.22 |
| SCAMP1 | Shwachman-Bodian-Diamond syndrome | 24.97 | 79.99 | 3.20 |
| NME2 | NIPA-like domain containing 3 | 13.73 | 43.97 | 3.20 |
| CALU | calumenin | 5.34 | 17.06 | 3.20 |
| ARL2BP | ADP-ribosylation factor-like 2 binding protein | 13.28 | 42.46 | 3.20 |
| CISD1 | CDGSH iron sulfur domain 1 | 354.71 | 1133.38 | 3.20 |
| AIF1 | allograft inflammatory factor 1 | 95.89 | 305.21 | 3.18 |
| DNAJB4 | DnaJ (Hsp40) homolog, subfamily B, member 4 | 48.84 | 154.38 | 3.16 |
| SF3B1 | stress-associated endoplasmic reticulum protein family member 2 | 7.22 | 22.74 | 3.15 |
| HSP90AA1 | #N/A | 43.10 | 135.28 | 3.14 |
| PGP | phosphoglycerate kinase 2 | 11.81 | 36.94 | 3.13 |
| LAPTM4A | #N/A | 55.64 | 173.85 | 3.12 |
| CYLC2 | cylicin, basic protein of sperm head cytoskeleton 2 | 138.04 | 427.82 | 3.10 |
| C1D | C1D nuclear receptor co-repressor | 16.78 | 51.94 | 3.09 |
| PFN2 | #N/A | 6.08 | 18.82 | 3.09 |
| CIAPIN1 | cytokine induced apoptosis inhibitor 1 | 10.35 | 32.02 | 3.09 |
| LOC100125913 | leukemia NUP98 fusion partner 1 | 9.85 | 30.50 | 3.09 |
| PDHB | pyruvate dehydrogenase (lipoamide) alpha 2 | 6.73 | 20.82 | 3.09 |
| ACOT7 | acyl-CoA thioesterase 7 | 5.15 | 15.94 | 3.09 |
| NDUFB2 | NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 1, 7kDa | 10.39 | 32.14 | 3.09 |
| CABYR | calcium binding tyrosine-(Y)-phosphorylation regulated | 21.15 | 65.21 | 3.08 |
| FAM92A1 | #N/A | 27.53 | 83.42 | 3.03 |
| C5H12orf50 | chromosome 12 open reading frame 50 ortholog | 28.76 | 86.78 | 3.02 |
| SPESP1 | spermatid maturation 1 | 11.12 | 33.56 | 3.02 |

| | | | | |
|---|---|---|---|---|
| CUL3 | cullin 3 | 15.33 | 45.85 | 2.99 |
| IQCF1 | integrator complex subunit 12 | 36.67 | 109.69 | 2.99 |
| TMEM258 | transmembrane and coiled-coil domains 5B | 101.28 | 301.68 | 2.98 |
| SERF1A | SERPINE1 mRNA binding protein 1 | 162.97 | 485.41 | 2.98 |
| ARL4A | ADP-ribosylation factor-like 4A | 75.36 | 224.41 | 2.98 |
| COIL | coilin | 9.14 | 27.08 | 2.96 |
| BRD2 | bromodomain containing 2 | 6.47 | 19.13 | 2.96 |
| RABEPK | RAB5A, member RAS oncogene family | 28.56 | 84.21 | 2.95 |
| HSPA5 | heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa) | 20.32 | 59.86 | 2.95 |
| VTI1B | very low density lipoprotein receptor | 34.70 | 101.74 | 2.93 |
| FAM229B | #N/A | 965.33 | 2823.04 | 2.92 |
| PFDN5 | phosphatidylethanolamine-binding protein 4 | 108.54 | 317.07 | 2.92 |
| LYZL1 | Lyrm7 homolog (mouse) | 33.52 | 97.51 | 2.91 |
| TMA16 | #N/A | 19.54 | 56.43 | 2.89 |
| HSP90AB1 | heat shock 90kDa protein 1, beta | 9.29 | 26.62 | 2.87 |
| HSP90AB1 | heat shock 90kDa protein 1, beta | 9.29 | 26.62 | 2.87 |
| MEIG1 | mediator complex subunit 6 | 117.34 | 334.29 | 2.85 |
| NDUFA4 | #N/A | 93.47 | 265.58 | 2.84 |
| SPINK2 | sperm equatorial segment protein 1 | 192.07 | 545.75 | 2.84 |
| UBB | #N/A | 384.20 | 1088.56 | 2.83 |
| TOX4 | translocase of outer mitochondrial membrane 70 homolog A (S. cerevisiae) | 9.26 | 26.20 | 2.83 |
| GSG1 | germ cell associated 1 | 24.89 | 70.19 | 2.82 |
| LELP1 | lactate dehydrogenase C | 36.39 | 102.60 | 2.82 |
| CDV3 | CDV3 homolog (mouse) | 27.46 | 77.42 | 2.82 |
| TERF2 | tektin 3 | 8.88 | 24.92 | 2.81 |
| RPS21 | #N/A | 62.43 | 173.88 | 2.79 |
| H3F3A | #N/A | 16.81 | 46.23 | 2.75 |
| BANF2 | barrier to autointegration factor 2 | 251.77 | 692.54 | 2.75 |
| AKTIP | AKT interacting protein | 14.65 | 40.21 | 2.75 |
| WBP2NL | Wiskott-Aldrich syndrome-like | 18.68 | 51.20 | 2.74 |
| BAZ2B | bromodomain adjacent to zinc finger domain, 2B | 14.22 | 38.93 | 2.74 |
| DCUN1D4 | DCN1, defective in cullin neddylation 1, domain containing 4 (S. cerevisiae) | 61.63 | 166.89 | 2.71 |
| MS4A13 | mitochondrial ribosomal protein S9 | 151.90 | 411.21 | 2.71 |
| CCDC54 | coiled-coil domain containing 54 | 69.36 | 186.97 | 2.70 |
| SRP54 | signal recognition particle 19kDa | 92.91 | 247.01 | 2.66 |
| DNAJA1 | DnaJ (Hsp40) homolog, subfamily A, member 1 | 15.92 | 42.16 | 2.65 |
| TPI1 | TOX high mobility group box family member 4 | 12.05 | 31.69 | 2.63 |
| RPS15 | #N/A | 22.07 | 58.05 | 2.63 |
| KIF2B | kinesin heavy chain member 2A | 12.48 | 32.83 | 2.63 |
| PTGR1 | proteasome (prosome, macropain) assembly chaperone 2 | 5.99 | 15.76 | 2.63 |
| SAA4 | S100 calcium binding protein G | 67.67 | 175.29 | 2.59 |
| ENO1 | enolase 1, (alpha) | 6.00 | 15.47 | 2.58 |
| WASL | vesicle transport through interaction with t-SNAREs homolog 1B | 5.82 | 15.01 | 2.58 |
| FAM103A1 | #N/A | 20.22 | 52.15 | 2.58 |
| PRPS1L1 | protein interacting with cyclin A1 | 6.03 | 15.45 | 2.56 |

| | | | | |
|---|---|---|---|---|
| FAU | Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed; similar to ubiquitin-like/S30 ribosomal fusion protein | 100.22 | 256.54 | 2.56 |
| FAU | Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed; similar to ubiquitin-like/S30 ribosomal fusion protein | 100.22 | 256.54 | 2.56 |
| C15H11orf88 | #N/A | 28.37 | 72.44 | 2.55 |
| AP2B1 | adaptor-related protein complex 2, beta 1 subunit | 94.45 | 240.14 | 2.54 |
| AP2B1 | adaptor-related protein complex 2, beta 1 subunit | 94.45 | 240.14 | 2.54 |
| TRIM36 | trafficking protein particle complex 3 | 20.97 | 53.24 | 2.54 |
| CNOT2 | CCR4-NOT transcription complex, subunit 2 | 8.19 | 20.73 | 2.53 |
| CLU | clusterin | 12.20 | 30.88 | 2.53 |
| MRPS18C | #N/A | 77.78 | 195.99 | 2.52 |
| RPL37A | #N/A | 375.70 | 945.04 | 2.52 |
| ZNF451 | zinc finger protein 432 | 9.83 | 24.70 | 2.51 |
| YWHAQ | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide | 82.64 | 206.93 | 2.50 |
| UQCRC2 | ubiquinol-cytochrome c reductase binding protein | 7.02 | 17.52 | 2.50 |
| RANGRF | #N/A | 95.84 | 238.84 | 2.49 |
| UBE2V1 | ubiquitin-conjugating enzyme E2R 2 | 17.63 | 43.66 | 2.48 |
| C19H17orf98 | #N/A | 6.02 | 14.90 | 2.48 |
| ARF4 | #N/A | 21.19 | 52.45 | 2.48 |
| CAP1 | similar to CAP1 protein; CAP, adenylate cyclase-associated protein 1 (yeast) | 7.49 | 18.53 | 2.48 |
| LOC504599 | histone H2B variant PT15 | 7.46 | 18.47 | 2.48 |
| RPS27 | #N/A | 120.56 | 298.47 | 2.48 |
| CCDC167 | #N/A | 8.88 | 21.98 | 2.48 |
| C11H2orf43 | #N/A | 8.07 | 19.97 | 2.48 |
| C15H11orf1 | #N/A | 39.83 | 98.29 | 2.47 |
| FSCN3 | fascin homolog 3, actin-bundling protein, testicular (Strongylocentrotus purpuratus) | 28.32 | 69.49 | 2.45 |
| FAM209B | #N/A | 328.88 | 805.05 | 2.45 |
| C21H14orf2 | #N/A | 339.04 | 826.64 | 2.44 |
| ATP5L | #N/A | 92.45 | 225.29 | 2.44 |
| GALNTL5 | UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 5 | 42.73 | 103.89 | 2.43 |
| EIF1 | #N/A | 87.71 | 212.72 | 2.43 |
| TEX101 | testis derived transcript (3 LIM domains); similar to testis derived transcript | 23.63 | 57.28 | 2.42 |
| PSME4 | #N/A | 5.55 | 13.46 | 2.42 |
| DNAJC21 | DnaJ (Hsp40) homolog, subfamily C, member 21 | 19.41 | 47.01 | 2.42 |
| C4H7orf62 | #N/A | 95.34 | 230.41 | 2.42 |
| PSMA6 | proteasome (prosome, macropain) subunit, alpha type, 4 | 537.25 | 1296.79 | 2.41 |
| MGC134093 | hypothetical protein MGC134066 | 128.66 | 310.48 | 2.41 |
| MGC134093 | #N/A | 128.66 | 310.48 | 2.41 |
| MGC134093 | #N/A | 128.66 | 310.48 | 2.41 |
| PIGF | protein interacting with PRKCA 1 | 58.77 | 141.32 | 2.40 |
| RAB5A | RAB28, member RAS oncogene family | 7.20 | 17.19 | 2.39 |
| KIF5C | kinesin family member 5B | 43.01 | 102.51 | 2.38 |
| TMEM30A | #N/A | 21.60 | 51.48 | 2.38 |
| TEX35 | #N/A | 111.07 | 262.35 | 2.36 |

| | | | | |
|---|---|---|---|---|
| HNRNPC | heterogeneous nuclear ribonucleoprotein C (C1/C2) | 11.00 | 25.95 | 2.36 |
| ARPM1 | actin related protein M1 | 21.94 | 51.76 | 2.36 |
| ERICH2 | #N/A | 364.32 | 858.46 | 2.36 |
| HNRNPU | heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A) | 8.51 | 20.06 | 2.36 |
| FNDC3A | fibronectin type III domain containing 3A | 11.09 | 26.11 | 2.35 |
| TSACC | hypothetical LOC540154 | 791.86 | 1857.74 | 2.35 |
| PRSS37 | protease, serine, 23 | 79.01 | 184.74 | 2.34 |
| CCDC101 | coiled-coil domain containing 101 | 24.12 | 56.39 | 2.34 |
| CCDC101 | coiled-coil domain containing 101 | 24.12 | 56.39 | 2.34 |
| RPS20 | #N/A | 82.82 | 193.11 | 2.33 |
| DBI | diazepam binding inhibitor (GABA receptor modulator, acyl-Coenzyme A binding protein) | 245.96 | 572.66 | 2.33 |
| DKKL1 | dickkopf-like 1 (soggy) | 8.27 | 19.18 | 2.32 |
| RPS7 | #N/A | 39.81 | 92.39 | 2.32 |
| LY6G6C | leucine zipper protein 1 | 18.64 | 43.26 | 2.32 |
| CALM2 | #N/A | 233.28 | 539.14 | 2.31 |
| BPIFA3 | #N/A | 119.02 | 272.68 | 2.29 |
| CDKL3 | cyclin-dependent kinase-like 3 | 10.09 | 23.10 | 2.29 |
| FBP1 | #N/A | 49.85 | 114.05 | 2.29 |
| KLHL2 | kelch-like 10 (Drosophila) | 8.41 | 19.23 | 2.29 |
| ALB | albumin | 6.09 | 13.91 | 2.28 |
| TMCO2 | transmembrane BAX inhibitor motif containing 6 | 152.54 | 348.14 | 2.28 |
| RPSA | ribosomal protein S8 | 11.98 | 27.18 | 2.27 |
| MED6 | mediator complex subunit 31 | 11.43 | 25.94 | 2.27 |
| HNRNPD | heterogeneous nuclear ribonucleoprotein D (AU-rich element RNA binding protein 1, 37kDa) | 7.24 | 16.44 | 2.27 |
| RPL13 | #N/A | 19.73 | 44.76 | 2.27 |
| RPL32 | similar to ribosomal protein L31; ribosomal protein L31 | 283.92 | 641.10 | 2.26 |
| RPS3A | ribosomal protein S3 | 224.15 | 506.05 | 2.26 |
| ATP6V1C1 | ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C1 | 33.73 | 75.89 | 2.25 |
| RPL22L1 | #N/A | 14.43 | 32.27 | 2.24 |
| UBXN1 | #N/A | 9.07 | 20.21 | 2.23 |
| KIF5B | kinesin family member 3A | 5.39 | 11.81 | 2.19 |
| RPL35A | ribosomal protein L35 | 238.98 | 519.68 | 2.17 |
| SPATA6 | spermatogenesis associated 4 | 314.58 | 681.94 | 2.17 |
| RBM4B | RNA binding motif protein 46 | 30.40 | 65.85 | 2.17 |
| C9H6orf165 | chromosome 6 open reading frame 165 ortholog | 6.14 | 13.30 | 2.17 |
| PLCZ1 | phospholipase C, delta 4 | 27.33 | 58.96 | 2.16 |
| C1H3orf38 | chromosome 3 open reading frame 38 ortholog | 88.13 | 190.05 | 2.16 |
| KLHDC10 | kinesin family member 5C | 19.54 | 42.01 | 2.15 |
| COPS4 | COP9 constitutive photomorphogenic homolog subunit 4 (Arabidopsis) | 32.03 | 68.68 | 2.14 |
| GTF2H5 | general transcription factor IIH, polypeptide 5 | 50.85 | 108.57 | 2.14 |
| LPIN1 | hypothetical protein LOC788205 | 28.98 | 61.85 | 2.13 |
| UQCRB | ubiquitin interaction motif containing 1 | 245.77 | 523.54 | 2.13 |
| CALM3 | #N/A | 28.81 | 61.37 | 2.13 |
| TXN | tumor suppressor candidate 3 | 36.83 | 78.15 | 2.12 |
| ATOX1 | ATX1 antioxidant protein 1 homolog (yeast) | 92.30 | 195.86 | 2.12 |

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

**Appendix 2. Transcripts with > 2:1 ratios in higher:lower fertility populations**

| Official Gene Symbol | Gene Long Name | Higher FPKM | Lower FPKM | Higher/ Lower Ratio |
|---|---|---|---|---|
| UCHL1 | UBX domain protein 6 | 142.15 | 5.23679 | 27.14 |
| ENY2 | enhancer of yellow 2 homolog (Drosophila) | 113.50 | 5.66497 | 20.04 |
| PLAC1 | #N/A | 119.24 | 8.02166 | 14.86 |
| CETN2 | centrin, EF-hand protein, 2 | 62.39 | 5.67822 | 10.99 |
| BCL2L11 | BCL2-like 11 (apoptosis facilitator) | 141.55 | 15.5432 | 9.11 |
| EMC2 | #N/A | 87.94 | 10.5031 | 8.37 |
| LSM12 | limbic system-associated membrane protein | 43.79 | 5.5314 | 7.92 |
| NAPEPLD | #N/A | 40.84 | 5.20358 | 7.85 |
| CHRAC1 | chromatin accessibility complex 1 | 72.12 | 9.76486 | 7.39 |
| UBE2R2 | ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast) | 143.09 | 21.255 | 6.73 |
| COPS3 | COP9 constitutive photomorphogenic homolog subunit 3 (Arabidopsis) | 58.32 | 8.87618 | 6.57 |
| SH3BGRL3 | #N/A | 64.91 | 10.4808 | 6.19 |
| ASB12 | ankyrin repeat and SOCS box-containing 12 | 56.81 | 9.58963 | 5.92 |
| ANP32B | acidic (leucine-rich) nuclear phosphoprotein 32 family, member B | 42.45 | 7.50689 | 5.66 |
| AHCY | adenosylhomocysteinase | 33.02 | 6.04565 | 5.46 |
| MANF | maelstrom homolog (Drosophila) | 69.42 | 12.7308 | 5.45 |
| GLYATL3 | #N/A | 60.05 | 11.4356 | 5.25 |
| KPNA4 | kelch-like 24 (Drosophila) | 59.37 | 11.5279 | 5.15 |
| C11H9orf9 | chromosome 9 open reading frame 9 ortholog | 106.29 | 21.0511 | 5.05 |
| CDKL4 | cyclin-dependent kinase-like 4 | 93.75 | 18.686 | 5.02 |
| SPAG4 | sperm associated antigen 17 | 46.37 | 9.30735 | 4.98 |
| PTMA | prostaglandin reductase 1 | 34.13 | 7.04153 | 4.85 |
| RPS28 | ribosomal protein S27-like | 35.99 | 7.42415 | 4.85 |
| MIR1307 | missing oocyte, meiosis regulator, homolog (Drosophila) | 420.45 | 86.7418 | 4.85 |
| MRPS16 | mitochondrial ribosomal protein L42 | 29.61 | 6.10836 | 4.85 |
| SNRPF | small nuclear ribonucleoprotein D2 polypeptide 16.5kDa | 24.60 | 5.37417 | 4.58 |
| LSM2 | LSM12 homolog (S. cerevisiae) | 277.89 | 60.9775 | 4.56 |
| FAM76B | #N/A | 48.69 | 10.7459 | 4.53 |
| PRSS23 | proline rich 13 | 24.72 | 5.56424 | 4.44 |
| CEP76 | centrosomal protein 76kDa | 25.96 | 5.91952 | 4.39 |
| AMAC1L3 | acyl-malonyl condensing enzyme 1-like 3 | 27.76 | 6.442 | 4.31 |
| TRAPPC3 | transformer 2 beta homolog (Drosophila) | 74.24 | 17.5969 | 4.22 |
| SYNE4 | SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae) | 21.54 | 5.12758 | 4.20 |
| UBE2J1 | ubiquitin-conjugating enzyme E2H (UBC8 homolog, yeast) | 101.78 | 24.4097 | 4.17 |
| FBXL2 | F-box and leucine-rich repeat protein 2 | 33.65 | 8.12796 | 4.14 |
| PNN | #N/A | 23.51 | 5.69404 | 4.13 |
| CCNB2 | similar to cyclin B2; cyclin B2 | 34.12 | 8.2962 | 4.11 |
| LYRM4 | #N/A | 52.41 | 12.9745 | 4.04 |
| C23H6orf25 | #N/A | 39.98 | 9.99186 | 4.00 |
| SEC61A2 | succinate dehydrogenase complex, subunit B, iron sulfur (Ip) | 70.32 | 17.871 | 3.94 |
| YAE1D1 | #N/A | 69.71 | 17.8064 | 3.92 |
| PMPCB | non-protein coding RNA 153 | 34.91 | 9.06111 | 3.85 |
| WDR53 | WBP2 N-terminal like | 64.89 | 17.7455 | 3.66 |
| GALNT3 | UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 3 (GalNAc-T3) | 39.51 | 10.8672 | 3.64 |
| EIF3G | eukaryotic translation initiation factor 3, subunit G | 19.08 | 5.36828 | 3.55 |
| ASB8 | ankyrin repeat and SOCS box-containing 8 | 50.71 | 14.4483 | 3.51 |
| HIST1H2AG | #N/A | 23.82 | 6.80562 | 3.50 |
| RPS3 | #N/A | 236.85 | 68.2526 | 3.47 |
| SSRP1 | #N/A | 65.61 | 19.085 | 3.44 |
| BTBD10 | BTB (POZ) domain containing 10 | 44.81 | 13.1859 | 3.40 |
| DYNLRB2 | dynein, light chain, roadblock-type 2 | 260.34 | 77.7852 | 3.35 |
| IQCF2 | IQ motif containing F1 | 120.31 | 35.9479 | 3.35 |
| GOSR1 | golgi SNAP receptor complex member 1 | 61.50 | 18.5172 | 3.32 |
| REEP6 | RNA binding motif protein 4B | 28.81 | 8.79414 | 3.28 |

| | | | | |
|---|---|---|---|---|
| SDHA | #N/A | 17.34 | 5.36623 | 3.23 |
| C11H9orf16 | chromosome 9 open reading frame 16 ortholog | 38.57 | 11.9349 | 3.23 |
| SRP19 | spermatogenic leucine zipper 1 | 24.08 | 7.45045 | 3.23 |
| ADK | #N/A | 23.34 | 7.22263 | 3.23 |
| SCP2D1 | #N/A | 743.18 | 232.871 | 3.19 |
| GTSF1L | gametocyte specific factor 1-like | 862.80 | 276.158 | 3.12 |
| RFWD2 | receptor accessory protein 6 | 35.98 | 11.5462 | 3.12 |
| CDC27 | cell division cycle 27 homolog (S. cerevisiae) | 20.91 | 6.86248 | 3.05 |
| WFDC15B | #N/A | 33.28 | 10.9419 | 3.04 |
| CCDC67 | coiled-coil domain containing 67 | 52.53 | 17.417 | 3.02 |
| AGPAT9 | 1-acylglycerol-3-phosphate O-acyltransferase 9 | 19.66 | 6.5189 | 3.02 |
| SERP1 | small EDRK-rich factor 2 | 56.38 | 19.0346 | 2.96 |
| LOC520057 | #N/A | 114.95 | 39.0595 | 2.94 |
| ZDHHC20 | zinc finger, CCHC domain containing 6 | 14.93 | 5.1975 | 2.87 |
| SERBP1 | SUMO1/sentrin specific peptidase 6 | 16.13 | 5.61632 | 2.87 |
| TMEM41B | transmembrane protein 30A | 16.67 | 5.80419 | 2.87 |
| ILF2 | interleukin enhancer binding factor 2, 45kDa | 207.66 | 72.6433 | 2.86 |
| GDE1 | glycerophosphodiester phosphodiesterase 1 | 33.40 | 11.7081 | 2.85 |
| CYB5R1 | cytochrome b5 reductase 1 | 21.55 | 7.55951 | 2.85 |
| DNAJC7 | DnaJ (Hsp40) homolog, subfamily C, member 7 | 19.15 | 6.77096 | 2.83 |
| RHOBTB1 | ring finger and WD repeat domain 2 | 18.93 | 6.69607 | 2.83 |
| CFL1 | cofilin 1 (non-muscle) | 60.20 | 21.4534 | 2.81 |
| MCFD2 | MARCKS-like 1; similar to MARCKS-related protein (MARCKS-like protein 1) (Macrophage myristoylated alanine-rich C kinase substrate) (Mac-MARCKS) (MacMARCKS) (Brain protein F52) | 17.65 | 6.32279 | 2.79 |
| WDR96 | WD repeat domain 74 | 14.69 | 5.33539 | 2.75 |
| TAF10 | synaptophysin-like 1 | 354.99 | 129.035 | 2.75 |
| FAM76A | family with sequence similarity 76, member A | 42.42 | 15.6108 | 2.72 |
| BTG1 | B-cell translocation gene 1, anti-proliferative | 30.42 | 11.2972 | 2.69 |
| GLOD4 | glyoxalase domain containing 4 | 64.59 | 24.1218 | 2.68 |
| AFF4 | AF4/FMR2 family, member 4 | 69.77 | 26.1796 | 2.67 |
| FAM98A | #N/A | 22.78 | 8.61676 | 2.64 |
| LIN7C | lipoic acid synthetase | 23.65 | 8.94574 | 2.64 |
| CCDC135 | coiled-coil domain containing 135 | 18.51 | 7.01963 | 2.64 |
| EIF5A | eukaryotic translation initiation factor 5A | 79.68 | 30.2392 | 2.63 |
| ATP5I | ATP synthase, H+ transporting, mitochondrial F0 complex, subunit e | 200.95 | 76.5374 | 2.63 |
| USP44 | ubiquitin specific peptidase 16 | 45.64 | 17.547 | 2.60 |
| UAP1 | thioredoxin-like 1 | 36.26 | 13.9643 | 2.60 |
| HNRNPH1 | #N/A | 22.54 | 8.71883 | 2.59 |
| CCDC70 | coiled-coil domain containing 70 | 25.22 | 9.93395 | 2.54 |
| ANKRD32 | ankyrin repeat domain 32 | 23.33 | 9.34499 | 2.50 |
| ACTL7A | actin-like 7A | 44.54 | 17.8814 | 2.49 |
| IQCG | IQ motif containing F5 | 115.05 | 46.2528 | 2.49 |
| SSR1 | #N/A | 23.27 | 9.37931 | 2.48 |
| TEKT3 | tektin 1 | 25.22 | 10.1809 | 2.48 |
| RNMT | similar to ring finger protein 38; ring finger protein 38 | 28.31 | 11.478 | 2.47 |
| MCL1 | #N/A | 19.43 | 8.01607 | 2.42 |
| NXN | #N/A | 16.39 | 6.76194 | 2.42 |
| SPATA6L | spermatogenesis associated 6 | 89.91 | 37.4999 | 2.40 |
| LSAMP | #N/A | 14.72 | 6.14617 | 2.40 |
| VIMP | vimentin | 101.90 | 42.9584 | 2.37 |
| VGLL3 | similar to voltage-dependent anion channel 2; voltage-dependent anion channel 2 | 94.26 | 39.7495 | 2.37 |
| POLB | pinin, desmosome associated protein | 63.50 | 26.7976 | 2.37 |
| FZD7 | frizzled homolog 7 (Drosophila) | 13.86 | 5.88325 | 2.36 |
| FN3KRP | fructosamine 3 kinase related protein | 40.66 | 17.254 | 2.36 |
| APOPT1 | #N/A | 291.65 | 124.488 | 2.34 |
| ACAT1 | acetyl-Coenzyme A acetyltransferase 1 | 99.09 | 42.3472 | 2.34 |
| LOC788205 | #N/A | 15.40 | 6.60052 | 2.33 |
| BAZ1A | bromodomain adjacent to zinc finger domain, 1A | 17.47 | 7.5322 | 2.32 |
| RN45S | RNA component of mitochondrial RNA processing endoribonuclease | 258.48 | 113.249 | 2.28 |
| RN45S | #N/A | 258.48 | 113.249 | 2.28 |

| | | | | |
|---|---|---|---|---|
| RN45S | #N/A | 258.48 | 113.249 | 2.28 |
| RN45S | #N/A | 258.48 | 113.249 | 2.28 |
| TM9SF2 | T-cell immunoglobulin and mucin domain containing 4 | 31.74 | 14.211 | 2.23 |
| TM9SF2 | #N/A | 31.74 | 14.211 | 2.23 |
| RAD21 | RAB interacting factor | 41.18 | 18.4569 | 2.23 |
| ZNF706 | hypothetical protein LOC100125413 | 180.19 | 81.1021 | 2.22 |
| PSMD1 | proteasome (prosome, macropain) subunit, beta type, 7 | 18.26 | 8.30929 | 2.20 |
| CNBP | CCHC-type zinc finger, nucleic acid binding protein | 30.43 | 13.921 | 2.19 |
| CHMP5 | chromatin modifying protein 5 | 4359.80 | 1995.9 | 2.18 |
| GSTO2 | glutathione S-transferase omega 2 | 121.97 | 56.3501 | 2.16 |
| CSN3 | casein kappa | 19.71 | 9.14845 | 2.15 |
| MND1 | methylmalonic aciduria (cobalamin deficiency) cblD type, with homocystinuria | 58.49 | 27.1507 | 2.15 |
| RORA | ropporin 1-like | 53.29 | 24.7343 | 2.15 |
| CSPP1 | centrosome and spindle pole associated protein 1 | 12.16 | 5.64413 | 2.15 |
| ZNF33B | zinc finger protein 24 | 23.93 | 11.3243 | 2.11 |
| HDGFRP3 | #N/A | 54.71 | 26.1645 | 2.09 |
| FGA | fibrinogen alpha chain | 14.62 | 7.03925 | 2.08 |
| BCKDHB | branched chain keto acid dehydrogenase E1, beta polypeptide | 20.81 | 10.1199 | 2.06 |
| SHOC2 | #N/A | 23.88 | 11.6139 | 2.06 |
| MLL5 | myeloid leukemia factor 1 | 10.48 | 5.131 | 2.04 |
| YWHAE | yippee-like 2 (Drosophila) | 24.40 | 11.9553 | 2.04 |
| RPL23A | ribosomal protein L23 | 34.30 | 16.8916 | 2.03 |
| PPM1G | #N/A | 40.24 | 19.8221 | 2.03 |
| APOO | apolipoprotein O | 59.03 | 29.2261 | 2.02 |
| TUBB | testis-specific serine kinase 3 | 10.77 | 5.33413 | 2.02 |
| ZNF428 | zinc finger protein 33B | 13.24 | 6.55637 | 2.02 |
| STRBP | #N/A | 21.34 | 10.5645 | 2.02 |
| MAP9 | mitogen-activated protein kinase kinase kinase kinase 3 | 40.52 | 20.1516 | 2.01 |
| MAP9 | microtubule-associated protein 9 | 40.52 | 20.1516 | 2.01 |

## APPENDIX I

**Spermatozoal RNA amount and its relationship with fertility and conventional semen parameters**

**Card CJ and Sartini BL**

**Abstract**

While current *in vitro* male fertility tests can identify infertile sires, they are insufficient at detecting differences in individuals that show similar levels of fertility or sub-fertility. The need for more accurate in vitro fertility tests for artificial insemination (AI) sires in the dairy industry is apparent. Currently *in vivo* fertility tests such as sire conception rate (SCR) require hundreds of inseminations per individual bull which requires an excess of time and resources. Mirroring *in vivo* fertility ratings accurately with a fast and reproducible *in vitro* fertility test would economically benefit the dairy industry while also aiding the development of new methods for measuring sperm quality in humans. The goal of this study was to determine if RNA content of transcriptionally and translationally silent bovine spermatozoa is an indicator of sperm motility, sperm DNA fragmentation, sperm head morphology and *in vivo* fertility. A column based RNA isolation procedure (Method #1) was developed for bovine spermatozoal RNA. Although this RNA isolation method yielded RNA with acceptable 260/280 ratios (1.85) and no detectable somatic cell contamination, transcripts known to be present in spermatozoa were not amplified in RNA samples consistently decreasing our confidence that this method yielded representative and pure spermatozoal RNA. Twenty-eight bulls with known *in vivo* fertility scores (CR) were used for analysis of RNA amount, motility, sperm head morphology, and DNA fragmentation. No significant correlations were found via Pearson correlation analysis between RNA/10 million spermatozoa and CR score, %

motility, percent DNA fragmentation (% DFI), or sperm head morphology harmonics. When ranking bulls by RNA/10 million spermatozoa, no significant differences in fertility or semen parameters were found when comparing splits of lowest 14/highest 14 or lowest 10/highest 10 individuals. These results show that RNA amount isolated with this method does not appear to be a good predictor of fertility or semen quality across a population of bulls with a wide range of fertility and semen quality parameters. The lack of validation, measured by the ability to amplify spermatozoal transcripts via PCR, of this spermatozoal RNA isolation method by amplification of known transcripts also brings into question the accuracy of these spermatozoal RNA amounts.

**Introduction**

Male *in vitro* fertility assays aim to correlate semen parameters with *in vivo* fertility as accurately as possible. Several sperm traits have been in used recently as in vitro fertility assays including scrotal circumference, sperm concentration, motility, morphology, and DNA fragmentation (Foote, 2003). Higher percent motility is associated with higher fertility as it increases the chances of sperm reaching the site of fertilization. Sperm head morphology is a conventional semen parameter often analyzed and has shown correlations with *in vivo* fertility (Al-Makhzoomi et al, 2008; Nagy et al, 2013). Large or misshaped sperm heads could hinder motility or be indicators of improper DNA compaction, leading to high susceptibility to DNA damage (Ostermeier et al, 2001b). Higher fertility individuals tend to have more elongated sperm heads than their lower fertility counterparts, possibly to due proper chromatin structure (Ostermeier et al, 2001b). Other sperm head measurements, classified as harmonics, have shown

correlations with fertility, and may be able to act as markers for bull fertility (Ostermeier et al, 2001a). Normal head morphology has been negatively correlated with DNA integrity as measured the sperm chromatin structure assay (SCSA), a flow cytometric based assay, in humans and bulls (Garcia-Macias et al, 2007; Nagy et al, 2013; Omran et al, 2013).

During sperm production, tight compaction of genomic DNA is necessary to prevent damage from free radicals, oxidative stress, nucleases, and other factors (Evenson et al, 2002). When DNA is not compacted properly, the chromatin is cleaved; causing DNA fragmentation and impaired gene transcription. Sperm DNA fragmentation has been correlated with decreased sperm function and can predict sire fertility more accurately than sperm motility, morphology, and concentration (Aoki et al, 2005; Agarwal et al, 2003; Madrid-Bury et al, 2005). DNA fragmentation is not only a problem for fertilization but post-fertilization as well. Embryo quality is negatively impacted when spermatozoa containing fragmented DNA are used for *in vitro* fertilization in mice, with no resulting embryos progressing to the blastocyst stage (Gawecka et al, 2013). Pregnancy is 3.5 times more likely to occur if lower fragmentation sperm is used for *in vitro* fertilization (Avendano et al, 2009). The percentage of DNA fragmentation in the sperm head is correlated with field fertility estimates for bulls, based on non-return rate. Semen with lower DNA fragmentation rates (1.6 - 3.8% fragmentation) have a 7% increase in successful artificial insemination (Waterhouse et al, 2006).

Although each of these sperm parameters have shown correlations with *in vivo* fertility, conventional semen parameters often cannot identify subfertile sires and are therefore not consistent indicators of high fertility (Colengrander et al, 2003). In the last

35 years, herd values for scrotal circumference, sperm concentration, and morphological traits have stayed consistent while fertility has steadily declined over the same period further highlighting the inability of these sperm traits to predict fertility (Dejarnette et al, 2004). More accurate fertility tests are needed, perhaps with more of a focus on sperm functionality, which could include looking at spermatozoal RNAs.

It has long been thought that the only function of spermatozoa was to deliver the paternal half of the genome to the egg for fertilization. However, more recently, spermatozoa have been shown to also deliver other components to the oocyte at fertilization that are necessary for successful early embryogenesis, including a population of RNAs from the transcriptionally-inert spermatozoa (as reviewed in Hamatani, 2012; Jodar et al, 2013). Spermatozoa contain lower amounts of RNA (5-400 fg/sperm depending on species) in contrast to oocytes (0.5-1.5 ng) and other cell types (10-30 pg) (Boerke et al, 2007; Das et al, 2010; Hamatani, 2012). The sperm RNA population includes coding and non-coding RNAs, such as mRNA, miRNA, siRNA, piRNA, and lncRNAs (Jodar et al, 2013).

The amount of total RNA is higher in morphologically normal sperm than that of abnormal sperm (Roudebush et al, 2004). Amounts of individual spermatozoal RNAs are correlated with sperm functional parameters such as sperm concentration, motility, morphology, and DNA fragmentation (Lambard et al, 2004; Aoki et al, 2005; Aoki et al, 2006). Specifically, sperm motility levels are correlated with the ratio of protamine transcripts, *Prm1/Prm2* ratios, in human sperm; high motility samples had a low ratio of *Prm1/Prm2* (Lambard et al, 2004). Increased *Prm1*/*Prm2* transcript levels and corresponding low protein levels in an individual could be an indicator of inefficient

translation during spermatogenesis (Aoki et al, 2006). Also, specific measurements of fertility, including *in vitro* fertilization, sperm concentration, motility, and morphology, are all negatively affected when the PRM1/PRM2 protein ratios are abnormally high or low (Aoki et al, 2006). DNA fragmentation, or the extent of damaged chromatin in the sperm nucleus, is significantly higher in sperm with low PRM1/PRM2 protein ratios (Aoki et al, 2005).

The goal of this study was to measure the amount of RNA per 10 million sperm cells and determine if the amount of sperm RNA is correlated with bovine semen parameters such as motility, morphology and DNA fragmentation as well as fertility rankings. Conventional semen parameters are not optimal for ranking individuals by fertility or detecting subtle differences between individuals of similar fertility ratings. This study is designed to explore a different parameter, sperm RNA amount, as a potential predictor for fertility.

**Materials and Methods**

*Semen Samples*

Cryopreserved semen from twenty-eight bull studs was obtained from Genex Cooperative Inc. (Shanawo, WI). The bulls had CR fertility scores that ranged from -4.7 to 2.2.

*DNA Fragmentation Analysis*

DNA fragmentation analysis was conducted on semen from the 28 bulls using the Sperm Chromatin Structure Assay (SCSA; Evenson et al, 2002)

*Head Morphology Analysis*

Morphology analysis was conducted on semen from the 28 bulls using Fourier harmonic amplitudes (Ostermeier et al, 2001a). Harmonic 0 is a measure of overall sperm head size. Harmonic 1 is a measure of the rounding of the anterior portion of the sperm head while harmonic 2 measures head elongation. The remaining harmonics (3-5) are measures of the shape of the posterior part of the head.

*Semen Thawing*

Frozen bovine semen straws (n=2 for each bull) were thawed in a 37°C water bath for one minute.

*Sperm-TALP Wash*

Semen was washed with 4 mls sperm-TALP media (100 mM NaCl, 3.1 mM KCl, 25 mM $NaHCO_3$, 0.3 mM $NaH_2PO_4$, 2 mM $CaCl_2 2H_2O$, 0.4 mM $MgCl_2 6H_2O$, 21.6 mM sodium lactate, 1 mM pyruvate, 6 mg/ml BSA, pH 7.3-7.4; Parrish et al, 1999) and centrifuged for 10 minutes at 800 *x g* to remove cryoprotective media. This sperm-TALP wash is repeated a second time, the supernatant is removed and the sperm pellet is subsequently used for RNA isolation.

*Sperm Motility/Concentration Analysis*

Sperm analysis was conducted to determine motility and concentration values on a Nikon Eclipse 50i microscope (MVI; Avon, MA). Five ul of 100 ul re-suspension solution was placed on a warmed slide to maintain motility. Percent motile sperm was subjectively

determined by the same person for all samples by recording motility of 100 sperm. Sperm concentration was determined with a hemocytometer. Numbers of sperm on each side of the hemocytometer were averaged together and the sample concentration was calculated using a standard hemocytometer formula based on dilution factor. Using concentration numbers, 10 x $10^6$ sperm cells from the remaining re-suspension solution were used in the RNA isolation for each bull.

*Sperm RNA Isolation (Method #1)*

The sperm pellets were re-suspended into 700 ul of buffer (10 ul 2-Mercaptoethanol + 1000 ul buffer RLT) for cell disruption (Qiagen; Valencia, CA). The samples were subsequently homogenized using a 26-gauge needle (3x) and extruded into RNase-free microfuge tubes. Subsequent RNA isolation was conducted with the RNeasy kit (Qiagen; Valencia, CA) per manufacturer's directions. Modifications to the standard RNeasy protocol included warming the elution water to 37° C and extending the final incubation of the sample on the column for 8 minutes per elution. RNA concentrations were measured using the NanoDrop UV/Vis Spectrometer (Thermo Scientific; Waltham, MA). RNA preparations with A260/A280 ratios over 1.80 (indicating pure RNA) were stored at -80°C until needed for further experiments. Amount of total RNA per $10^6$ spermatozoa for each sample was calculated by dividing the RNA concentration by the total number of sperm then multiplying by $10^6$

*Removal of genomic DNA*

RNA samples were DNAse treated using Turbo DNA-free Kit according to manufacturer's protocol (Ambion/Life Technologies; Carlsbad CA).

*Statistical Analysis*

For correlations between semen parameters, a Pearson's correlation test was used. When ranked by total RNA amount/10 million spermatozoa, motility, % DFI, and head morphology parameters H0-H5 were compared between various groupings of the 28 bulls. Groups analyzed included 14/14, and lowest 10/highest 10 when ranked by RNA amount/10 million spermatozoa and were subjected to unpaired t-tests.

**Results**

*Semen Analysis*

After RNA isolation and removal of genomic DNA, the total RNA per $10^6$ sperm ranged from 2,782 ng to 13,807 ng. During the RNA isolation, an aliquot of the sample was used for motility analysis, which ranged from 37.5% to 52.5% with one bull's DFI showed a variable range per bull from 6.1% to 30.5%. The CR, motility, % DFI, RNA/10 million sperm are presented in Table 1 and the sperm head morphology ratings are presented in Table 2.

*Relationships Between Fertility (CR score) and Semen Parameters*

Head morphology harmonics H0-H5 were not correlated with CR score, with the highest correlation at r = 0.164. CR score showed a moderate correlation with %

motility, $r^2 = 0.385$; $p = 0.043$ (Table 3). There was no correlation between CR score and % DNA fragmentation ($r^2 = -0.093$).

*Spermatozoal RNA Amount Related to Semen Parameters*

No high correlations were found between RNA/10 million sperm and any of the semen parameters (Table 3). No correlations were found between RNA/10 million spermatozoa and CR score as ($r^2 = 0.30$), % motility ($r^2 = 0.25$), and HA0 ($r^2 = 0.32$) were not significant. Correlations with other sperm parameters were between $-0.13 < r^2 > 0.13$ thus showing no correlation. A number of high correlations were found among the sperm morphology harmonics (HA0-HA5) ranging from 0.58 - 0.92 when compared to one another.

Bulls were ranked by RNA/10 million sperm and comparisons were analyzed for the 14 lowest RNA amount with 14 highest RNA amount then the 10 lowest RNA amount with the 10 highest RNA amount. No significant differences were found for CR score or any other semen parameters when ranked by RNA amount and separated into these lower and higher groups.

**Discussion**

A correlation of total spermatozoal RNA amount with bovine semen quality or fertility has not been explored previously. In this study, no significant correlations were found between total spermatozoal RNA amount and CR score or any semen parameters. A previous study reported a higher amount of total RNA in morphologically normal sperm than in of abnormal sperm (Roudebush et al, 2004). In this study however, RNA

amount had no correlation with the 6 sperm head morphological harmonics. It has been hypothesized that RNA amount could be related to morphology due to impaired transcriptional or translational pathways. If a greater amount of RNA is present, it may be an indicator of inefficient translation of key proteins that aid motility or chromatin packaging, thus causing abnormal morphology (Roudebush et al, 2004). There was only a very modest correlation between motility and RNA amount (approaching significance), not enough support for this hypothesis with the current data. Because transcripts known to be present in spermatozoa could not amplified in RNA samples consistently, the representation and purity of these RNA populations are questionable and definitive conclusions about sperm RNA amount and semen parameters cannot be made (see Appendix 2).

Though linear correlations were lacking between RNA amount and semen quality parameters, bulls were separated into groups based upon their spermatozoal RNA levels. There were still no significant differences between the low and high RNA groups for any parameters when the half and half split (14/14) and the lowest 10 and highest 10 (10/10) were compared. Total RNA amount seems to offer little relevance in predicting fertility or semen quality amongst bulls with variable fertility scores.

Compensable traits, those that can have deficiencies overcome to an extent by fertilizing with more semen, measured in this study included motility and head morphology harmonics. Though the focus of this study was to determine if RNA amount could predict fertility or any other semen parameters, other parameters were examined for relationships with fertility and the remaining semen parameters. Motility showed a modest correlation with CR score (r = 0.385) while none of the morphology harmonics

showed any correlation to CR score (highest at r = 0.164). Motility and morphology have long been used as predictors of sperm fertilization potential (Foote, 2003). However recent evidence supports the notion that these semen characteristics are very limited in their ability to predict fertility. Often these measures only pick out the most sub-fertile individuals from the population but have a hard time differentiating the rest of the individuals (Petrunkina et al, 2007).

Morphological status of semen has been shown to be a better predictor of fertility than some other compensable traits such as motility and concentration. Improper head morphology is detrimental to the fertilization potential of spermatozoa as it is often indicative of acrosomal defects and incomplete DNA compaction. Sperm head morphology is correlated with fertility in many species (Al-Makhzoomi et al, 2008; Gillan et al, 2008; Love, 2011; Nagy et al, 2013). Like motility, morphological measurements have shown variable levels of correlation with fertility in the past but individual head morphology harmonics did not show a link to fertility in this study.

DNA fragmentation, an uncompensable trait, has shown more promise as an indicator of semen quality and fertility. DNA fragmentation has shown correlations to both sperm head morphological status and motility in various studies (Giwercman et al, 2003; Moskovtsev et al, 2009). However, another study has shown that DNA fragmentation is only linked to sperm vitality and patient age while showing no relationship to motility, morphology, or concentration (Cohen-Bacrie et al, 2009). In this study there was no correlation between % DFI and fertility, motility, or RNA amount though negative correlations were found with three of the head morphology harmonics, possibly indicating that particular head shapes may be indicative of DNA fragmentation.

Although there was no link with fertility in this study, DNA fragmentation has been found to be a better predictor than compensable traits such as motility and morphology (Aoki et al, 2005; Agarwal et al, 2003; Madrid-Bury et al, 2005; Castilla et al, 2010).

The need for better *in vitro* fertility measurements that accurately predict *in vivo* fertility is readily apparent. Total RNA amount does not appear to predict sire fertility. However, a number of individual sperm mRNAs have shown links to fertility when expression levels are measured, including *PRM1*, *AK1*, *IB5*, *CD36*, *TIMP*, *SNRPN2*, *PLCZ1*, *CRISP2*, *CCT8*, and *PEBP1* (Feugang et al, 2009; Feugang et al, 2010; Arangasamy et al, 2011; Kasimanickam et al, 2012). It is apparent that sperm RNA can be a valuable diagnostic tool but it must be observed at the individual mRNA level and not in regards to total RNA quantity isolated with the column-based RNA isolation method.

**References**

Agarwal A and Said TM (2003): Role of sperm chromatin abnormalities and DNA damage in male infertility. Human Reproduction, 9(4): 331-345.

Al-Makhzoomi A, Lundeheim N, Haard M, Rodriguez-Martinez H (2008): Sperm morphology and fertility of progeny-tested AI dairy bulls in Sweden. Theriogenology, 70(4): 682-691.

Aoki VW, Moskovtsev SI, Willis J, Liu L, Mullen BM, Carrell DT (2005): DNA integrity is compromised in protamine-deficient human sperm. Journal of Andrology, 26: 741-748.

Aoki VW, Liu L, Jones KP, Hatasaka HH, Gibson M, Peterson MC, Carrell DT (2006): Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. Fertility and Sterility, 86: 1408-1415.

Arangasamy A, Kasimanickam VR, DeJarnette JM, Kasimanickam RK (2011): Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. Theriogenology, 76: 570-577.

Avendano C, Franchi A, Duran H, Oehninger S (2009): DNA fragmentation of normal spermatozoa negatively impacts embryo quality and intracytoplasmic sperm injection outcome. Fertility and Sterility, 94: 549-557.

Boerke A, Dieleman SJ, Gadella BM (2007): A possible role for sperm RNA in early embryo development. Theriogenology, 68S: 147-155.

Castilla JA, Zamora S, Gonzalvo MC, Luna del Castillo JD, Roldan-Nofuentes JA, Clavera A, Bjorndahl L, Martinez L (2010): Sperm chromatin structure assay and classical semen parameters: systematic review. Reproductive BioMedicine Online, 20: 114-124.

Cohen-Bacrie P, Belloc S, Menezo YJR, Clement P, Hamidi J, Benkhalifa M (2009): Correlation between DNA damage and sperm parameters: a prospective study of 1,633 patients. Fertility and Sterility, 91(5): 1801-1805.

Colenbrander B, Gadella BM, Stout TA (2003): The predictive value of semen analysis in the evaluation of stallion fertility. Reproduction in Domestic Animals, 38(4): 305-311.

Das PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, Varner DD, Chowdhary BP, Raudsepp T (2010): Total RNA isolation from stallion sperm and testis biopsies.Theriogenology, 74: 1099-1106

Dejarnette J, Marshall C, Lenz RW, Monke DR, Ayars WH, Sattler CG (2004): Sustaining the fertility of artificially inseminated dairy catle: the role of the artificial insemination industry. Journal of Dairy Science, 87: 93-104.

Evenson DP, Larson KL, Jost LK (2002): Sperm chromatin structure assay: its clinical use for detecting sperm DNA fragmentation in male infertility and comparisons with other techniques. Journal of Andrology, 23(1): 25-43.

Feugang JM, Kaya A, Page GP, 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 a potential role in bull fertility. BMC Genomics, 10: 176-185.

Feugang JM, Rodriguez-Osorio N, Kaya A, Wang H, Page G, Ostermeier GC, Topper EK, Memili E (2010): Transcriptome analysis of bull spermatozoa: implications for male fertility. Reproductive BioMedicine, 21: 312-324.

Foote RH (2003): Fertility estimation: a review of past experience and future prospects. Animal Reproduction Science, 75: 119-139.

Garcia-Macias, de Paz P, Martinez-Pastor F, Alvarez M, Gomes-Alves S, Bernardo J, Anel E, Anel L (2007): DNA fragmentation assessment by flow cytometry and Sperm-

Bos-Halomax (bright-field microscopy and fluorescence microscopy) in bull sperm. Animal Reproduction and Obstetrics, 2: 88-98.

Gawecka JE, Marh J, Ortega M, Yamauchi Y, Ward MA, Ward S (2013): Mouse zygotes respond to severe sperm DNA damage by delaying paternal DNA replication and embryonic development. PLOS One, 8(2): e56385.

Gillan L, Kroetsch T, Maxwell WMC, Evans G (2008): Assesment of *in vitro* sperm characteristics in relation to fertility in dairy bulls. Animal Reproduction Science, 103: 201-214.

Giwercman A, RIchthoff J, Hjollund H, Bonde JP, Jepson K, Frohm B, Spano M (2003): Correlation between sperm motility and sperm chromatin structure assay parameters. Fertility and Sterility, 80(6): 1404-1412.

Hamatani, T (2012): Human spermatozoal RNAs. Fertility and Sterility, 97(2): 275-281.

Jodar M, Selvaraju S, Sendler E, Diamond MP, Krawetz SA (2013): The presence, role and clinical use of spermatozoal RNAs. Human Reproduction Update, 0(0): 1-21.
Kasimanickam V, Kasimanickam R, Arangasamy A, Saberivand A, Stevenson JS, Kastelic JP (2012): Association between mRNA abundance of functional sperm function proteins and fertility of Holstein bulls. Theriogenology, 78(9): 2007-2019.

Lambard S, Galeraud-Denis I, Martin G, Levy R, Chocat A, Carreau S (2004): Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. Molecular Human Reproduction, 10: 535-541.

Love CC (2011): Relationship between sperm motility, morphology and the fertility of stallions. Theriogenology, 76: 547-557.

Madrid-Bury N, Perez-Guitierrez JF, Perez-Garnelo S, Moreira P, PintadoSanjuanbenito B, Gutierrez-Adan A, de la Fuente Martinez J (2005): Relationship between non-return rate and chromatin condensation of deep frozen bull spermatozoa. Theriogenology, 64: 232-241.

Moskovtsev SI, Willis J, White J, Mullen JBM (2009): Sperm DNA damage: correlation to severity of semen abnormalities. Infertility, 74(4): 789-793.

Nagy A, Johannisson A, Wahlsten T, Ijas R, Andersson M, Rodriguez-Martinez H (2013): Sperm chromatin structure and sperm morphology: Their association with fertility in the AI-dairy Ayrshire sires. Theriogenology, 79: 1153-1161.

Omran HM, Bakhiet M, Dashti MG (2013): DNA integrity is a critical molecular indicator for the assessment of male infertility. Molecular Medicine and Reproduction, 7(5): 1631-1635.

Ostermeier GC, Sargeant GA, Yandell BS, Parrish JJ (2001a): Measurement of bovine sperm nuclear shape using Fourier harmonic amplitudes. Journal of Andrology, 22(4): 584-594.

Ostermeier GC, Sargeant GA, Yandell BS, Evenson DP, Parrish JJ (2001b): Relationship of bull fertility to sperm nuclear shape. Journal of Andrology, 22(4): 595-603.

Parrish JJ, Susko-Parrish JL, Graham JK (1999): In vitro capacitation of bovine spermatozoa: role of intracellular calcium. Theriogenology, 51(2): 461-472.

Petrunkina AM, Waberski D, Gunzel-Apel AR, Topfer-Peterson E (2007): Determinants of sperm quality and fertility in domestic species. Reproduction, 134(1): 3-17.

Roudebush WE, Massey JB, Zhu J, Mitchel-Leef DE, Kort HI, Elsner CW (2004): Morphologically normal sperm have significantly greater total-RNA content than abnormal sperm. International Congress Series, 1271: 193-196.

Waterhouse KE, Haugan T, Kommisrud E, Tverdal A, Flatberg G, Farstad W, Evenson DP, De Angelis PM (2006): Sperm DNA damage is related to field fertility of semen from young Norwegian Red bulls. Reproduction, Fertility and Development, 18: 781-788.

**Table 1. CR score, RNA/10 million spermatozoa, % motility, and % DFI for Individual Bulls**

| Bull | CR Score | RNA/10 mill sperm (ng) | % Motility | % DFI |
|---|---|---|---|---|
| **1** | -4.7 | 2782.789 | 0 | 6.8 |
| **2** | -3.3 | 11648.07 | 37.5 | 11.7 |
| **3** | -2.9 | 11070.2 | 47 | 9.8 |
| **4** | -1.9 | 10650.5 | 48 | 14.4 |
| **5** | -1.5 | 5344.173 | 36.5 | 30.5 |
| **6** | -1.5 | 3037.502 | 51.5 | 9.1 |
| **7** | -1.4 | 9659.487 | 34 | 10.9 |
| **8** | -1 | 7424.739 | 51 | 13.1 |
| **9** | -0.9 | 10405.7 | 46.5 | 21.5 |
| **10** | -0.9 | 6003.03 | 49.5 | 21.1 |
| **11** | -0.7 | 7309.06 | 35.5 | 7.9 |
| **12** | -0.6 | 9042.087 | 58 | 10.9 |
| **13** | 0.7 | 11326.42 | 45 | 11.1 |
| **14** | 0.8 | 11936.45 | 38.5 | 14.4 |
| **15** | 0.8 | 8714.513 | 34 | 12.7 |
| **16** | 0.8 | 8105.15 | 23.5 | 12 |
| **17** | 1 | 10636.82 | 45.5 | 11.8 |
| **18** | 1.1 | 8535.056 | 46 | 12.7 |
| **19** | 1.2 | 6309.815 | 55 | 14 |
| **20** | 1.3 | 11983.17 | 46.5 | 8.7 |
| **21** | 1.3 | 12528.57 | 47 | 8 |
| **22** | 1.5 | 13807.37 | 42 | 16.5 |
| **23** | 1.5 | 6438.878 | 38 | 12 |
| **24** | 1.5 | 9606.453 | 41 | 10.5 |
| **25** | 1.5 | 8031.272 | 48.5 | 8.9 |
| **26** | 2 | 11465.24 | 39 | 12.6 |
| **27** | 2.1 | 7564.077 | 45.5 | 13.3 |
| **28** | 2.2 | 8185.131 | 52.5 | 6.1 |

**Table 2. Sperm Head Morphology Fourier Harmonic Measurements for Individual Bulls.**

| Bull | HA0 | HA1 | HA2 | HA3 | HA4 | HA5 |
|---|---|---|---|---|---|---|
| **1** | 3.067075 | 0.1196819 | 1.187841 | 0.09194465 | 0.2957738 | 0.09069261 |
| **2** | 3.118647 | 0.09430127 | 1.173004 | 0.09625085 | 0.2538151 | 0.0700948 |
| **3** | 3.087277 | 0.1175812 | 1.032603 | 0.1278038 | 0.2130631 | 0.08314797 |
| **4** | 3.072428 | 0.09753479 | 0.9832771 | 0.1177246 | 0.1815846 | 0.0669689 |
| **5** | 3.102513 | 0.08865031 | 0.9192595 | 0.113693 | 0.1282369 | 0.04416647 |
| **6** | 2.957772 | 0.1175632 | 1.075898 | 0.1203994 | 0.2338014 | 0.08099175 |
| **7** | 3.280553 | 0.1549128 | 1.232687 | 0.1491078 | 0.3107107 | 0.1087798 |
| **8** | 3.19624 | 0.09921172 | 1.102563 | 0.1136219 | 0.2066654 | 0.06916805 |
| **9** | 3.25312 | 0.1019572 | 1.051068 | 0.1159668 | 0.1896874 | 0.06847527 |
| **10** | 3.068156 | 0.09327969 | 1.079726 | 0.09733689 | 0.228799 | 0.06309004 |
| **11** | 3.066665 | 0.1514072 | 1.136857 | 0.1475045 | 0.2686795 | 0.1078571 |
| **12** | 3.098434 | 0.136584 | 1.099823 | 0.147715 | 0.2315209 | 0.09987295 |
| **13** | 3.209306 | 0.1015355 | 1.140226 | 0.1122499 | 0.2244274 | 0.06950404 |
| **14** | 3.23141 | 0.1080424 | 1.10632 | 0.1266937 | 0.216668 | 0.07381759 |
| **15** | 3.198411 | 0.09730248 | 1.115984 | 0.1052048 | 0.2122806 | 0.06275318 |
| **16** | 3.02259 | 0.1685147 | 1.223936 | 0.129008 | 0.3381399 | 0.1217816 |
| **17** | 3.322652 | 0.1088058 | 1.145906 | 0.1112953 | 0.2462428 | 0.09052296 |
| **18** | 3.127135 | 0.1310327 | 1.132599 | 0.1410027 | 0.2365529 | 0.09296565 |
| **19** | 3.127742 | 0.1146983 | 1.106599 | 0.1163031 | 0.2591915 | 0.09443299 |
| **20** | 3.068512 | 0.1122507 | 1.188197 | 0.1055449 | 0.2717835 | 0.08661904 |
| **21** | 3.230158 | 0.1039717 | 1.059529 | 0.1191768 | 0.1756475 | 0.05743645 |
| **22** | 3.064542 | 0.1011781 | 1.14048 | 0.1015061 | 0.2422783 | 0.07707373 |
| **23** | 3.096737 | 0.1008863 | 1.044269 | 0.1023432 | 0.1978347 | 0.06918425 |
| **24** | 3.214521 | 0.09727702 | 1.058098 | 0.1048204 | 0.1969743 | 0.0712937 |
| **25** | 3.253222 | 0.09543855 | 1.105198 | 0.1083746 | 0.1943067 | 0.05374382 |
| **26** | 3.018042 | 0.1025259 | 1.112031 | 0.1074448 | 0.2319929 | 0.07037682 |
| **27** | 3.221139 | 0.1255516 | 1.130806 | 0.1401007 | 0.2313195 | 0.08714826 |
| **28** | 2.986035 | 0.09526549 | 0.9917883 | 0.1170227 | 0.1850216 | 0.06695067 |

**Table 3. Pearson correlation values for all semen parameters tested as well as CR score**

| | CR | RNA/10 mill | % Motility | % DFI | HA0 | HA1 | HA2 | HA3 | HA4 | HA5 |
|---|---|---|---|---|---|---|---|---|---|---|
| **CR** | | 0.31 | 0.39 | -0.09 | 0.16 | -0.08 | 0.00 | 0.07 | -0.14 | -0.07 |
| **RNA/10 mill** | 0.31 | | 0.25 | -0.07 | 0.32 | -0.14 | 0.13 | 0.02 | -0.07 | -0.09 |
| **Motility** | 0.39 | 0.25 | | 0.08 | 0.08 | -0.27 | -0.40 | 0.22 | -0.45 | -0.24 |
| **DFI** | -0.09 | -0.07 | 0.08 | | 0.08 | -0.32 | -0.42 | -0.13 | -0.40 | -0.38 |
| **HA0** | 0.16 | 0.32 | 0.08 | 0.08 | | -0.10 | 0.16 | 0.11 | -0.14 | -0.13 |
| **HA1** | -0.08 | -0.14 | -0.27 | -0.32 | -0.10 | | 0.59 | 0.74 | 0.75 | 0.92 |
| **HA2** | 0.00 | 0.13 | -0.40 | -0.42 | 0.16 | 0.59 | | 0.12 | 0.90 | 0.67 |
| **HA3** | 0.07 | 0.02 | 0.22 | -0.13 | 0.11 | 0.74 | 0.12 | | 0.20 | 0.58 |
| **HA4** | -0.14 | -0.07 | -0.45 | -0.40 | -0.14 | 0.75 | 0.90 | 0.20 | | 0.86 |
| **HA5** | -0.07 | -0.09 | -0.24 | -0.38 | -0.13 | 0.92 | 0.67 | 0.58 | 0.86 | |

## APPENDIX 2

**Bovine Spermatozoal transcript profile from RNA isolated by RNeasy column-based method**

**Card CJ and Sartini BL**

**Abstract**

Sperm not only deliver the paternal half of the genome to the oocyte but also a population of various RNA types. These RNAs may be crucial for fertilization and embryonic development but also offer the potential as a diagnostic tool for fertility. To date the bovine spermatozoal transcript profile remains incomplete due to the reliance on hybridization-based technology such as microarrays. The goal of this study was to utilize Illumina RNA-Seq to sequence the bovine spermatozoal RNA profile. RNA was isolated via an RNeasy column based extraction (Method #1) from a pool of bulls with known fertility ratings of Conception Rate (CR) -2.9 to 3.5. Spermatozoal RNA isolated with this method had 260/280 ratios > 1.80, lacked contamination from somatic cells, and bioanalyzer profiles were consistent with previously published sperm RNA profiles. A total of 8,311 transcripts were found in the bovine spermatozoal RNA population when reads were aligned to the bovine genome (UMD 3.1/bosTau6). Some of the most abundantly expressed transcripts include *RPS3*, *YWHAE*, and *H2AFZ*. Gene ontology analysis revealed 762 Biological Processes (BP) categories represented by at least 6 transcripts and as many as 669 transcripts. The top three BP categories represented were regulation of transcription, regulation of RNA metabolic processes, and regulation of transcription (DNA-dependent). We were unable to validate the presence of a number of transcripts therefore the validity of this sequenced spermatozoal transcript profile is questionable. A second method for spermatozoal RNA isolation was developed and the

sequenced spermatozoal transcript profile is reported in Chapter 2 (Card and Anderson et al., 2013).

**Materials and Methods**

*Semen samples*

Cryopreserved semen from twenty-eight bull studs was obtained from Genex Cooperative Inc. (Shanawo, WI). The bulls had CR fertility scores that ranged from -2.9 to 3.5.

*Semen thawing*

Frozen bovine semen straws (n = 2 for each bull) were thawed in a 37°C water bath for one minute.

*Sperm-TALP wash*

Semen was washed with 4 ml sperm-TALP media (100 mM NaCl, 3.1 mM KCl, 25 mM $NaHCO_3$, 0.3 mM $NaH_2PO_4$, 2 mM $CaCl_2 2H_2O$, 0.4 mM $MgCl_2 6H_2O$, 21.6 mM sodium lactate, 1 mM pyruvate, 6 mg/ml BSA, pH 7.3-7.4; Parrish et al, 1998) and centrifuged for 10 minutes at 800 *x g* to remove cryoprotective media. This sperm-TALP wash is repeated a second time, the supernatant is removed and the sperm pellet is subsequently used for RNA isolation.

*Sperm RNA isolation (Method #1)*

The sperm pellets were re-suspended into 700 ul of buffer (10 ul 2-Mercaptoethanol + 1000 ul buffer RLT) for cell disruption (Qiagen; Valencia, CA). The samples were

subsequently homogenized using a 26-gauge needle (3x) and extruded into RNase-free microfuge tubes. Subsequent RNA isolation was conducted with the RNeasy kit (Qiagen; Valencia, CA) per manufacturer's directions. Modifications to the standard RNeasy protocol included warming the elution water to 37° C and extending the final incubation of the sample on the column for 8 minutes per elution. RNA concentrations were measured using the NanoDrop UV/Vis Spectrometer (Thermo Scientific; Waltham, MA). RNA preparations with A260/A280 ratios over 1.80 (indicating pure RNA) were stored in the -80°C freezer until needed for further experiments.

*Removal of genomic DNA*

RNA samples were DNAse treated using Turbo DNA-free Kit according to manufacturer's protocol (Ambion/Life Technologies; Carlsbad CA).

*mRNA amplification*

RNA samples were mRNA linearly amplified using the SMART mRNA amplification kit per manufacturer's protocol (Clontech; Mountain View, CA).

*Polymerase chain reaction (PCR)*

PCR conditions were 94° for 5 min, followed by 35 cycles of 94° for 30 sec, 60° for 30 sec, 72° for 30 seconds and a final extension at 72° for 10 min with transcript specific

primers. Negative controls included a no enzyme control as well as a no template control.

*RNA-Seq*

RNA-Seq analysis was performed on amplified mRNA from a pool of 10 bulls with fertility scores ranging from -2.9 to 3.5. This pooled sample was run on the Illumina HiSeq 2000 to generate paired-end 100 bp reads (Tufts University; Boston, MA). All read processing, alignment, and transcript expression analysis was performed using Galaxy (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010). Reads were aligned to the bovine genome (UMD 3.1/bosTau6) was performed using Tophat (Langmead et al., 2009; Trapnell et al., 2012). 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 their respective expression levels, denoted as Fragments Per Kilobase of exon per Million fragments mapped (FPKM) (Giardine et al., 2005).

*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).

## Results

*RNA quality*

Bull spermatozoal RNA isolated with the column-based method was devoid of contamination from germ cells and leukocytes as demonstrated by the lack of amplification for *C-KIT* and *CD45* respectively (Figure 1). Bioanalyzer analysis reveals a lack of 18S and 28S peaks compared to the control testis as expected (Figure 2). The amplified mRNA sperm sample showed a broad peak for its RNA profile with a greater presence of smaller RNAs (Figure 2). Nanodrop analysis results showed 260/280 ratios that ranged from 1.83 to 1.87.

*Illumina Sequencing*

Using the Illumina HiSeq 2000, a total of 126,797,470 x 2 paired-end reads were generated that were all 100 bp in length. When mapping the reads to the UMD 3.1/bosTau6 version of the bovine genome, 17.75% of the reads mapped to the bovine genome and 37% of these were uniquely mapped to a single transcript. Mapping statistics show that almost all fragments assigned to genes fell within coding exons and there was slightly more saturation at the 3' end over the 5' end with 107.43 reads/kb versus 80.74 reads/kb respectively. Very few reads mapped to intronic regions at just 0.82 reads/kb.

*Cryopreserved bovine spermatozoal transcript profile*

A total of 8,311 transcripts were identified as expressed, FPKM>0, in the pool of 10 bulls. This population includes nuclear-encoded transcripts as well as 4

mitochondrial-encoded transcripts. The top 10 transcripts, excluding transcripts coding for ribosomal proteins, based on FPKM expression levels are listed in Table 1. A number of transcripts from this profile were unable to be validated through PCR as their presence was not found in similarly isolated samples, including: *ACTINB*, *EIF1*, *EIF4A*, *GAPDH*, and *DBY*.

Despite the inability to validate the spermatozoal RNA isolated with the column based method, there are several similarities between the transcript profile of RNA isolated with Method 1 and Method 2 (TRIzol isolation) reported in Chapter 2 of this dissertation. This profile contained a total of 8,311 transcripts while the RNA isolation Method 2 contained 6,166 transcripts. When comparing the two populations a total of 4,126 transcripts were common between the two populations. Also, 2,040 transcripts were found unique to the Method 2 population while 4,206 were unique to this population (Figure 3).

*Gene ontology analysis*

When examining all expressed transcripts (FPKM>0) a total of 762 BP categories were present with each being represented by a range of 6 to 669 transcripts (Table 2). A total of 213 CC categories were found to be relevant with 7 to 786 transcripts representing each category. The final category, MF, had 191 represented categories with each including anywhere from 6 to 1226 transcripts. The top three BP represented by the dataset include regulation of transcription, regulation of RNA metabolic processes, and regulation of transcription (DNA-dependent).

Gene ontology analysis was also performed on the population of spermatozoal transcripts common between the profiles sequenced with the two RNA isolation methods. The top 3 BP categories in this common population are regulation of transcription, protein localization, and transcription (Table 3). For the transcript population unique to Method 1, the top 3 BP categories are regulation of transcription, regulation of RNA metabolic process, and regulation of transcription (DNA-dependent) while the unique to RNA isolation Method 2 had a top three BP category list of proteolysis, oxidation reduction, and translation.

**Discussion**

High-throughput sequencing of spermatozoal RNA isolated with a column-based procedure (Method #1) identified 8,311 transcripts. The isolated bovine spermatozoal RNA isolated with this method was devoid of somatic cell and testicular germ cell contamination, consistent with porcine spermatozoal RNA isolated with the same method (Yang et al, 2009). Bioanalyzer analysis of this isolated spermatozoal RNA showed a population of small RNAs that was comparable to previous literature showing a low concentration of RNA with a broad peak that is at its highest where smaller sized RNAs would be present (Das et al, 2010). This RNA profile differed from the profile of spermatozoal RNA isolated with a phenol-based method (Method #2) as it includes larger RNAs than those profiles (Gilbert et al, 2007; Card and Anderson et al, 2013). Isolation Method #1 also yielded higher RNA amounts per spermatozoa (600-1000 fg per spermatozoa) than was reported in other studies using a TRIzol RNA isolation (Method #2) method (5-400 fg RNA per spermatozoa; Boerke et al, 2007; Das et al, 2010;

Hamatani 2012). The increase RNA yield may represent the true amount of RNA in sperm or could include contamination from other sources that were not tested for. Though high quality RNA can be isolated through a variety of methods, the isolation method should be chosen to yield the desired subset of RNA. For example, the total spermatozoal RNA population contains mostly smaller RNAs, in which a phenol-based isolation method such as Method #2 would an optimal choice to efficiently isolate this segment of RNA populations as phenol based isolation methods generally yield more smaller sized RNAs than column based methods (Eldh et al, 2012).

Despite the high quality of the RNA isolated and the number of transcripts sequenced, we were unable to validate the sperm transcript profile by PCR amplification of individual transcripts in an aliquot of RNA from sequenced population or in RNA from other individuals that were isolated in the same manner. Several spermatozoal transcripts were identified from previous publications but were not identified in this transcript profile. Anticipated transcripts that were not present in the sperm transcript profile, including *TNP1*, *TNP2*, *PGK2*, *AKAP4*, *RPS4Y*, *PRM2*, *PLCZ1*, *CSN2*, and *CRISP2* (Swann et al, 2006; Gilbert et al, 2007; Feugang et al, 2010; Yao et al, 2010; Arangasamy et al, 2007; Ganguly et al, 2012). Only the *PRM1* and *CLU* transcripts were amplified, although inconsistently, in a similar sperm RNA population isolated via isolation Method #1. *PRM1* and *TNP1* were not identified in the sequenced sperm transcript profile. *PRM1* is one of the most studied sperm RNAs and has been reported in bovine spermatozoa in numerous studies, as well as *TNP1* (Gilbert et al, 2007; Feugang et al, 2010; Hecht et al, 2011). The absence of these transcripts in the spermatozoal RNA

population lead us to question the validity of the isolation Method #1, and the biological representation of the resulting spermatozoal transcript profile.

Due to the inability to validate the spermatozoal transcript profile sequenced from sperm RNA isolated with Method #1, a TRIzol RNA isolation method (Method #2) for bovine spermatozoal was developed and validated from a horse spermatozoal RNA protocol (Card and Anderson et al, 2013 and Das et al., 2010). The RNA from the TRIzol isolation (Method #2) had lower A260/A280 ratios (around 1.60) but was more consistent with previously published data. Though the ratios were lower in RNA isolated with this method, the resulting sperm RNA population was treated with a DNAse and cDNA amplified then sequenced and successfully validated (Card and Anderson et al., 2013).

There were many similarities between the sequences transcript profile for Method #1 with the transcript profile from Method #2. The percent of reads aligned to the genome was similar (at 16% from Method #1 versus the 14% in Method #2). Also, 4,126 transcripts are common between the two populations. There were also just over 2,000 transcripts unique to the isolation Method #2 population with just over 4,000 unique to this population from Method #1. A total of 8,311 transcripts were expressed in population from isolation Method #1, which is comparable to the 6,166 from isolation Method #2 (Card and Anderson et al, 2013). These differences must be due to not only sample preparation differences but also due to some different individual bulls used for the sequenced population.

Another reason for the difference in transcripts expressed is due to the data processing method. The number of transcripts expressed and similar/unique to

populations was based on accession numbers. A number of transcripts are represented by multiple accession numbers, in which there is potential that one transcript is represented by one accession number in a dataset, and then the other dataset contains that accession number but an additional accession number for the same transcript. This would result in the same transcript being represented by a common transcript and a unique transcript to one population. A number of transcripts in both populations were found expressed at very low levels, with only a few reads hitting on a transcript within the genome. These transcripts are likely insignificant to the datasets but were included in analysis. Although the fertility ranges were the same, individual variance has been seen in many species when comparing transcript profiles and is likely a large contributor to difference in the two transcript profiles (Ostermeier et al, 2002).

Some of the common transcripts that have been found between the RNA-Seq populations isolated using Method #1 and Method #2 include many found in previous literature such as *CHMP5*, *CLGN*, *H2AFZ*, *CLU*, *MYCBP*, *PEBP1*, *CCT8*, *DDX3Y*, *PPIH*, *STRBP*, *and FLOT1*. Two transcripts expressed in this population, *PEBPI and CCT8*, have already shown to be correlated with fertility in bulls (Arangasamy et al, 2011). *PEBP1* is known to have function associated with capacitation as it causes alterations in the plasma membrane located both in the tail and head of sperm. *CCT8* has been localized in centrioles and microtubules but appears to get discarded when the residual body is shed. This transcript has been negatively correlated with fertility; meaning that its presence indicates that residual bodies were not shed and the sperm is potentially immature (Arangasamy et al, 2011).

Some transcripts that were found in both populations had very different expression levels. For example, *H2AFZ* is expressed at 6,200 FPKM (one of the highest transcripts; Method #1) when the Method #2 population had an expression level of 166 FPKM. Another example is that of *CHMP5* in which this population, Method #1, had a low FPKM at 32 while it was one of the highest expression levels in the Method #2 population at 2,778. As the fertility range of the bulls was the same between populations, with many of the same bulls being used, this would be most likely attributed to sample preparation and sequencing differences.

Amongst the common transcript population gene ontology analysis showed a prevalence of transcripts involved in transcription, metabolic processes, and translation. Many of these transcripts are probably remnants of spermatogenesis and could be useful diagnostic tools for fertility. Some of the transcription and translation associated genes could also be interesting candidates for full-length examination as they could be crucial post-fertilization before the embryonic genome takes over, which occurs around the 8-cell stage in the cow (Boerke et al, 2007).

Some differences that remain concerns for this data set were also apparent in post-sequencing analysis. There were only 3 mitochondrial-encoded transcripts present in this population while 33 mitochondrial-encoded transcripts were identified in the spermatozoal transcript from RNA isolation Method #2. This number is lower than expected, as there is a high level of mitochondrial activity in the mid-piece of spermatozoa that provides energy for sperm motility. Only 37% of the aligned sequence was uniquely aligned to the genome, a much lower percentage than that of data from isolation Method #2, which was up at 80%.

The spermatozoal RNA profile obtained by RNA-Seq of spermatozoal RNA isolated with the column-based method (Method #1) resulted in a transcript profile that could not be validated. A TRIzol spermatozoal RNA isolation method was developed (Method #2) for bovine and that transcript profile was validated (Card and Anderson et al., 2013, Chapter 2 of this dissertation). Although many similarities are noted in the two transcript profiles from the different isolation method, differences apparent in the column-based profile (presented in this appendix) may demonstrate contamination of the RNA. These experiments highlight the variation that can occur with different RNA isolation methods.

**References**


Arangasamy A, Kasimanickam VR, DeJarnette JM, Kasimanickam RK (2011): Association of *CRISP2*, *CCT8*, *PEBP1* mRNA abundance in sperm and sire conception rate in Holstein bulls. Theriogenology, 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. Current Protocols in Molecular Biology, Chapter 19: Unit 19.10.1-21.

Boerke A, Dieleman SJ, Gadella BM (2007): A possible role for sperm RNA in early embryo development. Theriogenology, 68S: 147-155.

Card CJ, Baird EJ, Zamberlan S, Krieger KE, Kaproth M, Sartini BL (2013): Cryopreserved bovine spermatozoal transcript profile as revealed by RNA-Seq. Biology of Reproduction, 88(2): 1-9.

Das PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, Varner DD, Chowdhary BP, Raudsepp T (2010): Total RNA isolation from stallion sperm and testis biopsies.Theriogenology, 74: 1099-1106

Eldh M, Lotvall J, Malmhall C, Ekstrom K (2012): Importance of RNA isolation methods for analysis of exosomal RNA: Evaluation of different methods. Molecular Immunology, 50: 278-286.

Feugang JM, Rodriguez-Osorio N, Kaya A, Wang H, Page G, Ostermeier GC, Topper EK, Memili E (2010): Transcriptome analysis of bull spermatozoa: implications for male fertility. Reproductive BioMedicine, 21: 312-324.

Ganguly I, Gaur GK, Kumar S, Mandal DK, Kumar M, Singh U, Kumar S, Sharma A (2012): Differential expression of protamine 1 and 2 genes in mature spermatozoa of normal and motility impaired semen producing crossbred Frieswal (HF x Sahiwal) bulls. Research in Veterinary Science, 94(2): 256-262.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005): Galaxy: a platform for interactive large-scale genome analysis. Genome Research, 15(10):1451-5.

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 Biology, 2010 Aug 25; 11(8):R86.

Hamatani, T (2012): Human spermatozoal RNAs. Fertility and Sterility, 97(2): 275-281.

Hecht N, Cavalcanti MC, Nayudu P, Behr R, Reichenback M, Weidner W, Steger K (2011): Protamine-1 represents a sperm specific gene transcript: a study in Callithrix jacchus and Bos taurus. Andrologia, 43(3):167-173.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009): Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10(3): R25.

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA (2002): Spermatozoal RNA profiles of normal fertile men. The Lancet, 360: 772-777.

Swann K, Saunders CM, Rogers NT, Lai FA (2006): PLCz(zeta): A sperm protein that triggers $Ca^{2+}$ oscillations and egg activation in mammals. Seminars in Cell & Developmental Biology, 17: 264-273.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012): Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols, 7(3):562-578.

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

Yang CC, Lin YS, Hsu CC, Wu SC, Lin EC, Cheng WTK (2009): Identification and sequencing of remnant messenger RNAs found in domestic swine (*Sus scrofa*) fresh ejaculated spermatozoa. Animal Reproduction Science, 113: 143-155.

Yao C, Wang Z, Zhou Y, Xu W, Li Q, Ma D, Wang L, Qiao Z (2010): A study of Y chromosome gene mRNA in human ejaculated spermatozoa. Molecular Reproduction and Development, 77: 158-166.

Figure 1. Spermatozoal RNA isolated with a column based method (Method 1) lacks contamination from germ cells (C-KIT) and leukocytes (CD45). Three different sperm isolations were tested (Sp1-3). Testis (T) positive control and negative controls (-RT) containing no enzyme and no template respectively.

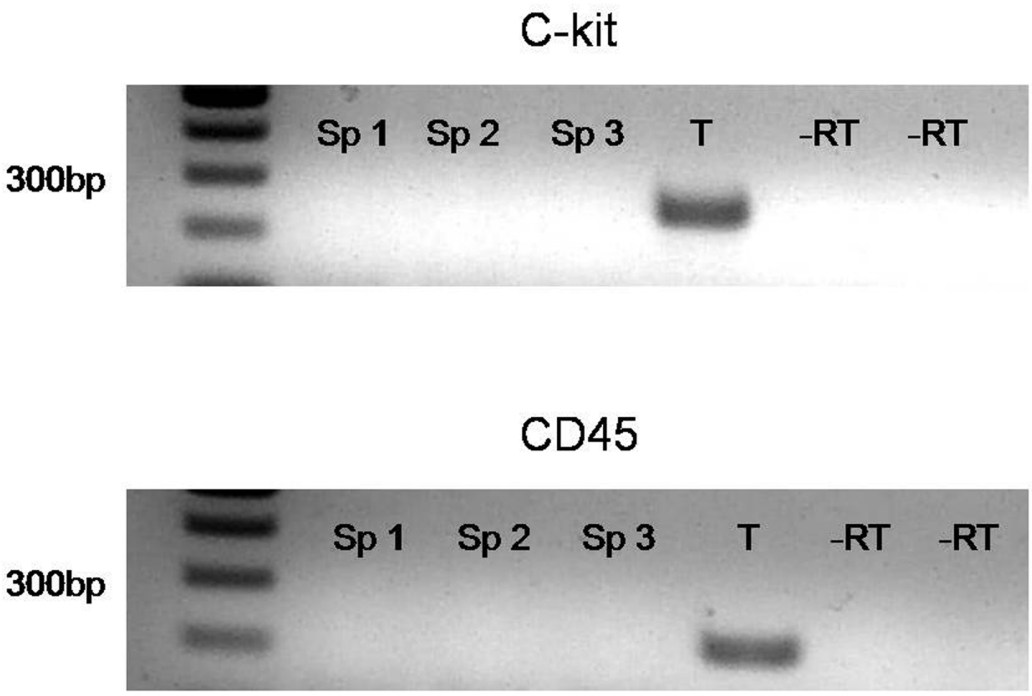


Figure 2. Bioanalyzer electropherograms for testis and amplified sperm mRNA (Method 1 isolation) samples. M = Marker; 18S = 18S ribosomal RNA peak; 28S = 28S ribosomal RNA peak.

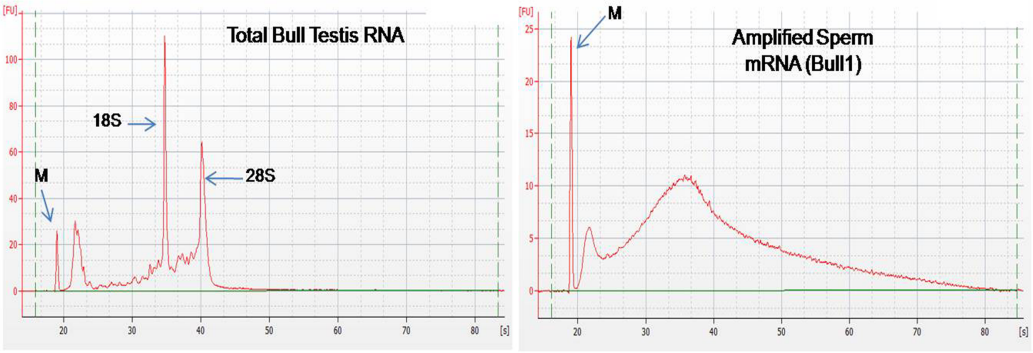


Figure 3. Common and unique spermatozoal transcripts between data from RNA isolated with the column based method (A) and the TRIzol method (B: Card and Anderson et al., 2013, Chapter 2 of this dissertation).

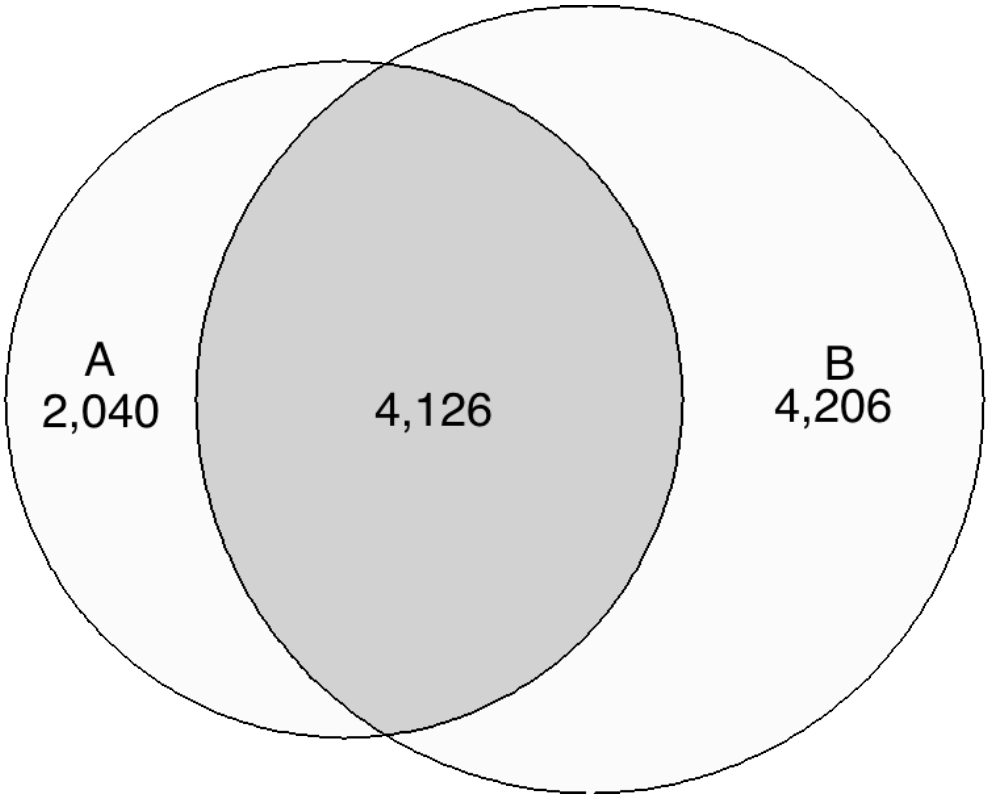


**Table 1. Top 10 bovine spermatozoal transcripts based on FPKM**

| Gene Symbol | Gene Name | Accession no. | FPKM |
|---|---|---|---|
| *HRAS* | similar to GTPase Hras | AJ437020 | 43728.7 |
| *YWHAE* | tyrosine 3-monooxygenase activation protein | BC102928 | 6290.88 |
| *H2AFZ* | H2A histone family, member Z | BC109743 | 6200.32 |
| *LOC782052* | basic transcription factor 3 | AB098942 | 4144.2 |
| *PTGES3* | prostaglandin E synthase 3 (cytosolic) | AY692440 | 4050.95 |
| *PCBP2* | poly(rC) binding protein 2 | BC103397 | 3681.9 |
| *HMGN2* | high-mobility group nucleosomal binding domain 2 | BC142241 | 3447.87 |
| *HNRNPC* | heterogenous nuclear ribonucleoprotein C (C1/C2) | BC104494 | 3378.44 |
| *LOC789867* | eukaryotic translation elongation factor 1 alpha 1 | AF013213 | 3348.21 |
| *PRDX1* | peroxiredoxin 1 | AB098940 | 3235.42 |

**Table 2. Top 10 gene ontology categories**

| | Transcripts | % of Total Transcripts |
|---|---|---|
| **Biological Process** | | |
| Reg. of Transcription | 669 | 9.06 |
| Reg. of RNA Metabolic Process | 451 | 6.11 |
| Reg. of Transcription, DNA-dependent | 439 | 5.94 |
| Transcription | 398 | 5.39 |
| Phosophorous Metabolic Process | 372 | 5.04 |
| Phosphate Metabolic Process | 372 | 5.04 |
| Protein Localization | 351 | 4.75 |
| Est. of Protein Localization | 323 | 4.37 |
| Intracellular Signaling Cascade | 323 | 4.37 |
| Protein Transport | 322 | 4.36 |
| | | |
| **Cellular Component** | | |
| Non-membrane-bounded Organelle | 786 | 10.64 |
| Intracellular Non-membrane-bounded Organelle | 786 | 10.64 |
| Mitochondrion | 557 | 7.54 |
| Membrane-enclosed Lumen | 518 | 7.01 |
| Intracellular Organelle Lumen | 499 | 6.76 |
| Organelle Lumen | 499 | 6.76 |
| Organelle Membrane | 420 | 5.69 |
| Nuclear Lumen | 371 | 5.02 |
| Cytoskeleton | 369 | 4.99 |
| Endoplasmic Reticulum | 333 | 4.51 |
| | | |
| **Molecular Function** | | |
| Ion Binding | 1226 | 16.59 |
| Cation Binding | 1214 | 16.44 |
| Metal Ion Binding | 1202 | 16.27 |
| Nucleotide Binding | 973 | 13.17 |
| Transition Metal Ion Binding | 864 | 11.69 |
| Purine Nucleotide Binding | 805 | 10.89 |
| Purine Ribonucleotide Binding | 775 | 10.49 |
| Ribonucleotide Binding | 775 | 10.49 |
| Zinc Ion Binding | 681 | 9.22 |
| Nucleoside Binding | 624 | 8.45 |

**Table 3. Top 10 gene ontology categories for common transcripts**

| | Transcripts | % of Total Transcripts |
|---|---|---|
| **Biological Process** | | |
| regulation of transcription | 319 | 8.21 |
| protein localization | 203 | 5.22 |
| transcription | 200 | 5.15 |
| phosphorus metabolic process | 190 | 4.89 |
| phosphate metabolic process | 190 | 4.89 |
| protein transport | 188 | 4.83 |
| establishment of protein localization | 188 | 4.84 |
| proteolysis | 185 | 4.76 |
| macromolecule catabolic process | 168 | 4.32 |
| cellular macromolecule catabolic process | 160 | 4.12 |
| **Cellular Component** | | |
| non-membrane-bounded organelle | 493 | 12.68 |
| intracellular non-membrane-bounded organelle | 493 | 12.68 |
| mitochondrion | 331 | 8.52 |
| membrane-enclosed lumen | 322 | 8.28 |
| intracellular organelle lumen | 311 | 8.01 |
| organelle lumen | 311 | 8.01 |
| nuclear lumen | 236 | 6.07 |
| organelle membrane | 231 | 5.94 |
| cytoskeleton | 221 | 5.69 |
| ribonucleoprotein complex | 208 | 5.35 |
| **Molecular Function** | | |
| metal ion binding | 615 | 15.82 |
| nucleotide binding | 545 | 14.02 |
| transition metal ion binding | 442 | 11.37 |
| purine nucleotide binding | 434 | 11.17 |
| purine ribonucleotide binding | 421 | 10.83 |
| ribonucleotide binding | 421 | 10.83 |
| nucleoside binding | 338 | 8.69 |
| purine nucleoside binding | 336 | 8.64 |
| adenyl nucleotide binding | 333 | 8.57 |
| adenyl ribonucleotide binding | 320 | 8.23 |

## APPENDIX 3: PROTOCOLS

### Percoll Separation and TRIzol Sperm Isolation

Sperm-TALP media (Parrish et al., 1998)

| | Final concentration | FW | g per 100 mls | |
|---|---|---|---|---|
| NaCl | 100 mM | 58.4 | 0.584 | |
| KCl | 3.1 mM | 74.55 | 0.023 | |
| $NaHCO_3$ | 25 mM | 84.01 | 0.21 | |
| $NaH_2PO_4$ | 0.3 mM | 120 | 0.0036 | |
| HEPES | 10 mM | 238.3 | 0.238 | |
| *Add $CaCl_2$ and $MgCl_2$ after all other powders are in solution | | | | |
| $CaCl_2 \cdot 2H_20$ | 2 mM | 111 | 0.022 | |
| $MgCl_2$.6H20 | 0.4 mM | 95.21 | 0.003 | |
| Lactate (sodium salt) | 21.6 mM | 112 | 0.308 ml | (fridge) |

Combine all chemical powders and solutions in 100 mls of $dH_2O$.
Sterile filter and store at 4°C in glass bottle.
Good for 1 month
On day of isolation: Supplement with pyruvate (1 mM). Be sure to adjust pH (7.3-7.4).
Then add BSA.

10X Tyrode's:
For 100 ml, mix the following in in 80 mls of $dH_20$:

| | Final concentration | FW | 100 mls |
|---|---|---|---|
| KCl | 31 mM | 74.55 | 0.230 g |
| NaCl | 800 mM | 58.4 | 4.675 g |
| $NaH_2PO_4$ | 3 mM | 142 | 0.04 g |
| HEPES | 100 mM | 238.3 | 2.09 g |

Bring volume to 100 mls with $dH_20$
pH to 7.3
Sterile filter and store in glass bottle in refrigerator indefinitely
Re-adjust pH as needed

| 90% isotonic Percoll | Conc. | FW | | |
|---|---|---|---|---|
| Percoll | | | 9 mls | |
| 10X Tyrode's | | | 1 ml | |
| $NaHCO_3$ | 25 mM | 84 | 0.021 g | |
| Lactic acid (sodium salt) | 21.6 mM | 112 | 37 ul | (fridge) |
| $CaCl_2 \cdot 2H_20$ | 2 mM | 111 | 20 ul | |

Stock = Add 0.735 g $CaCl_2 \cdot 2H_20$ to 5 ml $dH_20$
Sterile filter, aliquot and freeze at -20°C

| | | | |
|---|---|---|---|
| $MgCl_2$ | 0.4 mM | 95.21 | 39 ul |

Stock = Add 0.02 g $MgCl_2 \cdot 6H_20$ to 10 ml of $dH_20$

Sterile filter, aliquot and freeze at -20°C
Mix in a sterile 15 ml conical tube.
Add ingredients in order listed.
Cap and vortex to mix. Lactic acid may sink to bottom of tube. Be sure it is mixed in solution.
DO NOT FILTER. Store at 4°C for up to 2 weeks.

Percoll Washing Sperm Protocol for RNA and protein isolation from fresh sperm
This protocol will separate a population of motile viable sperm from dead sperm and extender and/or seminal fluid.

Need:
38.5°C water bath
Sperm-TALP with pyruvate and BSA
90% Percoll

Sperm-TALP Day of:
- Add pyruvate (0.0012g/10 mls sperm-TALP) (fridge)
- Check pH (7.3-7.4) and adjust if needed with acid or base. Add 10 to 20 uL at a time.
- Add BSA (0.06 g/10 ml sperm-TALP) (fridge)
- Sterile filter
- Warm sperm-TALP in 38.5°C water bath

Preparation for one Percoll gradient:
Add 100 ul of 90% Percoll to a 15 ml tube. Mark this fluid level on the tube with a sharpie.
Add another 200 ul and mark this level also.
Add 1.7 ml more Percoll to bring to a total volume of 2 mls.
In another 5 ml tube, dilute 1 ml 90% Percoll with 1 ml sperm-TALP (with pyruvate and BSA) to make 2 ml 45 % Percoll (*Always make fresh 45% Percoll day of separation).
Leave these tubes at room temperature until used.

Layering Percoll gradient:
Layer 2 mls of 45% Percoll onto the 2 mls 90% Percoll in the 15 ml tube by slowly pipetting using glass pipette with plastic bulb. Minimize splashing as much as possible. Make sure the two layers do not mix!
Cut 2 straws of sperm so contents flow into 1.5 mL centrifuge tube. If there is semen left in tip of straw, use pipette with tip at opposite end to push out into tube by pipetting up and down. Once straw contents are in centrifuge tube, pipette onto Percoll gradient.

Centrifugation and washing:
Centrifuge for 30 minutes at 340 x g
Remove supernatant from the live layer and discard.
Add 4 mL of warmed sperm-TALP to live and dead samples in separate conical tubes.
Centrifuge at 800 x g for 10 minutes.

Microscope Work:
Remove supernatant down to 300 uL mark on conical tube.
Reconstitute pellet in 300 uL of solution.
If doing more than one Percoll gradient then pool all of the 300ul solutions together in one 15ml conical tube and reconstitute
Set 5 uL aside on slide (warmed to 30$^{o}$C) for motility.
Remove another 3 uL of sample and put into 1.5 mL centrifuge tube with 27 uL water.
Mix well and put 10 uL of sample on each side of hemocytometer.

MICROSCOPE:

Hemocytometer: Count 5 squares (each square consists of 15 small squares), diagonally. Record numbers and save for later calculations.
Motility: Starting at 50%, determine if motility is over or under. From that, determine if motility is over or under half-way point of remaining percentile. Continue until accurate motility can be determined at a multiple of 5.

Use average hemocytometer count to dilute sample to 30,000,000 sperm cells/ul
See **Capacitation Protocol**

TRIzol Isolation
1. Add 1 mL TRIzol (or 750 uL TRIzol LS) reagent to sperm pellet + 3 µL Glycogen AFTER sperm has been washed twice for 10 minutes at 800 x g in sperm-TALP.
3. Lyse sample with 26 ga, 6cc needle 20 times and incubate for 30 minutes at room temp.
4. Add 200 µL chloroform per 1 mL TRIzol (or 750 uL TRIzol LS) reagent to sample. Shake for 20 seconds then let sit at room temperature for 10 minutes.
5. Centrifuge at 12,000x g (11,400 RPM) for 15 minutes at 4$^{o}$ C.
6. Remove clear aqueous layer at top (contains RNA) and put in new tube. Use a smaller sized pipette when closer to the division between the pink and clear layer.
7. Add 500 µL ice cold isoproanol and let sit for 10 minutes on ice.
Keep on ice for the remainder of isolation protocol.
8. Centrifuge at 12,000x g for 10 minutes at 4$^{o}$ C.
9. Remove and discard supernatant and add 1 mL 75% ethanol to pellet.
Vortex briefly, then centrifuge at 12,000x g for 5 minutes at 4$^{o}$ C.
10. Remove supernatant and air dry pellet on ice for 5 - 10 minutes.
After 5 minutes on ice, remove any accumulated supernatant again.
11. Add 60 µL of Nuclease-Free water to pellet.
Vortex until RNA pellet is dissolved in solution.
12. Nanodrop and store sample at -80$^{o}$ C.

## Sperm Wash and RNA Isolation (RNeasy Kit)

### Sperm Wash

1. Thaw frozen semen straws in water bath

2. Add 100 ml sperm-TALP to thin conical tube and mark line at 100ul mark and then add semen sample to tube

3. Add 4 ml of prepared/warmed sperm-TALP to sample and centrifuge for 10 minutes at 600 x g(RCF) in swinging bucket centrifuge

4. Remove supernatant from sperm pellet, add 4 ml of sperm-TALP and centrifuge again for 10 minutes at 600 x g(RCF)

5. Remove supernatant down to 100 ul mark and reconstitute pellet into remaining sperm-TALP

6. Put 5 ul of sample onto warmed slide and use microscope to determine motility

7. Dilute 3 ul of sample into 27 ul water and use this mixture on hemocytometer to determine sperm concentration (see hemocytometer protocol)

### RNA Isolation

-All steps should be done at room temperature
-Use RNA microcentrifuge and keep temperature between 20-25°C, making sure it does not cool below 20°C.

Before Starting:
- β-Mercaptoethanol (β-ME) must be added to the Buffer RLT before use. Add 10ul β-ME per 1ml Buffer RLT. Dispense in a fume hood and set aside for use during isolation steps.

-Make sure Buffer RPE has had ethanol added via kit instructions in RNeasy Kit before use.

Procedure:
1. Add 700ul of buffer RLT(with β-ME) to 10 million sperm cells worth of sample from wash and let sit for 10 minutes.

2. Suck sample in and out of 26 ½ gauge needle 3x using weighing boat to hold sample. After 3rd run through the needle, dispense sample into 1.5ml centrifuge tube.

3. Centrifuge the lysate for 3 minutes at 13,000 rpm.

4. Add 700ul of 70% ethanol to the cleared lysate and mix immediately by pipetting. DO NOT CENTRIFUGE. Proceed immediately to step 5.

5. Transfer up to 700ul of the sample, including any precipitate that may have formed to an RNeasy spin column placed in a 2ml collection tube from the kit. Close the lid gently and centrifuge for 15 seconds at 13,000 rpm. Discard the flow-through from the collection tube. Reuse the collection tube and repeat this process until the entire sample has been run through the column. Discard the flow-through each time and use the same collection tube for the next step.

6. Add 700ul of Buffer RW1 to the RNeasy spin column. Close the lid gently and centrifuge for 15 seconds at 13,000 rpm to wash the spin column membrane. Discard the flow-through and reuse the collection tube in the next step.

7. Add 500ul of Buffer RPE to the RNeasy spin column. Close the lid gently and centrifuge for 15 seconds at 13,000 rpm to wash the spin column. Discard the flow-through and reuse the collection tube in the next step.

8. Add 500ul of Buffer RPE to the RNeasy spin column. Close the lid gently and centrifuge for 2 minutes at 13,000 rpm to wash the spin column. Discard the flow-through.

9. Place the RNeasy spin column into a new 2ml collection tube and discard the old collection tube. Centrifuge for 1 minute at 13,000 rpm.

11. Place the RNeasy spin column into a new 1.5ml microcentrifuge tube. Add 55ul of RNase-free water (heated to 37.5°C) directly onto the spin column membrane. Let the water sit on the column 8 minutes and then close the lid gently and centrifuge for 1 minute at 13,000 rpm to elute the RNA.

12. Take the 55ul of sample from the collection tube and reapply it to the RNeasy spin column. Let the sample once again sit for 8 minutes and then centrifuge for 1 minute at 13,000 rpm to elute the RNA.

13. Put the RNA sample on ice and bring to Nanodrop to determine the concentration of the RNA (see Nanodrop protocol).

**RNA Quality Gel**

To check the integrity of your RNA sample and for Northern analysis

**After RNA isolation:**
Retreive aqueous RNA sample from -80°C
OR
Do ethanol precipitation of RNA sample that was stored at -20°C
Check concentration on Nanodrop
Make aqueous solution sample in RNAase free water
To check RNA sample 5 ug in 5 ul $H_20$
For Northern analysis 15 ug in 5 ul $H_20$
***KEEP SAMPLE ON ICE AT ALL TIMES UNLESS NOTED******

**10X MOPS**
41.8 g MOPS
in 700 ml of DEPC-treated $H_20$
Adjust pH to 7.0 with 2 N NaOH
Add 20 ml of DEPC-treated 1 M sodium acetate
Add 20 ml of DEPC-treated 0.5 M EDTA (pH 8.0)
Adjust the volume to 1 liter with DEPC-treated $H_20$

**Gel preparation:**
Use only DEPC-treated solutions and baked glassware to analyze RNA
Bake glassware: 200°C for 4 hours
Clean gel box, gel bed and comb with RNase away and wipe dry
Position gel box with rubber gaskets at wall of gel box

For 50 ml gel:
Weigh out 0.5 g agarose (Fisher BP1356-100 Northerns) (for a 1% gel)
Place agarose in a RNase-free 125 ml flask
Add 44 ml DEPC $H_20$ and swirl to mix
Cover flask with double layer of plastic wrap
Heat in microwave for 1 minute on high power.
Open microwave and swirl flask ****Flask is hot** Tilt flask away from your face*****
Heat in microwave for another 30sec to 1 minute until agarose is completely melted (no bubbles or "little contact lenses")
Remove from microwave and cool solution until you can touch the bottom with your hand
After solution has cooled, add 5 ml of 10X MOPS, pH 7.0
Add 0.885 ml 12.3M formaldehyde (final concentration in gel is 0.22M)
Pour gel, pop any air bubbles with a RNase-free pipet tip, ADD COMB and let gel polymerize

Running buffer:
55 ml of 10 X MOPS, pH 7.0 + 9.8 ml formaldhyde (final 0.22 M) + 485.2 ml DEPC $H_20$

After gel has set, remove comb and turn gel
Pour tank buffer to just cover gel
Replace gel box cover until ready to use to prevent formaldehyde fumes from coming

Sample preparation: Add 5 ul RNA loading buffer (stored in -20°C freezer) to 5 ul aqueous RNA sample. Heat to 65C for 10 minutes then chill on ice for 5 minutes. Load into wells of gel and note sample loading in lab notebook.

Plug in red and black leads to power supply (make sure these are in the correct position!) and run get at 80 V for approximately 45 minutes or until the sample is 2/3 the length of the gel. Look for bubbles to make sure you have current.

TURN OFF THE POWER!!!! AND UNPLUG THE UNIT FROM THE POWER SUPPLY!!!!!
With gloves, remove the gel bed onto a piece of saran wrap. The sample lane contains Ethidium Bromide so avoid touching the gel. Analyze the gel with the UV box and check the ratio of the 28S to 18S ribosomal RNA. Save an image of your gel.

**SMARTer Pico ds-cDNA Synthesis (Amplification Kit)**

**First-strand cDNA synthesis**

1. For each sample and Control Mouse Liver Total RNA, combine the following reagents in separate 0.5 ml reaction tubes:

| | |
|---|---|
| 1–50 µl | RNA (1-1,000 ng of total RNA)* |
| 7 µl | 3' SMART CDS Primer II A (12 µM) |
| x µl | Deionized H2O |
| **57 µl** | **Total Volume** |

   *For the control synthesis, add 10 ng of Control Mouse Liver Total RNA. PCR-Select users should start with >10 ng of total RNA.
2. Mix contents and spin the tubes briefly in a microcentrifuge.
3. Incubate the tubes at 72°C in a hot-lid thermal cycler for 3 min, and then cool the tubes to 42°C.
   **NOTE:** The initial reaction steps (Step 4-6) are critical for first-strand synthesis and should not be delayed after Step 3. You can prepare your master mix (for Step 4) while your tubes are incubating (Step 3) in order to jump starts the cDNA synthesis.
4. Prepare a Master Mix for all reaction tubes at room temperature by combining the following reagents in the order shown:

| | |
|---|---|
| 20 µl | 5X First-Strand Buffer |
| 2 µl | DTT (100 mM) |
| 10 µl | dNTP Mix (10 mM) |
| 7 µl | SMARTer II A Oligonucleotide (12 µM) |
| 5 µl | RNase Inhibitor |
| 5 µl | SMARTScribe™ Reverse Transcriptase (100 U)* |
| **49 µl** | **Total Volume added per reaction** |

   * Add the reverse transcriptase to the master mix just prior to use. Mix well by vortexing and
   spin the tube briefly in a microcentrifuge.
5. Aliquot 49 µl of the Master Mix into each reaction tube. Mix the contents of the tubes by gently
   pipetting, and spin the tubes briefly to collect the contents at the bottom.
6. Incubate the tubes at 42°C for 1 hour.
   **NOTE:** If you plan to use a downstream application that requires long transcripts, *extend the incubation time to 90 min*.
7. Terminate the reaction by heating the tubes at 70°C for 10 min.
8. If necessary, cDNA samples can be stored at –20°C (for up to three months) until you are ready to proceed with spin-column purification.

**Column purification of cDNA using NucleoSpin gel and PCR clean-up**

1. Add 350 µl of Buffer NT to each cDNA synthesis reaction; mix well by pipetting.
2. Place a NucleoSpin Gel and PCR Clean-Up Column into a 2 ml collection tube. Pipette the sample into the column. Centrifuge at 8,000 rpm for 1 min. Discard the flowthrough.
3. Return the column to the collection tube. Add 600 µl of Wash Buffer NT3 to the column. Centrifuge at 14,000 rpm for 1 min. Discard the flowthrough.
4. Return the column to the collection tube. Add 250 µl of Wash Buffer NT3 to the column. Centrifuge at 14,000 rpm for 1 min. Discard the flowthrough.
5. Place the column back into the collection tube. Centrifuge at 14,000 rpm for 2 min to remove any residual Wash Buffer NT3.
6. Transfer the NucleoSpin Columns into a fresh 1.5 ml microcentrifuge tube. Add 50 µl of sterile Milli-Q H2O to the column. Allow the column to stand for 2 min with the caps open.
7. Close the tube and centrifuge at 14,000 rpm for 1 min to elute the sample.
8. Repeat elution with 35 µl of sterile Milli-Q H2O in the same 1.5 ml microcentrifuge tube. The recovered elution volume should be 80–85 µl per sample. If necessary, add sterile Milli-Q H2O to bring the total volume up to 80 µl.
9. For PCR-Select cDNA subtraction, proceed with the protocols provided in Appendix A of this User Manual. For all other applications, proceed with Section D. Samples can be stored at –20°C (for up to three months) until you are ready to proceed with cDNA amplification by LD PCR.

**cDNA amplification by LD PCR**

1. Preheat the PCR thermal cycler to 95°C.
2. For each reaction, aliquot the appropriate volume (see T able II) of each diluted first-strand cDNA into a labeled 0.5 ml reaction tube. If necessary, add deionized H2O to adjust the volume to 80 µl.

| Table II: Guidelines for Setting Up PCR Reactions | | | |
|---|---|---|---|
| Total RNA (ng) | Volume of Diluted ss cDNA for PCR (uL) | Volume of H20 (uL) | Typical Optimal No. of PCR Cycles* |
| 1000 | 2.5 | 77.5 | 18-20 |
| 250 | 10 | 70 | 18-20 |
| 100 | 25 | 55 | 18-20 |
| 50 | 40 | 40 | 18-20 |
| 20 | 80 | none | 19-21 |
| 5 | 80 | none | 21-23 |
| 1 | 80 | none | 24-27 |

3. Prepare a Master Mix for all reactions, plus one additional reaction. Combine the following reagents in the order shown:

| | |
|---|---|
| 4 µl | Deionized H2O |
| 10 µl | 10X Advantage 2 PCR Buffer |
| 2 µl | 50X dNTP Mix (10 mM) |
| 2 µl | 5' PCR Primer II A (12 µM) |
| 2 µl | 50X Advantage 2 Polymerase Mix |
| **20 µl** | **Total Volume per reaction** |

4. Mix well by vortexing and spin the tube briefly in a microcentrifuge.
5. Aliquot 20 µl of the PCR Master Mix into each tube from Step 2.
6. Cap the tube, and place it in the preheated thermal cycler. If you are NOT using a hot-lid thermal cycler, overlay the reaction mixture with two drops of mineral oil.

| Table III: Cycling Guidelines Based on Starting Material | | |
|---|---|---|
| No. of Cells (e.g. HeLa) | Typical Yield of Total RNA (ng) | Typical No. of PCR Cycles |
| ~10 | 0.15 | 27 |
| ~100 | 1.5 | 24 |
| ~1,000 | 15 | 20 |
| ~10,000 | 150* | 18 |

7. Commence thermal cycling using the following program:
   - 95°C 1 min
   - X cycles at:

95°C 15 sec
65°C 30 sec
68°C 3 min

a) Consult Tables II & III for guidelines. **Subject all tubes to 15 cycles.** Then, divide the PCR reaction mix between the "Experimental" and "Optimization" tubes, using the Optimization tube for each reaction to determine the optimal number of PCR cycles, as described in Step 8.
b) For applications requiring longer cDNA transcripts, increase to 6 min.

8. Subject each reaction tube to 15 cycles, then pause the program. Transfer 30 µl from each tube to a second reaction tube labeled "Optimization". Store the "Experimental" tubes at 4°C. Using the Tester PCR tube, determine the optimal number of PCR cycles (see Figure 3):
   a) Transfer 5 µl from the 15 cycle PCR reaction tube to a clean microcentrifuge tube (for agarose/EtBr gel analysis).
   b) Return the Optimization tubes to the thermal cycler. Run three additional cycles (for a total of 18) with the remaining 25 µl of PCR mixture.
   c) Transfer 5 µl from the 18 cycle PCR reaction tube to a clean microcentrifuge tube (for agarose/EtBr gel analysis).
   d) Run three additional cycles (for a total of 21) with the remaining 20 µl of PCR mixture.
   e) Transfer 5 µl from the 21 cycle PCR to a clean microcentrifuge tube (for agarose/EtBr gel analysis).
   f) Run three additional cycles (for a total of 24) with the remaining 15 µl of PCR mixture.
   g) Transfer 5 µl from the 24 cycle PCR to a clean microcentrifuge tube (for agarose/EtBr gel analysis).
   h) Run three additional cycles (for a total of 27) with the remaining 10 µl of PCR mixture.
   i) Transfer 5 µl from the 27 cycle PCR to a clean microcentrifuge tube (for agarose/EtBr gel analysis).
   j) Run three additional cycles (for a total of 30) with the remaining 5 µl of PCR mixture.
9. Electrophorese each 5 µl aliquot of the PCR reaction alongside 0.1 µg of 1 kb DNA size markers on a 1.2% agarose/EtBr gel in 1X TAE buffer. Determine the optimal number of cycles required for each experimental and control sample (see Figure 4, Section VI).
10. Retrieve the 15 cycle Experimental PCR tubes from 4°C, return them to the thermal cycler, and subject them to additional cycles, if necessary, until you reach the optimal number.
11. When the cycling 11. is completed, analyze a 5 µl sample of each PCR product alongside 0.1 µg of 1 kb DNA size markers on a 1.2% agarose/EtBr gel in 1X TAE buffer. Compare your results to Figure 4 to confirm that your reactions were successful.
12. Add 2 µl of 0.5 M EDTA to each tube to terminate the reaction.

**Column purification of PCR products using NucleoSpin gel and PCR clean-up**

1. Add 300 µl Binding NT Buffer to each 70 µl PCR reaction. Mix well by pipetting.
2. Place a NucleoSpin column into a 2 ml Collection Tube, and pipette the sample onto the filter. Centrifuge at 8,000 rpm for 1 min. Discard the Collection Tube and flowthrough.
3. Insert the NucleoSpin column into a fresh 2 ml Collection Tube. Add 600 µl Wash Buffer NT3 to the column. Centrifuge at 14,000 rpm for 1 min. Discard the flowthrough.
4. Return the column to the Collection Tube. Add 250 µl Wash Buffer NT3 to the column. Centrifuge at 14,000 rpm for 1 min. Discard the flowthrough.
5. Discard the flowthrough and spin again at 14,000 rpm for 1 min to remove the final traces of ethanol to dry the filter.
6. Transfer the NucleoSpin column to a clean 1.5 ml microcentrifuge tube. Pipette 50 µl Elution Buffer NE directly onto the filter, being careful not to touch the surface of the filter with the tip of the pipette. Allow the filter to soak for 2 min with the lid open.
7. Close the tube and centrifuge at 14,000 rpm for 1 min to elute PCR product. Save the column.
8. Determine the yield of each PCR product by measuring the A260. For each reaction, we usually obtain 1–2 µg of SMARTer cDNA after purification.
9. If no product is detected, perform elution (Steps 6 and 7) a second time, using a fresh 1.5 ml microcentrifuge tube.

**Qiagen Reverse Transcriptase (RT)**

1. Thaw gDNA wipeout buffer, Quantiscript reverse transcriptase, Quantiscript RT buffer, RT primer mix, and RNase-free water at room temperature. Thaw RNA in ice.

2. Prepare the genomic DNA elimination reaction on ice, making an extra tube without template RNA as a control, just gDNA wipeout buffer and water:

| | |
|---|---|
| • gDNA wipeout buffer | 2µl |
| • Template RNA | up to 2µg |
| • RNase-free water | up to 14 ul total volume |
| • TOTAL VOLUME: | 14µl |

3. Incubate for 2 minutes at 42°C. Place immediately back on ice. Do not exceed an incubation time of 10 minutes.

4. Add the remaining reagents to each tube. Add RNase-free water to the tube that is meant to be enzyme free:

| | |
|---|---|
| • Quantiscript reverse transcriptase | 1µl |
| • Quantiscript RT buffer | 4µl |
| • RT primer mix | 1µl |
| • TOTAL VOLUME: | 20 µl |

5. Incubate for 15 minutes at 42°C
   - May run for 30 minutes if RT-PCR product is greater than 200 bp or to increase cDNA yields.

6. Incubate for 3 minutes at 95°C

7. Add 1 µl of each sample to corresponding PCR mix. Store at -20°C long term.

**Diluting Primers from IDT**

To make freezer stock:

1. Spin down tubes
2. Add $dH_20$ water to 50 µM (50 pmoles/ul)
    - Divide amount of oligo in nMoles by 50 µM
    - Typical volumes range from 300- 800 µl
3. Vortex well
4. *Store at -20°C*
5. Note primer location on Primer Inventory Sheet

To make 10 µM PCR Stock:

100 µl of 50 µM primer
400 µl of $dH_20$
*Store at -20°C*

To make 2.5 µM working PCR stock

125 µl of 10 µM primer
375 µl of $dH_20$
*Store at -20°C*

## Polymerase Chain Reaction: PCR

- Wear gloves
- Make aliquots of kit components for own use – except for Taq
- Make sure to vortex $MgCl_2$ well
- Make master mix for number of samples + 1

**Sample Types:**

1.) cDNA from RT reaction
2.) RT negative control: no enzyme
3.) RT negative control: no template RNA
4.) PCR negative control: no template RT added

**PCR Reaction Master Mix:**

| Reagents | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | x10 | x11 | x12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Std Taq Buffer | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 |
| Forward GSP | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 32 | 36 | 40 | 44 | 48 |
| Reverse GSP | 4 | 8 | 12 | 16 | 20 | 24 | 28 | 32 | 36 | 40 | 44 | 48 |
| MgCl2 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | 36 |
| dNTPs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Taq | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 | 6 |
| Water | 31.5 | 63 | 94.5 | 126 | 157.5 | 189 | 220.5 | 252 | 283.5 | 315 | 346.5 | 378 |
| Total Per Tube | 49 | 49 | 49 | 49 | 49 | 49 | 49 | 49 | 49 | 49 | 49 | 49 |
| Template Per Tube | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| Reagents | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | x10 | x11 | x12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Std Taq Buffer | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 |
| MgCl2 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | 36 |
| dNTPs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Taq | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 | 6 |
| Water | 31.5 | 63 | 94.5 | 126 | 157.5 | 189 | 220.5 | 252 | 283.5 | 315 | 346.5 | 378 |
| Total Per Tube | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 41 |
| Forward GSP | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Reverse GSP | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Template Per Tube | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

- Mix well and centrifuge
- Make sure no bubbles are present before PCR reaction

**PCR conditions:**

- 1 cycle: 94°C for 3 min
- 35 cycles: 94°C for 30 sec, ____°C for 30 sec, 72°C for 30 sec
- 1 cycle: 72°C for 10 min
- Hold at 4°C

*Store at -20℃ until analysis*

**MgCl2 Gradient Option:**

| **Reagents** | **x1** | **x2** | **x3** | **x4** | **x5** | **x6** |
|---|---|---|---|---|---|---|
| Std Taq Buffer | 5 | 10 | 15 | 20 | 25 | 30 |
| Forward GSP | 4 | 8 | 12 | 16 | 20 | 24 |
| Reverse GSP | 4 | 8 | 12 | 16 | 20 | 24 |
| MgCl2 | 1 | 2 | 3 | 4 | 5 | 6 |
| dNTPs | 1 | 2 | 3 | 4 | 5 | 6 |
| Taq | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| Water | 33.5 | 32.5 | 31.5 | 30.5 | 29.5 | 28.5 |
| Total Per Tube | 49 | 49 | 49 | 49 | 49 | 49 |
| Template Per Tube | 1 | 1 | 1 | 1 | 1 | 1 |

## DNA Agarose Gel

During electrophoresis, water is electrolyzed, which generates protons at the anode, and hydroxyl ions at the cathode. The cathode end of the electrophoresis chamber then becomes basic and the anodal end acidic. The use of a buffering system therefore required when charged molecules are electrophoresed through a separation medium. The pH of the buffers is basic and the phosphate backbone of DNA has a net negative charge and migrates towards the anode.

**Make certain to wear gloves. Ethidium Bromide is a known mutagen**

**Need –**
**Distilled water**
**Clean gel box, casting tray and comb with $dH_20$**
**Ethidium bromide solution**
Stock: 10 mg/ml $dH_20$
0.1 g to 10 ml $dH_20$
Store at 4°C
Working solution: 2.5 ug/ul (4X dilution of stock)
(125 ul into 375 ul $dH_20$)

**Running buffer:**
**50X TAE**
242g Tris base
57.1 ml glacial acetic acid
100 ml 0.5 M EDTA pH 8.0 (FW 372 = 18.6 g)
$dH_20$ to 1 Liter

**1X TAE**
20 ml 50X TAE
980 ml $dH_20$

**Gel preparation: (Agarose PCR grade BP 2410-100)**

| Agarose (%) buffer (g) | Effective range of resolution of linear DNA fragments (kb) | Amount in 50 mls 1X TAE |
|---|---|---|
| 0.5 | 30 to 1 | 0.25 |
| 0.7 | 12 to 0.8 | 0.35 |
| 1.0 | 10 to 0.5 | 0.5 |
| 1.2 | 7 to 0.4 | 0.6 |
| 1.5 | 3 to 0.2 | 0.75 |
| 2.0 | | 1.0 |

For 50 ml gel, place agarose in a 125 ml flask
(choose a flask that is 2-4 times that size of the gel volume)
Add 50 ml 1X TAE buffer and swirl to mix

Cover flask with double layer of plastic wrap
Heat in microwave for 1 minute on high power.
Open microwave and swirl flask ****Flask is hot** Tilt flask away from your face*****
Heat in microwave for another 30sec to 1 minute until agarose is completely melted (no bubbles or "little contact lenses")
Remove from microwave and cool solution until you can touch the bottom with your hand
After solution has cooled, add 10 ul of 2.5 ug/ul ethidium bromide solution
(Final gel concentration is **0.5 ug/ml)**
Pour gel into casting tray with comb, pop any air bubbles with a pipet tip, and let gel polymerize

After gel has set, remove comb and turn gel
Pour tank buffer 1X TAE to just cover gel

**Sample preparation:**
Agarose gel loading dye is 6X. Add to samples so concentration is 1X
Run 12 ul of PCR reaction + 2 ul of gel loading dye = 14 ul total volume

**Running gel:**
Add 10 ul of ready made DNA ladder to first well to assist in determining the size of the DNA fragments observed.
Load samples into wells of gel
Note sample loading in lab notebook.

Plug in red and black leads to power supply (make sure these are in the correct position!) and run get at 80 V for approximately 45 minutes or until the dye front is near the bottom of the gel. Look for bubbles to make sure you have current.

TURN OFF THE POWER!!!! AND UNPLUG THE UNIT FROM THE POWER SUPPLY!!!!!
With gloves, remove the gel bed onto a piece of saran wrap. Observe stained gel on UV gel documentation system and save an image of your gel. Make certain to wear gloves when handling gel due to ethidium bromide.

WASTE: Gel is placed in ethidium bromide waste container and running buffer is poured down the sink

**QIAquick Gel Extraction Kit Protocol**

1.) Excise the DNA fragment from the agarose gel with a clean, sharp scalpel. Minimize the size of the gel slice by removing extra agarose.

2.) Weigh the gel slice in a colorless tube. Add **3 volumes of Buffer QG** to 1 volume of gel (100mg ~100 ul).

3.) Incubate at **50°C for 10 min** (or until the gel slice has completely dissolved). Vortex the tube every 2-3 min during the incubation to mix.

4.) After the gel slice has dissolved completely, check that the color of the mixture is yellow.

5.) Add **1 gel volume of isopropanol** to the sample and mix.

6.) Place a QIAquick spin column in a provided 2 ml collection tube (already done).

7.) To bind DNA, **apply the sample to the QIAquick column, and centrifuge for 1 min**. The maximum volume of the column reservoir is 800 ul. For sample volumes of more than 800 ul, simply load and spin again.

8.) Discard flow-through and place QIAquick column back in the same collection tube.

9.) Add **500ul of Buffer QG** to QIAquick column and centrifuge for 1 min. This step will remove all traces of agarose.

10.) To wash, add **750ul of Buffer PE** to QIAquick column and centrifuge for 1 min.

11.) Discard flow-through and **centrifuge the QIAquick column for an extra 1 min** at ≥10,000 X g (~13,000 rpm).

12.) Place a QIAquick column into a **clean 1.5 ml microcentrifuge tube**.

13.) To elute DNA, add **50 ul autoclaved H2O** to the center of the QIAquick membrane and centrifuge the column for 1 min at maximum speed. Reapply the flow-through and centrifuge again for 1 min.

14.) Check concentration of samples by nanodrop or running a 2% agarose gel

***Store samples at -20°C***

**Sequencing Gel Cutouts**

Following URI GSC Instructions:

Target amounts for dsDNA templates:

PCR products: 2.5 ng DNA per 100 bases per reaction

Plasmids: 300-500 ng DNA per reaction

Primer amount:

Use one primer only; either forward or reverse, but not both!

5 pmol per reaction (Note: 5 pmol = 2.0 µl of a 2.5 µM stock)

Single sample volume:

12 µl per reaction; add template plus one primer in the amounts above to MB grade water.

To facilitate pipetting, submit your sample in duplicate with a total volume of 24 µl. Submit your template and primer combined in a 0.5 or 1.5ml tube. DO NOT submit samples in individual 0.2 ml (200µl) tubes. When submitting 16 or more samples, please submit them in 8-tube strip-tube(s) (capped) or a 96-well plate (capped or sealed).

**qPCR: Standard Curve Protocol**

1. Run your primers that will be used in QPCR on sperm/testis samples using standard PCR protocol. Whichever sample type gives you the darkest/best product band should be used regardless of what sample type you want to investigate using QPCR. (Example: testis bands can be used to generate standard curve even if you are only going to be running sperm samples on QPCR for your critical data).

2. Run DNA agarose gel of your PCR and cut out product band for your gene of interest.

3. Use the gel extraction kit to purify the gel cut out.

4. Bring the purified DNA product to Nanodrop and get a concentration for you DNA solution, which is purified DNA of your gene of interest.

5. Enter your concentration (ng/ul) from the Nanodrop and the size of your product into the DNA copy calculator on the URI Genomics and Sequencing Center website (http://www.uri.edu/research/gsc/resources/cndna.html). This will give you the number of copies of your gene/ul of sample.

6. Once you have found the # of copies/ul of your sample, you must dilute this sample out to get $1X10^8$ copies/ul which will become your highest standard for the QPCR standard curve. Note: this dilution is usually a very large dilution factor so splitting it up into 2 separate dilutions may help pipetting accuracy.

7. From your $1X10^8$ copies/ul sample you need to do a serial dilution down to $1X10^1$ copies/ul, which will become your final standard in the curve.

> Example: 20ul $1X10^8$ copies/ul sample +180 ul $H_2O$ => 200 ul of $1X10^7$ copies/ul
> Repeat this dilution starting with 20ul of the $1X10^7$ copies/ul sample and add the 180ul of water, which will result in 200ul of your $1X10^6$ copies/ul standard
>
> - This should be done until you have 8 standards made, $1X10^8$ down to $1X10^1$ copies/ul

8. Once you have all standards diluted out put aside for use in QPCR runs. Standards are in solution with a concentration based on one ul of sample therefore all standards will only have 1ul added to tubes for QPCR samples as opposed to unknowns which have 2.5ul of sample added to each tube. This just requires adjusting water between unknowns and standards when running QPCR sample.

*Proceed to QPCR protocols once standard curve is diluted out accordingly.*

**qPCR (Stratagene MX3005P Machine)**

1. Dilute ROX reference dye 1:500 using nuclease-free PCR grade water. ***Keep all solutions containing reference dye away from light.**

2. Prepare the experimental reaction by combining components IN ORDER. Combine reagents to form a master mix in the amount of number of samples + one to facilitate pipetting.

| Reagent | Standards | Unknown |
|---|---|---|
| PCR water | 9.125 µl | 7.625 µl |
| SYBR master mix | 12.5 µl | 12.5 µl |
| Forward primer | 1 µl | 1 µl |
| Reverse primer | 1 µl | 1 µl |
| Diluted reference dye | 0.375 µl | 0.375 µl |
| Total volume (per sample) | 24 µl | 22.5 µl |

3. Gently mix without creating bubbles (do not vortex), then distribute mixture to each reaction tube.

4. Add template cDNA to each sample. (1 µl to standards, 2.5 µl to unknowns)

5. Gently mix without creating bubbles (do not vortex)

6. Centrifuge briefly.

7. Place reactions in instrument (Mx 3005) and run qPCR program.

qPCR cycling conditions

| Cycles | Duration | Temperature (°C) |
|---|---|---|
| 1 | 10 min | 95 |
| 40 | 30 sec<br>1 min | 95<br>59.0 (Annealing) |
| 1 (Dissociation curve) | 1 min<br>30 sec<br>30 sec | 95<br>55<br>95 |

Samples can be run on a 2% agarose gel afterward to check products.

## qPCR (Roche LightCycler 480 Machine)

Important notes before starting:
- work with lab lights off if others don't mind
- wrap all master mix tubes with aluminum foil to block light

1. Prepare the experimental reaction by combining components in order. Combine reagents to form a master mix in the amount of number of samples + one additional sample for every 8 samples to facilitate pipetting (for example: 40 samples run on qPCR = 45x master mix).

| Reagent | Amount/Sample |
|---|---|
| PCR water | 5 µl |
| SYBR master mix | 10 µl |
| Forward primer | 2 µl |
| Reverse primer | 2 µl |
| Total volume (per sample) | 19 µl |

2. Gently mix without creating bubbles (pulse vortex), and then distribute 19 µl of master mix to each well of qPCR plate (white) that will be used for samples.

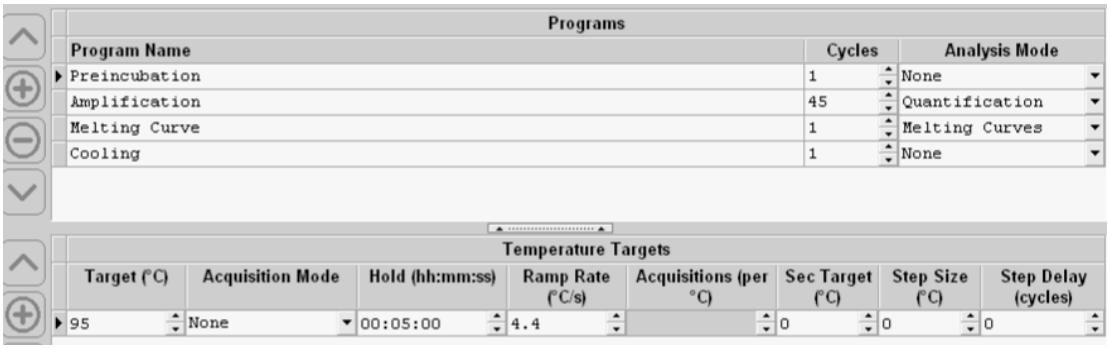
3. Add 1 µl of sample (standard or unknown) to appropriate wells.

4. Cover plate with plate sealing film and firmly seal sticky cover onto plate with plastic sealing applicator.

5. Bring plate up to plate centrifuge in GSC and spin plate at 1500 rcf for 2 minutes (make sure plate is balanced with balance plate from drawer below centrifuge).

6. Cycling conditions are to be set as follows within the LightCycler 480 software:

**Pre-incubation**



**Amplification** (60 degree temperature is annealing and could vary pending primers used)

| Program Name | Cycles | Analysis Mode |
|---|---|---|
| Preincubation | 1 | None |
| Amplification | 45 | Quantification |
| Melting Curve | 1 | Melting Curves |
| Cooling | 1 | None |

Amplification Temperature Targets

| Target (°C) | Acquisition Mode | Hold (hh:mm:ss) | Ramp Rate (°C/s) | Acquisitions (per °C) | Sec Target (°C) | Step Size (°C) | Step Delay (cycles) |
|---|---|---|---|---|---|---|---|
| 95 | None | 00:00:10 | 4.4 | | 0 | 0 | 0 |
| 60 | None | 00:00:10 | 2.2 | | 0 | 0 | 0 |
| 72 | Single | 00:00:15 | 4.4 | | 0 | 0 | 0 |

## Melting Curve

| Program Name | Cycles | Analysis Mode |
|---|---|---|
| Preincubation | 1 | None |
| Amplification | 45 | Quantification |
| Melting Curve | 1 | Melting Curves |
| Cooling | 1 | None |

Melting Curve Temperature Targets

| Target (°C) | Acquisition Mode | Hold (hh:mm:ss) | Ramp Rate (°C/s) | Acquisitions (per °C) | Sec Target (°C) | Step Size (°C) | Step Delay (cycles) |
|---|---|---|---|---|---|---|---|
| 95 | None | 00:00:05 | 4.4 | | | | |
| 65 | None | 00:01:00 | 2.2 | | | | |
| 97 | Continuous | | 0.11 | 5 | | | |

## Cooling

| Program Name | Cycles | Analysis Mode |
|---|---|---|
| Preincubation | 1 | None |
| Amplification | 45 | Quantification |
| Melting Curve | 1 | Melting Curves |
| Cooling | 1 | None |

Cooling Temperature Targets

| Target (°C) | Acquisition Mode | Hold (hh:mm:ss) | Ramp Rate (°C/s) | Acquisitions (per °C) | Sec Target (°C) | Step Size (°C) | Step Delay (cycles) |
|---|---|---|---|---|---|---|---|
| 40 | None | 00:00:30 | 2.2 | | 0 | 0 | 0 |

Analysis Notes:
- perfect standard curve has efficiency of 2.00
- standard curve error should remain under 0.20
- melting curve should show one distinct peak per sample for each primer set
    - small peak at lower temperature usually indicates primer dimers

**qPCR Troubleshooting:**
**Primer Matrix**

**Purpose:**
The primer matrix is done to find the best primer concentrations that offer optimal results on QPCR runs. The idea of the primer matrix is to combine different amounts of the Forward and Reverse primers to find the best combo which could be one in which the primers differ in concentration with each other.

**Why:**
The SYBR Green dye we use in QPCR binds to any double stranded DNA products and emits fluorescence. This makes it critical that when doing QPCR you are only amplifying your target product and no non-specific secondary products or primer dimers (forward and reverse primers sticking to each other). These other products will have SYBR bind to them and emit fluorescence thus throwing off your final data.

**Procedure:**
The standard QPCR protocol is followed when doing a primer matrix. The only variations are altering the amount of primer added between tubes. You must use the same sample(template) for all tubes so that they only thing that is altered across the run are primer concentrations. A standard primer matrix that should be sufficient is a 4x4 primer matrix which is represented below.

| | **F 50nM (.5ul)** | **F 100nM (1ul)** | **F 300nM (3ul)** | **F 600nM (6ul)** |
|---|---|---|---|---|
| **R 50nM (.5ul)** | F 50nM (.5ul) R 50nM (.5ul) | F 100nM (1ul) R 50nM (.5ul) | F 300nM (3ul) R 50nM (.5ul) | F 600nM (6ul) R 50nM (.5ul) |
| **R 100nM (1ul)** | F 50nM (.5ul) R 100nM (1ul) | F 100nM (1ul) R 100nM (1ul) | F 300nM (3ul) R 100nM (1ul) | F 600nM (6ul) R 100nM (1ul) |
| **R 300nM (3ul)** | F 50nM (.5ul) R 300nM (3ul) | F 100nM (1ul) R 300nM (3ul) | F 300nM (3ul) R 300nM (3ul) | F 600nM (6ul) R 300nM (3ul) |
| **R 600nM (6ul)** | F 50nM (.5ul) R 600nM (6ul) | F 100nM (1ul) R 600nM (6ul) | F 300nM (3ul) R 600nM (6ul) | F 600nM (6ul) R 600nM (6ul) |

- All amounts listed in the primer matrix are amount of 2.5uM primer stock that needs to be added. This calculation is based on the final volume of each QPCR tube being 25ul.

- After adding various amounts of primer to the solutions the $H_2O$ amounts will need to be altered accordingly to maintain a final volume of 25ul in each QPCR tube.

**Analysis:**
A dissociation curve(melting curve) should be done during all QPCR reactions with the primer matrix being no exception. There are a few criteria that must be looked at to be able to pick the best primer pair to move forward with in your QPCR assay. Based on analyzing the amplification plots and dissociation curve you can choose the best pair based on the following criteria:

1. There must be only one product being amplified.
    - Only one peak in the dissociation curve around 85-88° usually
    - Primer dimers will produce a small peak around 10° lower than your main product (unacceptable)
    - Secondary non-specific products can produce second higher temperature peaks or a "shoulder" off of your main product peak (unacceptable)
    - Samples can be run on DNA Agarose gels after QPCR to check for single product amplification as well as do sequence analysis

2. The lower the Ct the better when comparing combinations.
    - If multiple primer pairs show single product amplification you should choose the one with the lowest Ct(cycle where the amplification plot crosses the threshold line).

3. Amplification must of course be present on the amplification plot
    - Sometimes primers that are too low in concentration (example: Forward 50nM, Reverse 50nM combination) do not amplify due to insufficient primer amounts

**Once a usable primer pair has been determined you can move ahead with your QPCR assay**

**RNA-Seq Analaysis via Galaxy (usegalaxy.org)**

Notes before using Galaxy:

- Galaxy is very user friendly and will only let you choose files that are in correct format for any analysis you are trying to do.
- Analysis takes much longer on Galaxy server than through command line
- All output files get named with generic names; renaming after running them is beneficial
- Making new histories for different sample analysis is helpful for organizational purposes
- For the most part you can leave all parameters as default unless you start doing detailed analysis and become very familiar with advanced options

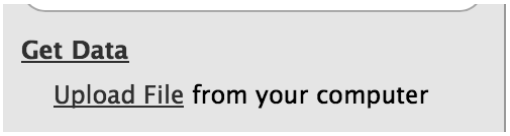
Files you will need to upload to galaxy server:

- Raw Illumina sequencing files (file 1 and file 2 for paired-end reads; fasta/fastq)
- Annotated transcript file (igenomes; .gtf format)
- You do not need to upload bovine genome as galaxy has genomes already on it

**From raw reads => Cufflinks/Cuffdiff analysis**

**Uploading Files**

Left Tools Panel: Get Data => Upload File

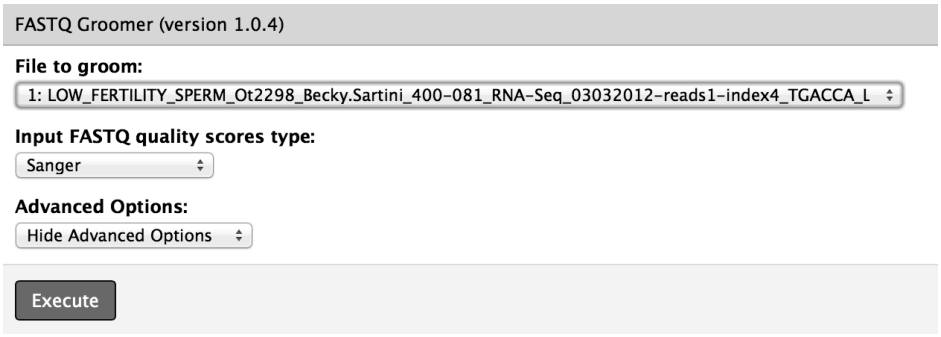



Browse for file(s) you want to upload and submit the job.

**FASTQ Groomer**

Left Tools Panel: NGS:QC and manipulation => FASTQ Groomer

What it does: Changes format of raw sequence files into usable format for downstream applications

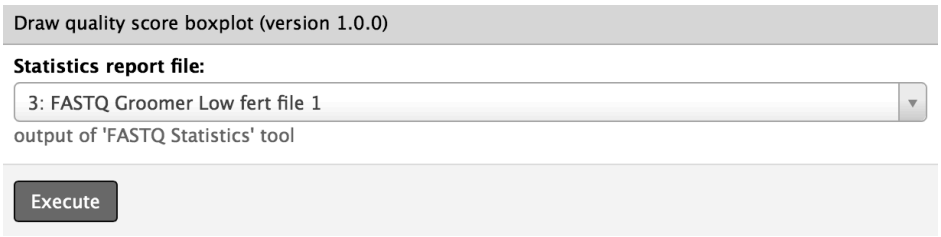



File to groom is your raw data sequence file. Choose Sanger for input type and no need to adjust any advanced options.

**Draw Quality Score Boxplot**

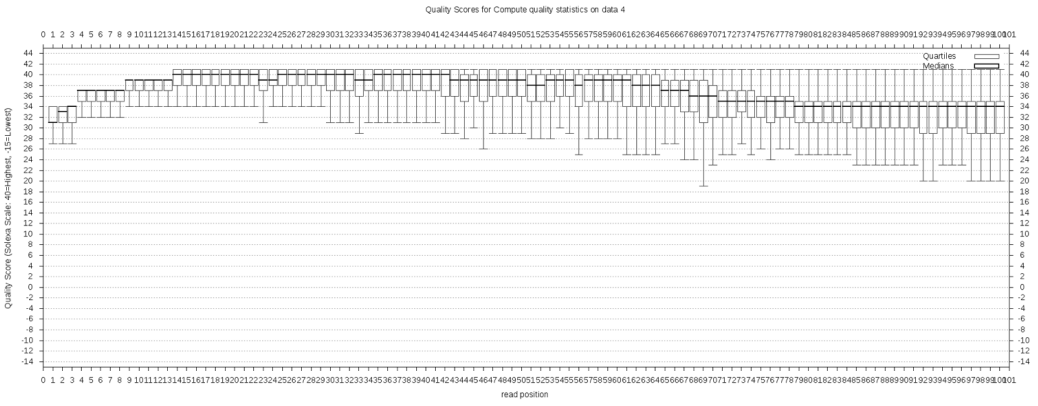
Left Tools Panel: NGS: QC and manipulation => Draw quality score boxplot

What it does: Scores each base of your reads across entire dataset with quality score to see if your read accuracy is consistent throughout all bases of sequencing.




Your input file is any FASTQ groomed file that you want to check read quality on.

Results look like this:

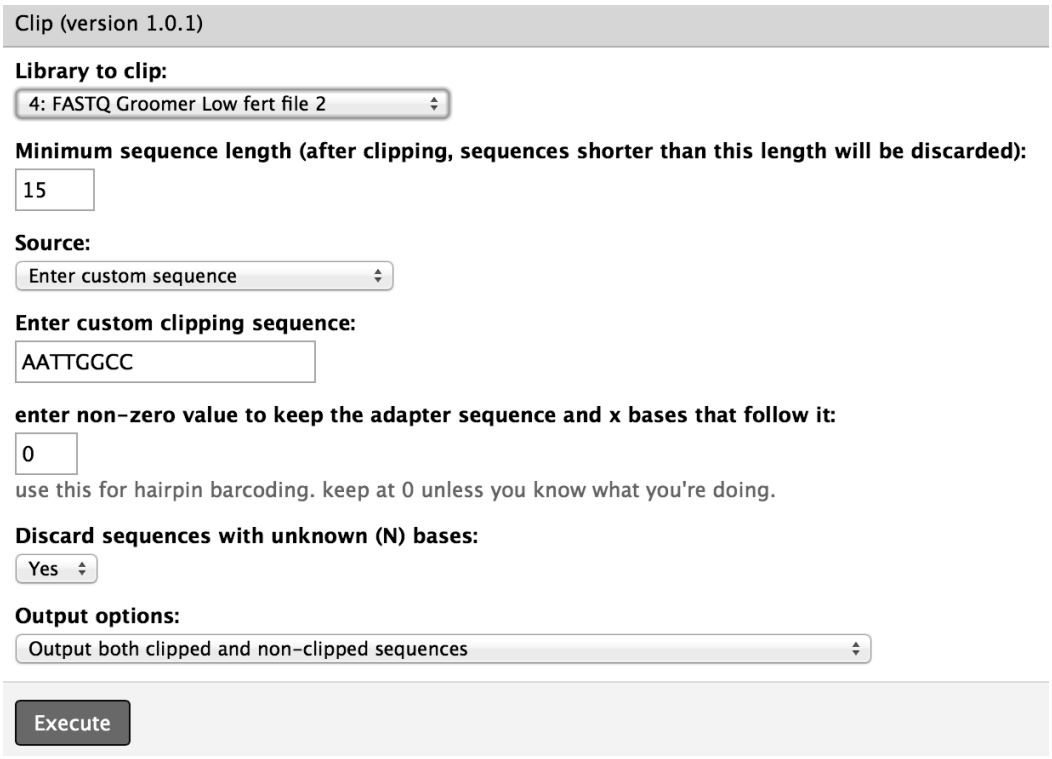



The higher the score, the better. A huge drop off on one end of the sequence could point to 3' bias with the 5' end not sequencing very accurately.

**Clipping adapter sequences**

Left Tools Panel: NGS: QC and manipulation => Clip

What it does: Removes adapters from ends of reads; removes reads with Ns; removes reads that are too short.

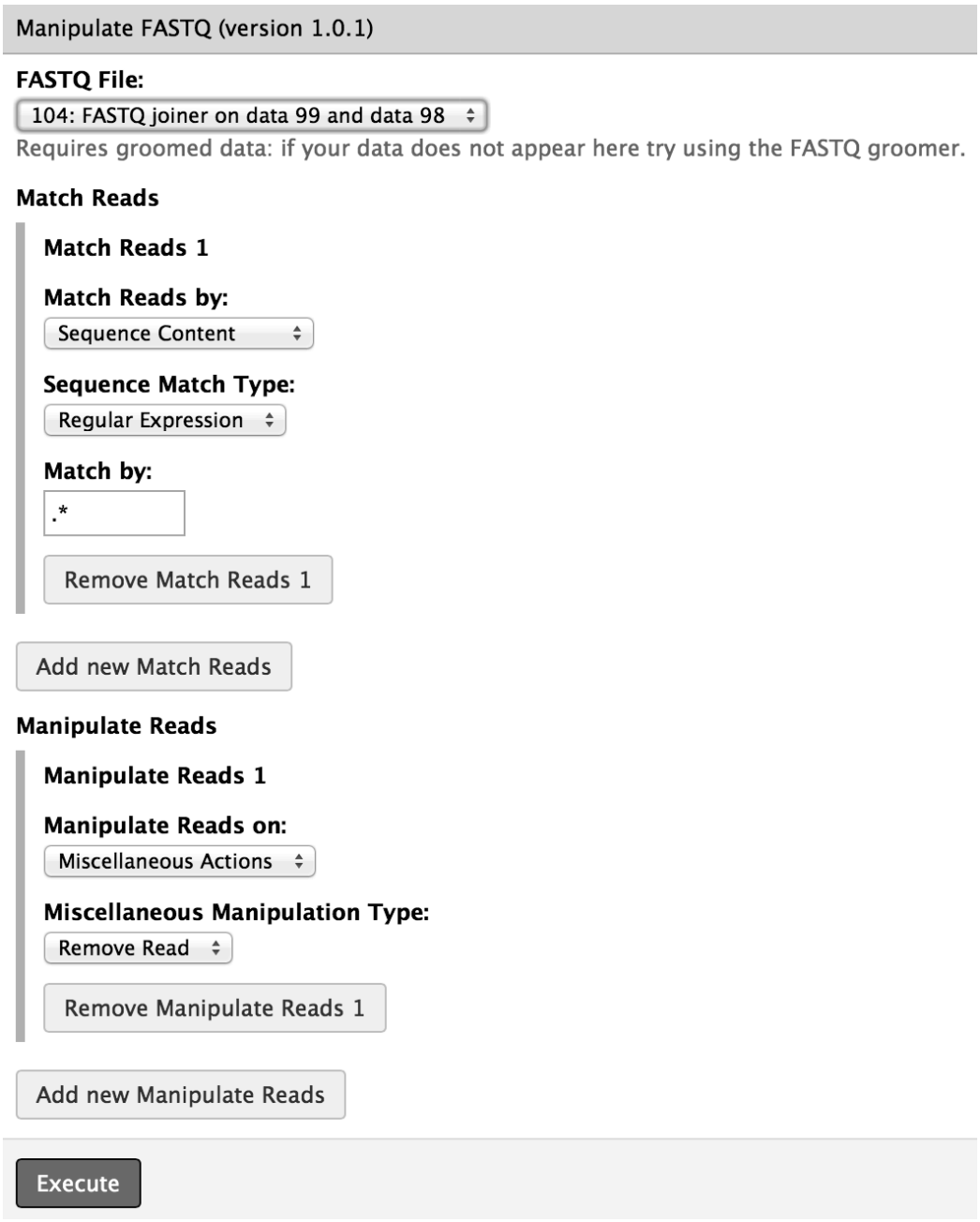



Library to clip input file is your FASTQ groomed file. You must change the source to "enter custom sequence" and then type in the sequencing adapter that you want clipped (example here is just random bases typed in). You want to change output options to "output both clipped and non-clipped sequences" so that you keep reads that the adapter was not maintained on as well as those you clip.

**Manipulate Reads (Delete reads overrepresented from contamination; concatemers)**

Left Tools Panel: NGS: QC and manipulation => Manipulate FASTQ

What it does: You can remove reads that contain a certain sequence that is due to some form of contamination.

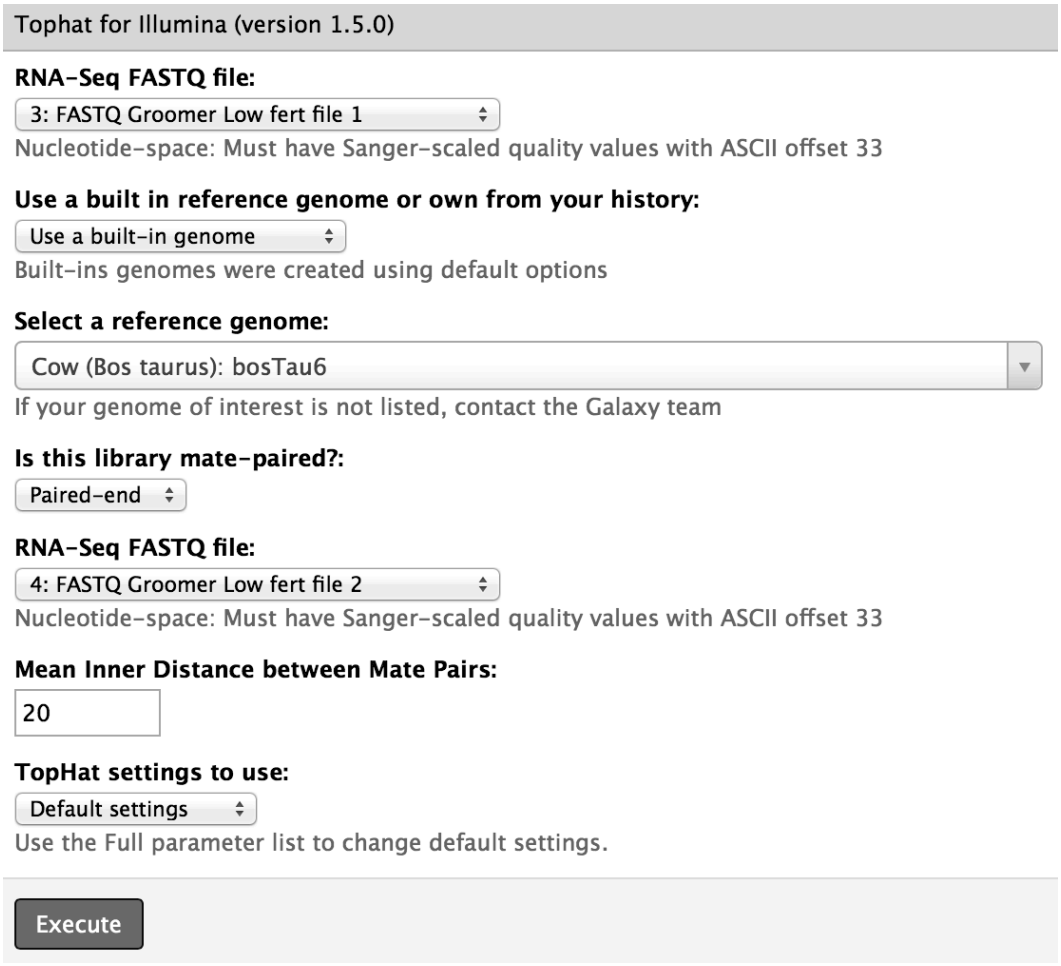



You must match reads by "sequence content" and type in the sequence you are looking for in contaminated reads. For the manipulation section you go to manipulate reads on "miscellaneous actions" and for manipulation type choose "remove read".

**Tophat**

Left Tools Panel: NGS: RNA Analysis => Tophat for Illumina

What it does: Aligns your reads to the bovine (or species you are studying) genome.




Galaxy has a large selection of built in genomes so they likely have what you need; in this example bosTau6. This prevents you from having to upload your own genome file.

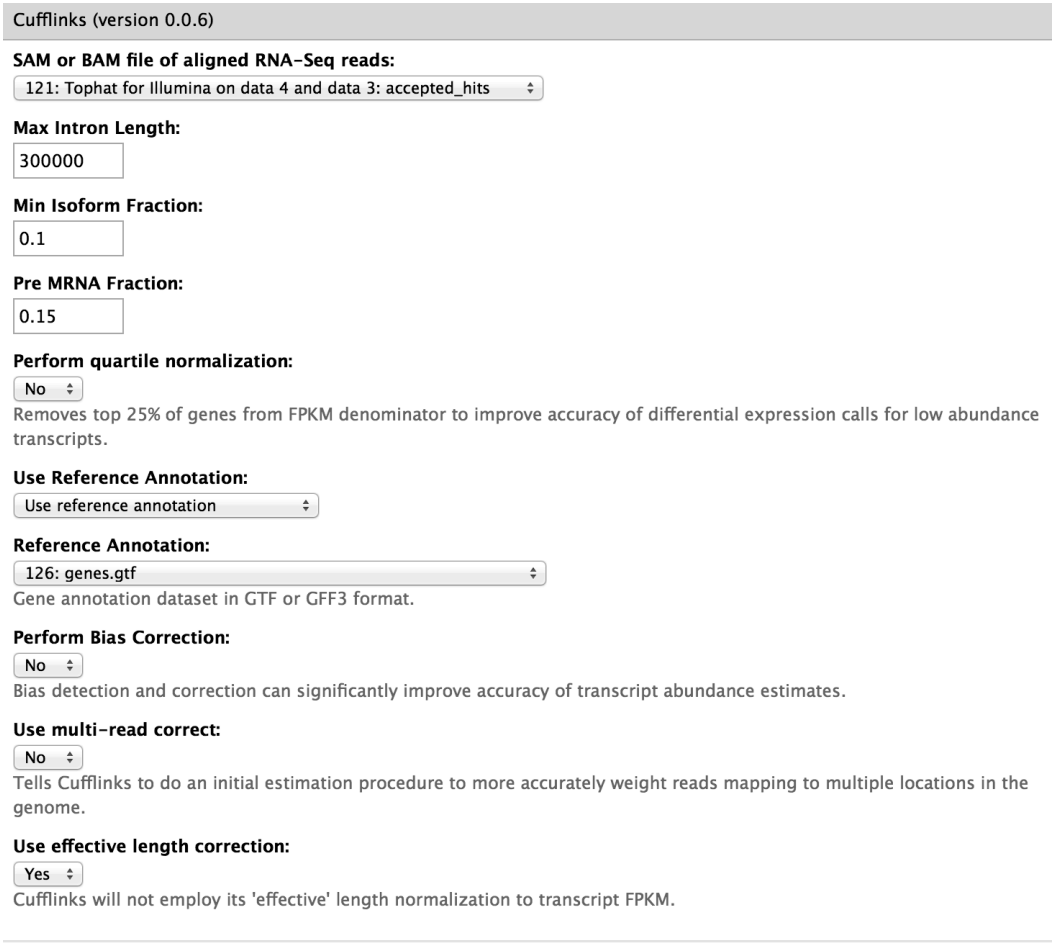
If you have paired-end reads you want to change the library mate-paired section to paired-end. This will cause a second RNA-Seq FASTQ file section to pop up. Put your first groomed/clipped/manipulated file in the first section and the second file of the paired-end sequencing in the second section. If you know mean inner distance between pairs you can change this but if not leave it as default.

Results: You can use the accepted hits file that is one of the output files for subsequent cufflinks or cuffdiff analysis. This file can also be downloaded and used for subsequent sequencing statistic/quality measures through other analysis packages such as ever-seq.

**Cufflinks**

Left Tools Panel: NGS: RNA Analysis => Cufflinks

What it does: Takes your alignment file from tophat and tells you what transcripts on the genome that reads aligned to as well as quantifies them.




The SAM or BAM input aligned RNA-Seq reads file you need to use is the accepted hits file from your tophat analysis.

The only change you need to make to default parameters is that you want to "use reference annotation" in the use reference annotation setting. Then you have to choose which file in your galaxy history to use in the reference annotation pulldown that now appears. This will be the igenomes .gtf file that you uploaded originally along with your raw sequence files.

Results: You can download the transcript expression file and open it through excel to sort by FPKM and that is your list of expressed transcripts.

**Cuffdiff**

Left Tools Panel: NGS: RNA Analysis => Cuffdiff

What it does: Compares alignment files of two different samples (for example: low and high fertility) and determines differentially expressed transcripts.

Cuffdiff (version 0.0.6)

**Transcripts:**

129: Cufflinks on data 121 and data 126: assembled transcripts

A transcript GFF3 or GTF file produced by cufflinks, cuffcompare, or other source.

**Conditions**

**Condition 1**

**Name:**

High Fert

**Replicates**

**Replicate 1**

**Add replicate:**

88: Tophat for Illumina on data 4 and data 3: accepted_hits

Add new Replicate

**Condition 2**

**Name:**

Low Fert

**Replicates**

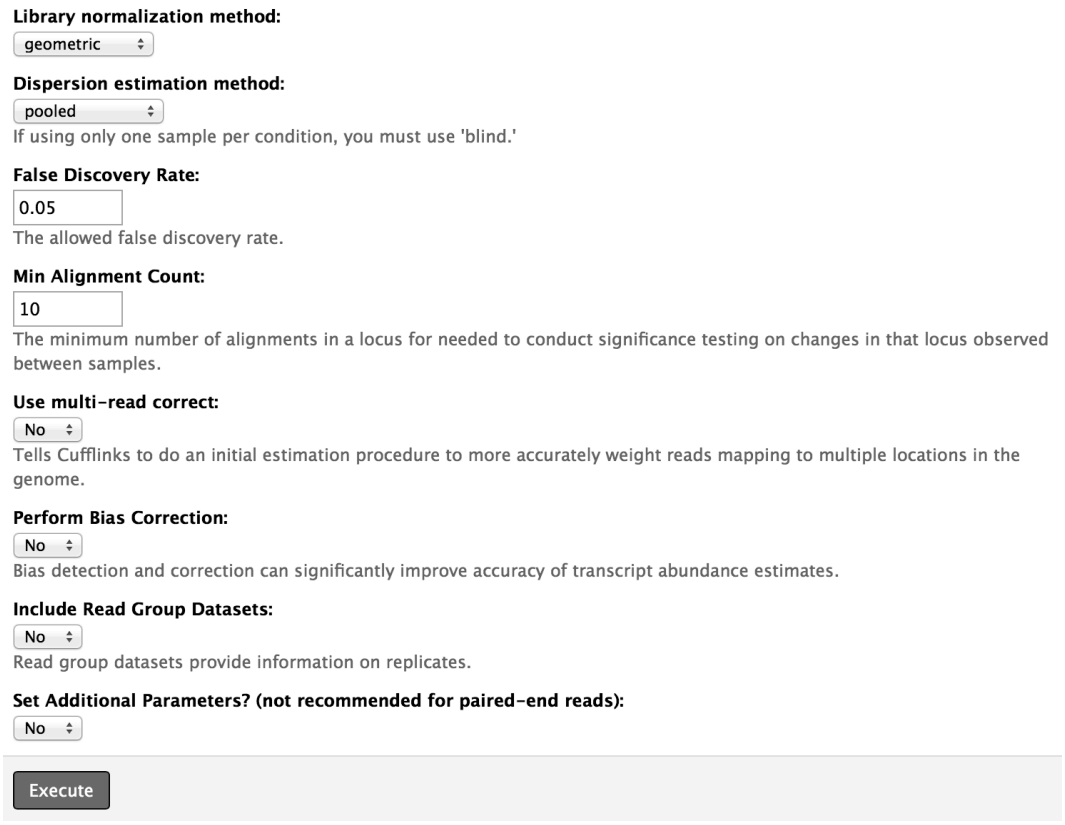
**Replicate 1**

**Add replicate:**

80: Tophat for Illumina on data 76: accepted_hits

Add new Replicate

Add new Condition




For the transcripts file you can use the igenomes .gtf file that you would likewise use in cufflinks analysis.

Each condition is your different sample types (for example: low and high fertility). You must use the accepted hits files from tophat for each sample type.

Results: You can download the differential expression file and sort through it in excel to determine what transcripts were significantly differentially expressed. This output will also give you FPKM of these transcripts in each sample but they will slightly differ from cufflinks FPKMs.

## Command Line RNA-Seq Analysis

Notes:

Installation of various software packages for command line use can differ depending on the package but most instructions can be found online.

For example, instructions for installation of cufflinks package on new can be found in the user manual online at: http://cufflinks.cbcb.umd.edu/tutorial.html#inst

You can drag and drop analysis/input files from their folders on the computer and it will automatically fill in the path to the file so that it can be used in your analysis

To just type in a file name you must be in the correct directory within the command line which can be accessed through the "cd" command (this is not necessary if you drag and drop each file you are using in the appropriate place within the command as it will fill in the pathing for you)

For example: **cd ~/desktop** will change the command directory to the desktop in which case you could type in a file or folder name from the desktop without having to drag and drop or write in the entire path to that folder

For almost all command line analysis you can direct where the output goes by using the command –o

For example: **-o /Users/chriscard/Desktop/seqoutput** on the end of command will send output to the folder named "seqoutput" on the desktop (after –o you can just drag the folder from the desktop and it automatically fills in the rest of the pathing to that folder)

This will change the heading for the command line from the home directory to the desktop as you can see in the following screenshot

```
Chriss-MacBook-Pro-4:~ chriscard$ cd ~/desktop
Chriss-MacBook-Pro-4:desktop chriscard$
```

**Cufflinks from command line**

Helpful information can be found online in the cufflinks user manual at: http://cufflinks.cbcb.umd.edu/manual.html

Input files you will need: alignment file (accepted hits tophat file; .bam format); annotated transcript file (igenomes; .gtf format)

First change directory to location where your input files are located:
**cd ~/Desktop/Card_LowHigh_Files**

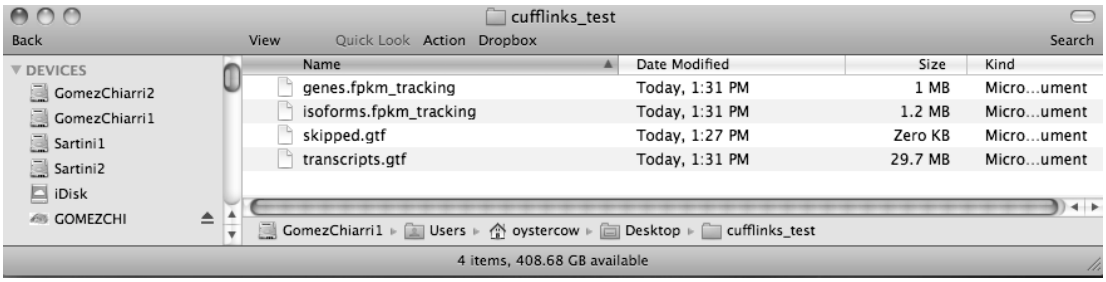
then you are ready for the cufflinks command which consists of the aligned .bam accepted hits file from tophat, an annotated transcript file, and an output location:

**cufflinks lowfertacchits.bam –G igenomegenes.gtf –o /Users/oystercow/Desktop/cufflinks_test** (all one continuous line just didn't fit here)

If done properly cufflinks should start to run in the command line. Below are screenshots showing the command line window and then the resulting files in the "cufflinks_test" folder on the desktop that you specified the output to be sent to

```
oystercows-Mac-Pro:~ oystercow$ cd ~/Desktop/Card_LowHigh_Files
oystercows-Mac-Pro:Card_LowHigh_Files oystercow$ cufflinks lowfertacchits.bam -G igenomegenes.gtf -o /Users/oystercow/Desktop/cufflinks_test
You are using Cufflinks v2.1.1, which is the most recent release.
[13:27:34] Loading reference annotation.
[13:27:36] Inspecting reads and determining fragment length distribution.
> Processed 13159 loci.                        [*************************] 100%
> Map Properties:
>       Normalized Map Mass: 1019444.77
>       Raw Map Mass: 1019444.77
>       Fragment Length Distribution: Truncated Gaussian (default)
>                     Default Mean: 200
>                  Default Std Dev: 80
[13:27:45] Estimating transcript abundances.
> Processed 13159 loci.                        [*************************] 100%
```

and the resulting files:




**Cuffdiff from command line**

Additional helpful information can be found in the cufflinks user manual linked above.

Cuffdiff is installed when the cufflinks package is installed so if cufflinks is working you are ready to go with cuffdiff.

Input files you will need include: annotated transcript file (igenomes; .gtf format); alignment file (low fertility accepted hits from tophat; .bam format); second condition alignment file (high fertility accepted hits from tophat; .bam format)

Stay in the same directory as cufflinks as the files you will use are in the same location.

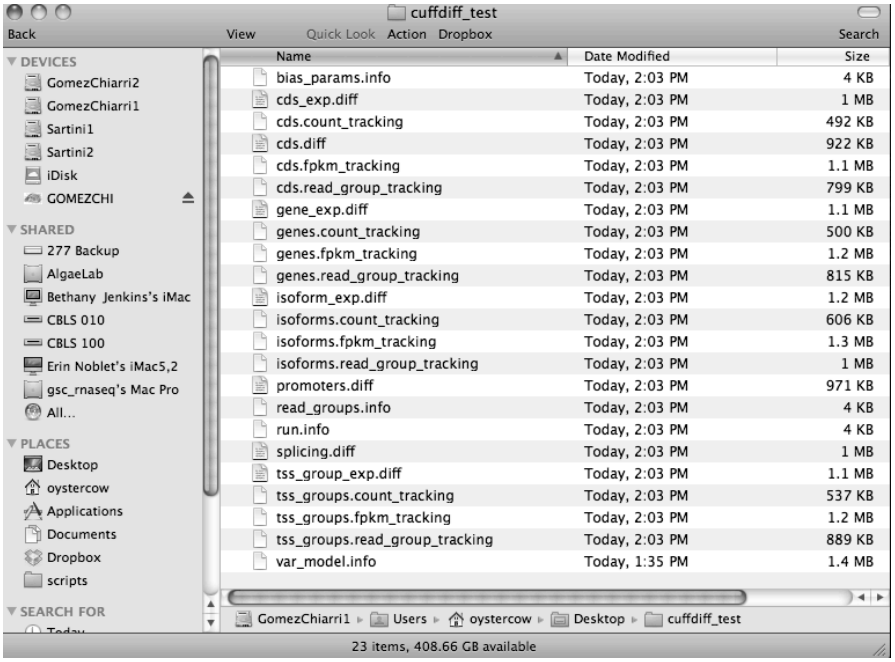
The command order for cuffdiff is annotated transcript reference file, alignment file #1, alignment file #2, and output location.

**Cuffdiff igenomegenes.gtf lowfertacchits.bam highfertacchits.bam –o /Users/oystercow/Desktop/cuffdiff_test** (all one continuous line again)

The final commands are as follows:

```
oystercows-Mac-Pro:Card_LowHigh_Files oystercow$ cuffdiff igenomegenes.gtf lowfertacchits.bam highfertacchits.bam -o /Users/oystercow/Desktop/cuffdiff_test
You are using Cufflinks v2.1.1, which is the most recent release.
[13:33:58] Loading reference annotation.
Warning: No conditions are replicated, switching to 'blind' dispersion method
[13:33:59] Inspecting maps and determining fragment length distributions.
[13:34:17] Modeling fragment count overdispersion.
> Map Properties:
>       Normalized Map Mass: 96186.84
>       Raw Map Mass: 166235.87
>       Fragment Length Distribution: Truncated Gaussian (default)
>                     Default Mean: 200
>                  Default Std Dev: 80
> Map Properties:
>       Normalized Map Mass: 96186.84
>       Raw Map Mass: 55935.02
>       Fragment Length Distribution: Truncated Gaussian (default)
>                     Default Mean: 200
>                  Default Std Dev: 80
[13:35:31] Calculating preliminary abundance estimates
[13:35:31] Testing for differential expression and regulation in locus.
> Processed 13159 loci.                          [*************************] 100%
Performed 1533 isoform-level transcription difference tests
Performed 1532 tss-level transcription difference tests
Performed 1503 gene-level transcription difference tests
Performed 1526 CDS-level transcription difference tests
Performed 0 splicing tests
Performed 0 promoter preference tests
Performing 0 relative CDS output tests
Writing isoform-level FPKM tracking
Writing TSS group-level FPKM tracking
Writing gene-level FPKM tracking
Writing CDS-level FPKM tracking
Writing isoform-level count tracking
Writing TSS group-level count tracking
Writing gene-level count tracking
Writing CDS-level count tracking
Writing isoform-level read group tracking
Writing TSS group-level read group tracking
Writing gene-level read group tracking
Writing CDS-level read group tracking
Writing read group info
Writing run info
oystercows-Mac-Pro:Card_LowHigh_Files oystercow$
```

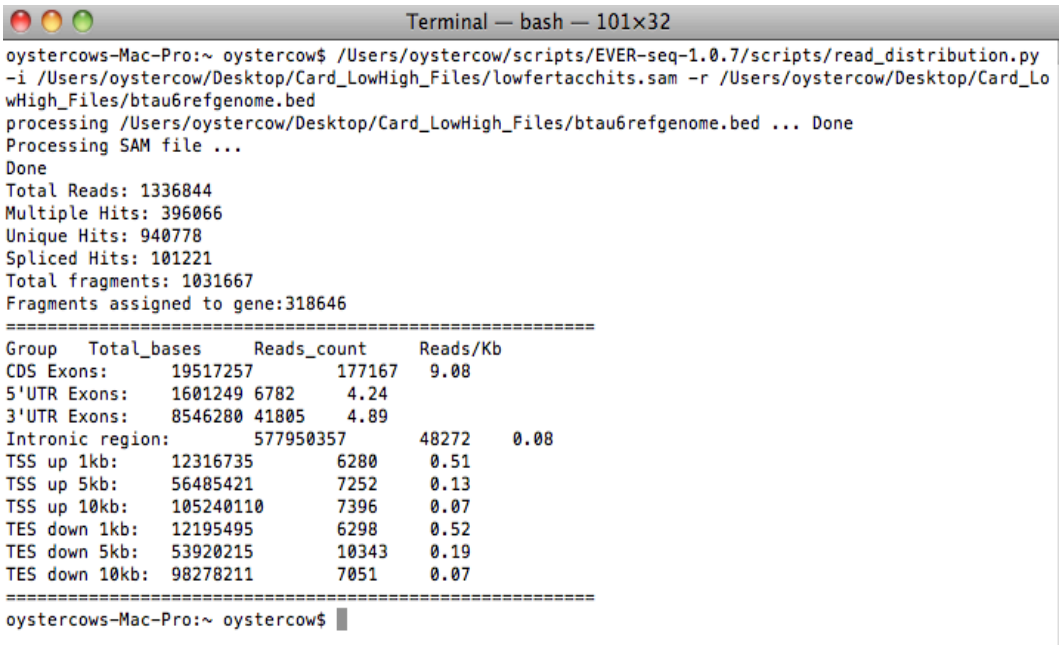
The resulting files are in the "cuffdiff_test" folder they were directed to:



**Determining % reads mapped via Everseq v.1.0.7 (read_distribution.py)**

Everseq commands are not formatted as shortcuts so unlike typing "cufflinks" as your first command line text and cufflinks running properly, you must drag and drop the location of the everseq script you want run. Go into finder > press command+shift+h (brings you to base directory for entire system) > go to scripts folder and find everseq folder (which contains a scripts folder of its own with all applications in it)

Files you will need to run read_distribution.py: alignment file (accepted hits from tophat; .sam format); bovine genome reference (btau6; .bed format)




Take the total reads output # at top of output (1,336,844 in example) and divide by total number of reads that you input into tophat via galaxy to determine % reads aligned to genome.

If reads/Kb in chart shows much higher value for 3'UTR exons it is indicative of 3' bias which is a common pitfall of RNA-Seq.

**DAVID Gene Ontology (GO) Analysis**

- Logon to the DAVID database ( http://david.abcc.ncifcrf.gov/)
- Select "Start Analysis" from top menu
- Copy transcript list of interest into the "Paste a list" box
    - Use Genbank IDs from RNA-Seq study
  - For the "Select Identifier" box, choose appropriate type of sample accession submitted
    - Use "GENBANK_ACCESSION" for RNA-Seq study
  - List Type: "Gene List"
  - Hit "Submit List"
- Select species "Bos taurus"
- On right, choose "Functional Annotation Tool"
- Click "Gene_Ontology"
  - Should automatically have "GOTERM_BP_FAT," "GOTERM_CC_FAT ," "GOTERM_MF_FAT " selected
- Click "Chart" next to each checked box
- Click "Download File" in top right corner
- Copy entire window into a .txt file and save
- Open the .txt file with excel, which should automatically insert tab delimiters
  - Can sort file according to target information

**Getting official gene symbols & long names from accession numbers**

- Logon to the DAVID database ( http://david.abcc.ncifcrf.gov/)
- Select “Start Analysis” from top menu
- Copy transcript list of interest into the “Paste a list” box
    - Use Genbank IDs from RNA-Seq study
  - For the “Select Identifier” box, choose appropriate type of sample accession submitted
    - Use “GENBANK_ACCESSION” for RNA-Seq study
  - List Type: “Gene List”
  - Hit “Submit List”
- On top menu, click “Shortcut to DAVID Tools”
- Click “Gene ID Conversion”
- Select “OFFICIAL_GENE_SYMBOL” from drop down menu
- Click “Submit to Conversion Tool”
- Click “Download File” in top right corner
- Copy entire window into a .txt file and save
- Open the .txt file with excel, which should automatically insert tab delimiters
  - Can sort file according to target information

**Pairing gene symbols & long names with known accessions**

- Open file that you’re annotating these names onto, referred to here as “FILE 1”

| | A |
|---|---|
| 1 | **Genbank ID** |
| 2 | AY646356 |
| 3 | BC108207 |
| 4 | BC102216 |

- Open file with accessions/official gene IDs/long names in it, referred to here as "FILE 2"

| | A | B | C | D |
|---|---|---|---|---|
| 1 | AY646356 | PLCZ1 | Bos taurus | phospholipase C, zeta 1 |
| 2 | AB099079 | LOC782924 | Bos taurus | similar to elongation factor 1 alpha |
| 3 | BC108207 | PRM1 | Bos taurus | protamine 1 |
| 4 | BC102713 | GTSF1 | Bos taurus | gametocyte specific factor 1 |
| 5 | BC102216 | PSMA1 | Bos taurus | proteasome (prosome, macropain) subunit, alpha type, 1 |
| 6 | AB098752 | LOC782525 | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 7 | AB098752 | EEF1G | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 8 | BC112491 | GSTM3 | Bos taurus | glutathione S-transferase mu 3 (brain) |
| 9 | BC109790 | Hmgb4 | Bos taurus | high-mobility |

- Temporarily copy all the information from "FILE 2" into blank columns to the right of the data in "FILE 1"
- In "FILE 1," add two additional columns to the right of the Genbank ID column
    - Title one "Gene ID" and the other "Long Name"

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **Genbank ID** | **Gene ID** | **Long Name** | | AY646356 | PLCZ1 | Bos taurus | phospholipase C, zeta 1 |
| 2 | AY646356 | | | | AB099079 | LOC782924 | Bos taurus | similar to elongation factor 1 alpha |
| 3 | BC108207 | | | | BC108207 | PRM1 | Bos taurus | protamine 1 |
| 4 | BC102216 | | | | BC102713 | GTSF1 | Bos taurus | gametocyte specific factor 1 |
| 5 | | | | | BC102216 | PSMA1 | Bos taurus | proteasome (prosome, macropain) subunit, alpha type, 1 |
| 6 | | | | | AB098752 | LOC782525 | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 7 | | | | | AB098752 | EEF1G | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 8 | | | | | BC112491 | GSTM3 | Bos taurus | glutathione S-transferase mu 3 (brain) |
| 9 | | | | | BC109790 | Hmgb4 | Bos taurus | high-mobility |

- One cell to the right of the first accession number (B2), type "=VLOOKUP(A2,$E$1:$H$9, 2, FALSE)"

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **Genbank ID** | **Gene ID** | **Long Name** | | AY646356 | PLCZ1 | Bos taurus | phospholipase C, zeta 1 |
| 2 | AY646356 | =VLOOKUP(A2,$E$1:$H$9, 2, FALSE) | | | | LOC782924 | Bos taurus | similar to elongation factor 1 alpha |
| 3 | BC108207 | | | | BC108207 | PRM1 | Bos taurus | protamine 1 |
| 4 | BC102216 | | | | BC102713 | GTSF1 | Bos taurus | gametocyte specific factor 1 |
| 5 | | | | | BC102216 | PSMA1 | Bos taurus | proteasome (prosome, macropain) subunit, alpha type, 1 |
| 6 | | | | | AB098752 | LOC782525 | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 7 | | | | | AB098752 | EEF1G | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 8 | | | | | BC112491 | GSTM3 | Bos taurus | glutathione S-transferase mu 3 (brain) |
| 9 | | | | | BC109790 | Hmgb4 | Bos taurus | high-mobility |

    - A1 the cell that contains what you're searching for

- $E$1:$H$100 — highlighting the entire table you're taking information from, with $ signs added to lock the entire thing in place for when you start dragging it
- 2 — column number in the table you're searching in containing the information you're looking for
- FALSE — to tell it to search for a perfect match to the accession only

- Copy this EXACT formula into cell under "Long Name", but change column 2 to column 4, in this example.

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **Genbank ID** | **Gene ID** | **Long Name** | | AY646356 | PLCZ1 | Bos taurus | phospholipase C, zeta 1 |
| 2 | AY646356 | PLCZ1 | =VLOOKUP(A2,$E$1:$H$9, 4, FALSE) | | | | Bos taurus | similar to elongation factor 1 alpha |
| 3 | BC108207 | | | | BC108207 | PRM1 | Bos taurus | protamine 1 |
| 4 | BC102216 | | | | BC102713 | GTSF1 | Bos taurus | gametocyte specific factor 1 |
| 5 | | | | | BC102216 | PSMA1 | Bos taurus | proteasome (prosome, macropain) subunit, alpha type, 1 |
| 6 | | | | | AB098752 | LOC782525 | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 7 | | | | | AB098752 | EEF1G | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 8 | | | | | BC112491 | GSTM3 | Bos taurus | glutathione S-transferase mu 3 (brain) |
| 9 | | | | | BC109790 | Hmgb4 | Bos taurus | high-mobility |

- Drag down to fill in the remainder of the accessions from column A

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **Genbank ID** | **Gene ID** | **Long Name** | | AY646356 | PLCZ1 | Bos taurus | phospholipase C, zeta 1 |
| 2 | AY646356 | PLCZ1 | phospholipase C, ze | | AB099079 | LOC782924 | Bos taurus | similar to elongation factor 1 alpha |
| 3 | BC108207 | PRM1 | protamine 1 | | BC108207 | PRM1 | Bos taurus | protamine 1 |
| 4 | BC102216 | PSMA1 | proteasome (proson | | BC102713 | GTSF1 | Bos taurus | gametocyte specific factor 1 |
| 5 | | | | | BC102216 | PSMA1 | Bos taurus | proteasome (prosome, macropain) subunit, alpha type, 1 |
| 6 | | | | | AB098752 | LOC782525 | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 7 | | | | | AB098752 | EEF1G | Bos taurus | eukaryotic translation elongation factor 1 gamma |
| 8 | | | | | BC112491 | GSTM3 | Bos taurus | glutathione S-transferase mu 3 (brain) |
| 9 | | | | | BC109790 | Hmgb4 | Bos taurus | high-mobility |

- Select all cells, then copy and "Paste Special" into same cells after clicking "Values" on the pop-up menu


- Delete reference table, leaving just the annotated original file.