2013

Are Full-Length mRNA In *Bos taurus* Spermatozoa Transferred to the Oocyte During Fertilization?

Elizabeth Jane Anderson
University of Rhode Island, ejb887@gmail.com

Recommended Citation

Anderson, Elizabeth Jane, "Are Full-Length mRNA In *Bos taurus* Spermatozoa Transferred to the Oocyte During Fertilization?" (2013). *Open Access Master's Theses.* Paper 84.
https://digitalcommons.uri.edu/theses/84

ARE FULL-LENGTH mRNA IN *Bos taurus* SPERMATOZOA TRANSFERRED TO THE OOCYTE DURING FERTILIZATION?

BY

ELIZABETH J ANDERSON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

BIOLOGICAL AND ENVIRONMENTAL SCIENCE

UNIVERSITY OF RHODE ISLAND

2013

MASTER OF SCIENCE IN BIOLOGICAL AND ENVIRONMENTAL SCIENCE

THESIS

OF

ELIZABETH J ANDERSON

APPROVED:

Thesis Committee:

Major Professor
Becky L. Sartini, Ph.D.

Katherine Petersson, Ph.D.

Steven Irvine, Ph.D.

Nasser H. Zawia, Ph.D.

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2013

ABSTRACT

This thesis focuses on the discovery of full-length mRNA transcripts in *Bos taurus* spermatozoa. The primary aim of this study is to identify and validate full-length mRNA primarily from RNA-Sequencing of bovine spermatozoa. The secondary aim is to determine if full-length spermatozoal transcripts are delivered to the oocyte at fertlilization, allowing for future studies to track their inheritance from paternal sources to the embryo. The main hypothesis of this thesis is that full-length mRNA transcripts exist within the spermatozoal transcript profile in *Bos taurus.* The secondary hypothesis is that if spermatozoal mRNA is functional after fertilization, then full-length transcripts should be present in the early stage embryo. To examine these hypotheses, this thesis is divided into three main chapters.

The first is a literature review, discussing the process of spermatogenesis, the unique properties of spermatozoal mRNAs, including some hypothesized functions of spermatozoal mRNAs. A summary of a new technique, RNA-Sequencing, will be discussed in this review as well as comparisons to previous literature techniques for identifying mRNA transcripts of interest.

The second chapter is the manuscript published in the journal Biology of Reproduction in January 2013, co-first-authored by Christopher Card. This manuscript uses the technique RNA-Seq to examine the transcript profile of nine *Bos taurus* bulls, and highlights several transcripts of interest for further study. This study found 6,166 total transcripts, and performed Gene Ontology analysis of the transcripts to categorize them into functional categories for further examination, the top most category of interest being translation.

The third chapter of this thesis is a manuscript in preparation, formatted for submission to the journal of Molecular Reproduction and Development. This manuscript evaluates twenty four target mRNA transcripts to see if they are full-length. These transcripts were identified through four main methods: their location on the Y chromosome, their high expression in the RNA-Seq data set from chapter 2, their presence in Gene Ontology categories of interest from chapter 2, and their discovery from previous literature studies. Sixteen transcripts are found to be full-length, eight are degraded, and four have alternative polyadenylation ends.

In conclusion, several full-length transcripts were found in this study, which have the potential to create functional proteins downstream in the fertilized oocyte. Several transcripts were also proved to be degraded in the mature spermatozoa. This has confirmed the need for this type of study, and elucidates new transcript targets for further research to pursue.

ACKNOWLEDGMENTS

I would like to thank Dr. Becky Sartini for her continuous support both scientifically and emotionally, throughout the process of developing my master's thesis. Additionally, the members of my lab have made my years here at URI much more enjoyable and productive: Justin Richard, Erin Aparicio, Krystle Schultz, and Jazmin Zamberlan. Particular to thanks to Christopher Card, who has been an excellent collaborator and mentor throughout my masters degree.

Without the personal support of my amazing husband, and love of my life, Edward Anderson, my commitment to these manuscripts would have been much more difficult. My parents and sister have been a continual source of support throughout this process as well.

Last, but certainly not least, I would like to thank all of the mentors that I have had the pleasure of working with during my time at URI. To my committee members, Dr. Katherine Petersson, Dr. Steven Irvine, and Dr. Ingrid Lofgren, I appreciate the time, effort, and patience involved with reading and working through my thesis. I would also like to thank the teachers I've worked with: Fred Launer, Emma Kaiser, and Marta Gomez-Chiarri.

PREFACE

The content of this thesis is subdivided into two different manuscripts. The first is a manuscript co-authored with Christopher Card and published in January 2013 in the journal “Biology of Reproduction.” Liz was a co-first author on this manuscript with a primary focus on identification of full-length transcripts in the bovine spermatozoal transcript profile. She was also responsible for all gene ontology analysis. Liz was responsible for tables 3, 5, 6, and figure 3, and collaborated with Chris Card on figures 1, 3, 4, and tables 1, 2, and 4. The writing and editing of the manuscript was shared equally between Chris and Liz.

The second manuscript here is in preparation for submission to the journal “Molecular Reproduction and Development.” This work is done entirely by Liz.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2:

CHAPTER 3:

TABLE PAGE

LIST OF FIGURES

CHAPTER 1:

FIGURE PAGE

CHAPTER 2:

FIGURE PAGE

CHAPTER 3:

FIGURE PAGE

CHAPTER 1: LITERATURE REVIEW

I. Introduction

The presence of messenger RNA (mRNA) in transcriptionally-silent spermatozoa was initially discovered in 1989 (Pessot et al. 1989). While most information is known about spermatozoa creation and spermatozoal DNA, little is known about the mRNAs contained within spermatozoa and their function. The primary focus of this thesis is the composition of the spermatozoal mRNA and the potential contributions of the spermatozoa to the oocyte at fertilization. Specifically of interest to this study is whether or not the mRNAs contained within spermatozoa are functional, or simply degraded remnants. A summary of what RNAs are present in spermatozoa is presented here, as well as several hypotheses as to the functions that these spermatozoal mRNAs might serve. Following discussion of the spermatozoal RNAs, this review will examine a new method for identifying RNA transcripts of interest, called RNA-Seq, which was used for the Chapter 2 Biology of Reproduction publication. This will include a comparsion with previous methods for identifying RNAs. This review will conclude with the hypotheses and aims of this study.

II. Gametogenesis

Both males and females create specialized reproductive cells through the process of gametogenesis. Many mechanisms are shared between spermatogenesis (the generation of male spermatozoa) and oogenesis (generation of female oocytes). Foremost, diploid stem cells in both males and females divide through mitosis and meiosis I and II into four haploid cells. In spermatogenesis, all four haploid cells mature into individual spermatozoa. In oogenesis, three of the four haploid cells form polar bodies that are discarded as the oocyte matures, until it contains only a single haploid nucleus. Meiosis also increases the genetic variation between the haploid cells through homologous recombination (Schlecht and Primig 2003; P.L.Senger 2005). Similar hormonal mechanisms are also used for spermatogenesis and oogenesis, with gonadotropin releasing hormone and lutenizing hormone acting as triggers for germ cell release in both processes (Holstein et al. 2003; P.L.Senger 2005).

Despite these basic similarities between spermatogenesis and oogenesis, the two processes achieve their reproductive goals through two very different methods. The main difference is that only males have a self-renewing system, which enables them to create much greater numbers of gametes compared to females (Holstein et al. 2003), discussed below. Unlike spermatozoa, the stem cells that create oocytes are incapable of self-renewing post-natally (Kocabas et al. 2006). For an oocyte to mature, they have to halt cell death, activate maternal transcription, unpack paternal DNA, and then kick-start embryo development (Potireddy et al. 2006). Once the spermatozoa fuse with the oocyte, the spermatozoa may be able to assist the oocyte with some of these functions. The maternal contribution produces a limited number of oocytes, but invests more energy into the production of oocytes with large biomass (Hayward and Gillooly 2011). In contrast, the paternal strategy focuses on the production of massive quantities of spermatozoa.

II. A. Spermatogenesis

Spermatogenesis is the process that spermatogonial stem cells undergo in the testis to mature into functional spermatozoa which is supported by Sertoli nurse cells (Figure 1; Petersen et al., 2006; Zhang et al., 2012),) To compensate for the need to make large quantities of spermatozoa, males use a self-renewing stem cell system to allow for production of spermatozoa from sexual maturity up until death. To this end, spermatozoa develop from a group of stem cells in the testis, which are capable of self-renewal through continuous mitosis (Holstein et al. 2003). The developing germ cells then enter meiosis I and II, now called spermatocytes, and, after a final homologous recombination event, become haploid round spermatids (Iguchi et al. 2006).

During the early round spermatid stage of the second meiosis, a significant increase of transcription and translation also occurs, depositing all the mRNAs the mature spermatozoa will maintain (Braun 2000; Eddy 2002; Holstein et al. 2003). From this point forward in spermatogenesis, the spermatids become transcriptionally silent and cease making mRNAs, although new proteins are later produced for the morphological changes necessary for mature spermatozoa formation. Several transcripts that remain in round spermatid are modified through post-transcriptional mechanisms in the 5’ and 3’ untranslated regions to hold them in an inactive state for later translation (Braun 1998). This is regulated by accessory proteins such as *Tarbp2* (Braun 2000), and is known to be used on the transcripts *PRM1* and *PRM2* (Mali et al. 1989).

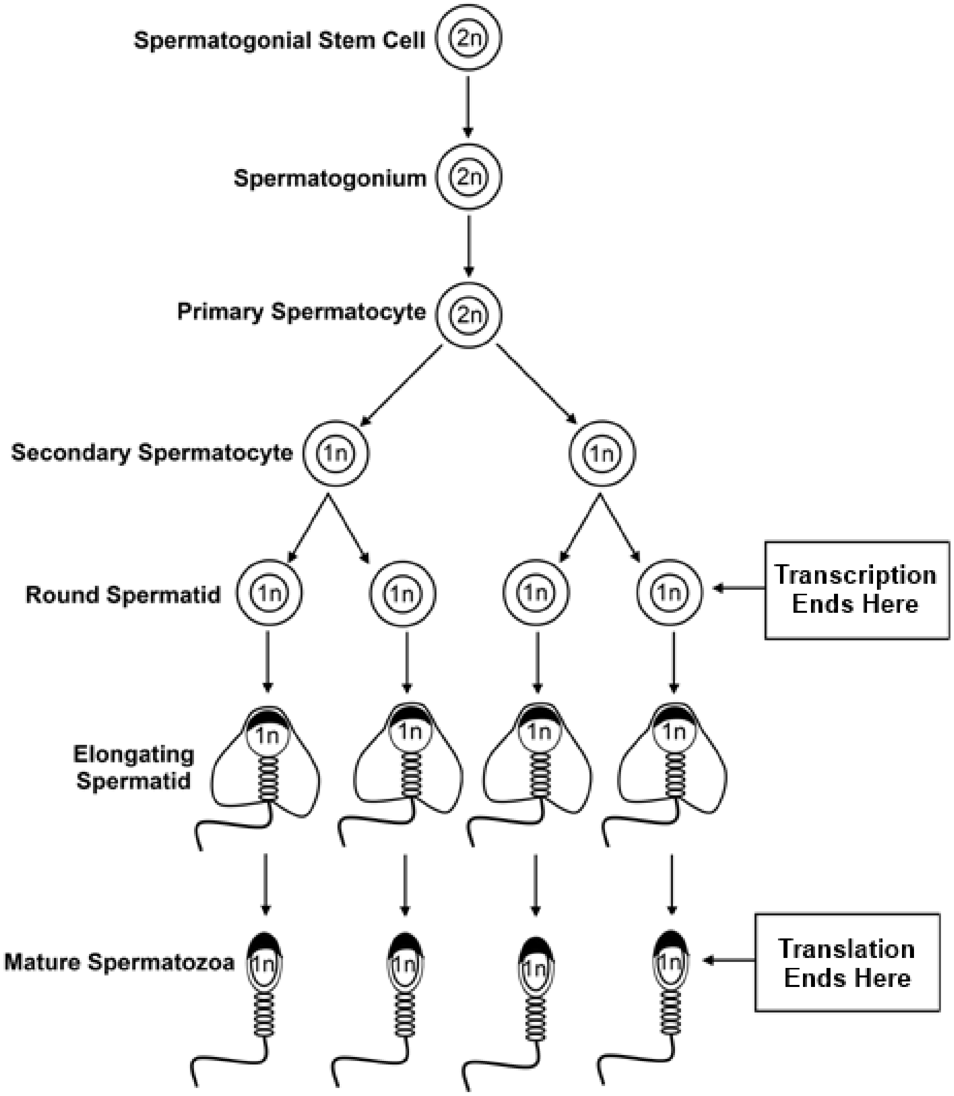


Figure 1: *Spermatogenesis*

The final differentiation of the developing spermatozoa cell occurs as the morphological changes create the spermatozoa head and tail. During this differentiation, round spermatids undergo a final process called cyto-differentiation, where the cells condense their nucleus, form an enzyme cap to help with fertilization, and develop the motor flagella needed for locomotion (Holstein et al. 1988). As the cytoplasm is lost in this elongating spermatid, the spermatozoa lose the ability to translate mRNAs into proteins, but the population of silenced mRNAs remain. Some silenced transcripts, from the elongating spermatid stage, are translated at this stage of development in order create new proteins to facilitate head and tail development. The final packing of the DNA occurs as spermatogenesis concludes, and the spermatozoa are released into the seminiferous tubule for their journey through the female reproductive tract (Dadoune et al. 2004).

Throughout the process of spermatogenesis, several unique mechanisms are utilized to regulate temporal and quality control of the spermatozoal mRNAs. Transcriptional silencing, discussed above, is one of the standard mechanisms used to regulate timing of mRNA transcript use during spermatogenesis (Braun 2000). Around the same time that selective mRNAs are being silence, the developing spermatozoa also performs DNA-repair and apoptosis to ensure quality of the mature spermatozoa (Smirnova et al. 2006). Prior to nuclear condensation ubiquitin-mediated proteolysis removes selective proteins from the cell, is responsible for replacing histones with sperm-specific protamines (Sutovsky 2003), and degrades selective mRNAs in the developing embryo (Thompson et al. 2003). The exact percentage of spermatozoa mRNAs that are degraded is unknown. When these regulatory mechanisms malfunction, the spermatozoal quality and fertility decrease (Foote 2003).

III. What is spermatozoal mRNA?

III. A. Composition

The focus of this thesis is investigating the existence of full-length transcripts in bovine spermatozoa thus providing more evidence that spermatozoa mRNAs may be functional. Spermatozoa carry not only paternal DNA, but also paternal RNAs to the oocyte at fertilization, but no previous studies have examined whether these mRNAs are full-length, a prerequisite for functionality as a protein. Finding the functionality of these transcripts is particularly interesting since spermatozoa are translationally silent at maturity, indicating that they are incapable of utilizing these mRNAs themselves (Miller and Ostermeier 2006).

Although the pool of spermatozoal RNAs is small, it contains a variety of different RNAs, including: mRNAs, rRNA, and siRNAs (Boerke et al. 2007). The function of some spermatozoal RNAs are known. For example, selective rRNA transcripts for ribosome proteins are retained or degraded through specific cleavage in the spermatozoa (Johnson et al. 2011b). The lack of complete ribosomes is one of the reasons why translation is not possible in the mature spermatozoa. Another type of RNAs found in spermatozoa is siRNAs, which have been demonstrated to be stable all the way until the activation of the embryonic genome (Kono et al. 2004). In addition, sperm-derived siRNAs have been shown to imprint on male germ cells (Reik and Walter 2001), and can modulate embryonic gene expression (Mao et al. 2002). Spermatozoal microRNAs such as microRNA-34c are also known to impact functionality in embryo development by regulating the first cellular division of the embryo (Liu et al. 2012).

This thesis focuses on the population of mRNAs in spermatozoa. The precise number of mRNAs contained per spermatozoa is unknown, but has been estimated to be approximately 3000-7000 transcripts (Ostermeier et al. 2002; Gilbert et al. 2007; Das et al. 2010), or approximately 5-10 fg per spermatozoa. This minimal amount of mRNA per spermatozoa makes detection of the mRNA difficult (Boerke et al. 2007).

The location of the spermatozoal mRNAs may provide a clue to the function. While the majority of spermatozoal mRNAs are located in the nucleus, a limited number of mRNAs are located in other areas. For example, *SP17* is expressed in the flagellar fibrous sheath (Chiriva-Internati et al. 2009). The protein made by the transcript *SMCY* is present on the surface of spermatozoa (Yao et al. 2010a). The mRNAs located on the outside of the nucleus may be more likely to have an impact on the act of fertilization or spermatozoa survival, rather than on embryogenesis, but it is not known if and where they are translated into protein.

III. B. Potential Functions of Spermatozoal mRNA

It was originally thought that the mRNAs in spermatozoa were just a remnant of the spermatogenesis process, but newer evidence supports the translation and function for these silenced mRNAs during fertilization and early embryonic development. The presence of regulatory mechanisms for mRNA timing also lends credence to the idea that the maturing spermatozoa may in fact be preserving them for later use.

There are a variety of hypothesized functions for spermatozoal mRNAs. Spermatozoa from several species have been examined using microarrays, although the most common are humans (Ostermeier et al. 2005; Chalmel et al. 2007; García-Herrero et al. 2010), rodents (Smirnova et al. 2006; Chalmel et al. 2007; Yao et al. 2010b; Liu et al. 2012), and bulls (Gilbert et al. 2007; Bissonnette et al. 2009; Vigneault et al. 2009; Feugang et al. 2010).

Some common spermatozoal transcripts among these species include *PRM1, CLU*, *PGK2*, *AKAP4*, *PRM2*, *H2AFZ*, *COX7A2*, and *DNMT1*. Several of the functions hypothesized from these mRNAs are discussed below, and largely focus on mRNA functions in fertilization, embryogenesis, or as a tool for fertility assays.

III.B.1. Spermatozoal survival in the female reproductive tract and fertilization

It is hypothesized that the translation of some spermatozoa mRNAs facilitate spermatozoa maturation in the female tract, known as capacitation, rather than impacting the embryo's development. By assisting with spermatozoa survival and the act of fertilization, these spermatozoa may influence fertility without ever impacting the embryo directly (Killian 2012). As stated previously, mature spermatozoa lack the cytoplasmic ribosomes to translate spermatozoal mRNAs but limited data suggest translation may occur in the spermatozoa mitochondria. Interestingly, there is some limited evidence that some spermatozoal nuclear-encoded mRNAs are translated in the mitochondria in the tailpiece of the spermatozoa, only occurring right before fertilization and right after the spermatozoa undergo capacitation (Gur and Breitbart 2008). Other spermatozoal proteins such as CRISP2, CCT8, and PEBP1 may assist with sperm-egg fusion during fertilization although it is not yet known if the spermatozoal transcripts themselves are also translated and contribute to this role in capacitation and fertilization (Arangasamy et al. 2011).

III.B. 2. Early embryogenesis

After the introduction of the paternal genome at fertilization, there is a lag time before the activation of the embryonic genome when the maternal genome maintains activity in the early zygote. In *Bos taurus*, the transition from maternal gene expression to zygotic gene expression occurs approximately 62 hours after fertilization (Memili and First 2000). Prior to this transition, the oocyte is responsible for mRNA and protein synthesis, helping with both maternal and paternal gene organization, and aiding the fusion of two pronuclei to form the zygotic nucleus (Potireddy et al. 2006). Due to this lag time in the activation of the embryonic genome, it has been hypothesized that spermatozoal mRNAs might function before activation of the embryonic genome (Ostermeier et al. 2002). This is supported by experiments that demonstrated that specific spermatozoal mRNAs are transferred to the oocyte at fertilization including clusterin (*CLU),* protamine 2 (*PRM2),* protamine 1 *(PRM1),* and *DDX3Y* (Ostermeier et al. 2002; Swann et al. 2006; Kempisty et al. 2008a) although a specific function for these spermatozoal transcripts has not been determined, and several are likely degraded rapidly upon entering the oocyte.

III. B.2.a. Oocyte activation

One of the only definitive functions for spermatozoal mRNA is oocyte activation at fertilization, caused by the specific transcript *PLCZ1*. *PLCZ1* triggers calcium oscillations in the oocyte. This activation occurred both when the RNA was injected into the oocyte (Saunders et al. 2002; Rogers et al. 2004), and when the protein was injected (Swann et al. 2006). The egg activation causes the oocyte to finish maturing and triggers the transition to embryonic genome control (Boerke et al. 2007). This transcript appears to be highly conserved, occurring in animals from drosophila to humans (Fischer et al. 2012).

III. B.2.b. Protamines: chromatin repackaging & regulation

During spermatogenesis, histones are replaced by transition proteins and then finally by protamines. Protamines *PRM1* and *PRM2* are responsible for replacing histones to package DNA at the end of spermatogenesis, so regulation of these specific transcripts will have a large impact on how chromatin gets packaged (Mali et al. 1989). The manner in which genes are packaged depends on how they might be used, with selective components of the DNA remaining loosely wound for access by translational machinery (Miller et al. 2005; Miller and Ostermeier 2006; Carrell and Hammoud 2010; Arangasamy et al. 2011; Ellis et al. 2011; Johnson et al. 2011b; Hamatani 2012). Spermatozoal mRNAs might assist with this organization through the use of the mRNAs as a structural component, acting to expose areas of DNA and preventing them from being wrapped up as tightly in the mature spermatozoa (Wykes et al. 1997).

Aside from repackaging the spermatozoal DNA, the protamines are responsible for temporal control of protein synthesis in developing spermatozoa. This is accomplished by creating the protamines early in spermatogenesis, and then silencing them until they are needed for nuclear condensation at the end of spermatogenesis (Braun 2000). *PRM1* is highly conserved and has been found in spermatozoa of mice (Eddy 2002; Rassoulzadegan et al. 2006; Johnson et al. 2011a), cows (Gilbert et al. 2007; Lalancette and Miller 2008; Bissonnette et al. 2009; Feugang et al. 2010; Hecht et al. 2010; Ganguly et al. 2012; Card et al. 2013), humans (Ziyyat et al. 1999; Avendaño et al. 2009; Carrell and Hammoud 2010; Johnson et al. 2011a; Hamatani 2012; Jodar et al. 2012), and marmosets (Hecht et al. 2010). Of the transcripts previously studied, protamine 1 (*PRM1)* is also one of only a few that have been investigated in multiple species, and proven to be transferred to the oocyte in *Bos taurus* (Hecht et al. 2010). Although *PRM1* mRNA is transferred from the spermatozoa to the oocyte at fertilization, a functional role is unlikely because related *PRM2* has been demonstrated to be rapidly degraded in the oocyte (Avendaño et al. 2009). The ability to detect *PRM1* in the oocyte is likely due to its critical function in spermatozoa maturation, which leaves many copies of the *PRM1* and *PRM2* transcripts untranslated in the spermatozoa after maturation.

Despite *PRM1* likely being a non-functional spermatozoal mRNA in the oocyte and embryo, it has a potential for use in a fertility assay that represents gene expression during spermatogenesis (see Fertility section below), even if found in a degraded state (Gilbert et al. 2007). Protamines are also useful candidates for tracking mRNA inheritance patterns because they are sperm-specific (Kempisty et al. 2008b; Hecht et al. 2010).

III. B. 2. c. Oocyte meiotic division

One way that spermatozoal mRNAs may impact embryogenesis is through control of the cell cycle, such as preventing or promoting the timing of cell divisions. For example, the microRNA-34c has been demonstrated to be responsible for the first cleavage of the embryo in mice. MicroRNA-34c is also known to be carried by the spermatozoa rather than the oocyte (Liu et al. 2012). This demonstrates that at a basic level, spermatozoa determine the timing of development.

Other transcripts directly regulate the cell cycle, such as *CKS2* that is involved in the MI anaphase transition in the cell cycle. Mutations present in this gene are responsible for sterility in both men and women (Donovan and Reed 2003). Cell cycle regulation is very important in early embryo development because it helps to determine when the embryo will overtake its own gene expression and cease using paternal or maternal sources of mRNAs and proteins (Hecht et al. 2009).

III. B. 2. d. Embryo imprinting

In reproduction, an eternal arms race exists between which copy of a gene will be used by the embryo: the maternal copy or the paternal copy? A large portion of this is controlled by a process called imprinting, which marks which copy of the allele to use by methylating the unused gene copy (Jenkins and Carrell 2012). This process is thought to be partially controlled by spermatozoal antisense RNAs, which act to maintain and protect the paternal copies of genes from degradation mechanisms as they enter the oocyte. The RNAs are hypothesized to work through the formation of a transcriptional silencing complex, which tags paternal DNA for imprinting in the oocyte (Miller and Ostermeier 2006).

III. B. 2. e. Epigenetic influences

There are many different ways that spermatozoal mRNAs might help the earliest stages of fertilization to establish and maintain the paternal genome (Miller et al. 2005). This touches on the idea of selfish genes: that the mRNAs of the father may be acting to further paternal interests, while maternal mRNAs compete against them.

This may also explain why certain transcripts are expressed highly and selectively in spermatozoa versus oocytes (Kleene 2005).

An expanding field of interest as to the mechanism of these changes is the field of epigenetics. Epigenetic changes are changes in the genetics or phenotype resulting from modifications other than to the underlying DNA. This is to say that epigenetics are post-processing modifications made to DNA, mRNAs, and proteins that affect areas other than the DNA coding (see Figure 2 below). Epigenetic modifications have been linked to cases of male infertility, and commonly occur through histone modifications and DNA methylation (Carrell and Hammoud 2010). RNAs are partially responsible for the control of this DNA methylation, with *Kit* mRNA knockouts resulting in heritable epigenetic changes in mice offspring, although the precise mechanism of change is unknown (Rassoulzadegan et al. 2006).

III. B. 2. f. Regulating proper embryo development

After fertilization, some transcripts from spermatozoa may still exist that can impact the developing embryo. Transcripts for the proteins HLA-E, PSGI, and PRM2 are found in spermatozoa from fertile men have been demonstrated to have an impact on embryo implantation and development, although the length of duration of the action is limited to approximately 24 hours after fertilization, at which point the first polar body is released (Avendaño et al. 2009).

Another spermatozoal transcript that is delivered to the oocyte at fertilization is *DDX3Y.* DDX3Y is a DEAD-box RNA helicase, and is one of 33 total genes found on the Y chromosome (Marshall Graves 2000), which also has an X chromosome homolog (Vong et al. 2006). The X homolog of *DDX3Y* also shares in similar functions, but it has been demonstrated that when the X-encoded isoform is mutated, that *DDX3Y* is capable of rescuing some of the functionality for the embryo (Sekiguchi et al. 2004). *DDX3Y* is located in a known azoospermia region on spermatozoa, a region known for causing infertility in the spermatozoa when damaged (Session et al. 2001; Vong et al. 2006). In mature spermatozoa, *DDX3Y* is localized to the post-acrosomal region of sperm, and injection of as-*DDX3Y* into the oocyte nucleus has been correlated with decreased embryo development rates (Yao et al. 2010b).

III. B. 3. Spermatozoal mRNA use as a fertility assay

Assessing fertility of spermatozoa has been limited to tests of morphology, motility, and concentration (Lalancette and Miller 2008; Feugang et al. 2010). These measurement of spermatozoa function are incapable of predicting whether the spermatozoa possess the ability to survive the female reproductive tract, fertilize an egg, or even produce a viable embryo (Sutovsky 2003; Bissonnette et al. 2009). Due to this problem, finding more quantitative measures of fertility is highly desirable. By identifying spermatozoal mRNAs, we are not only provided with information about their potential function, but may also find uses for them as diagnostic measures of fertility for individuals. The mRNAs produced by an individual are consistent from ejaculate to ejaculate, and vary between individuals, making them a stable target for a fertility assay (Das et al. 2010). In addition, the amount of individual transcripts *CRISP2*, *CCT8*, and *PEBP1* has been correlated with relative fertility of individual bulls (Arangasamy et al. 2011). The efficiency with which PRM1 proteins package

DNA has also been shown to impact fertility, and is differentially expressed in fertile versus infertile bulls (Carrell and Hammoud 2010; Feugang et al. 2010; Ganguly et al. 2012; Jodar et al. 2012). The spermatozoal mRNAs may therefore be used as a snapshot of fertility, providing better diagnostic tools to examine infertility (Dadoune et al. 2004; Ostermeier et al. 2005; Lalancette et al. 2008).

As previously discussed, many of the mRNAs in spermatozoa are degraded or will be degraded rapidly in the oocyte. However, even incomplete mRNA transcripts may be used as a snapshot of fertility, and also as a predictor of infertility (Dadoune et al. 2004; Ostermeier et al. 2005; Lalancette et al. 2008). This is because the mRNAs are deposited in spermatozoa at a single timepoint, so they reflect how efficiently the spermatozoa were developing at that point. Many of the mutations in spermatozoa that would render them incapable of fertilization occur around this time (Miller and Ostermeier 2006). A few studies have indicated that degraded mRNAs are specific to an individual; they may serve as a useful fertility assay to measure relative fertility (Suri 2004; Feugang et al. 2010; Park et al. 2012). This is particularly useful for mutations that alter the genetics and transcript profiles without altering the general phenotype or motility (Foote 2003). For example, the calmegin gene is required for allowing spermatozoa to progress past the uterotubal junction in the female reproductive tract, making knockouts of the calmegin gene infertile. However, these knockouts fail to express any morphological or motility abnormalities, so they would not be detected using standard fertility assays (Yamagata et al. 2002).

IV. Full-Length mRNA Transcripts

While some specific spermatozoal transcripts have been identified, functionally depends on the presence of intact, full-length transcripts that can be translated into proteins either in the spermatozoa or in the early stage embryo. Spermatozoal transcript profile methods, primarily microarray studies, have so far only identified the presence of transcripts but have not designated between whether the mRNA transcripts are full-length or degraded remnants of spermatogenesis. If spermatozoal mRNA is functional after fertilization, then full-length transcripts should exist within the spermatozoa transcript profile.

Figure 2: *Diagram of a full-length mRNA transcript. UTR= Untranslated Region*

A full-length transcript requires several parts to be translated into a functional protein (Figure 2). The coding region provides the nucleotide sequences to be translated into protein while much of the protein regulation is done by the UTRs, the 5’ cap and 3’ poly(A) tail. At the 5’ end, a modified guanine cap is responsible for recognition of the mRNAs by ribosomes, providing the attachment for their translation and determining the efficiency with which mRNAs will be translated (Parker and Sheth 2007). The cap also influences the survival of the mRNA, by stabilizing it against degradation (Gallie 1991).

The untranslated regions (UTRs) flanking both sides of the coding region contain regulatory elements that control gene expression including post-transcriptional gene expression in the 3’UTR. Regulatory elements in the UTR impact degradation by setting the location of the stop codon that completes the mRNA, by controlling the tertiary structure of the protein being formed, and by regulating decay of the mRNA (Isken and Maquat 2007). The 3' UTR can also be modified or lengthened, a process called alternative polyadenylation (Danckwardt et al. 2008). These modifications can lead to several different transcript isoforms , which may be expressed simultaneously within a given individual's transcriptome (Kleene 2005). At the end of the mRNA, the 3' end poly(A) tail also contributes to the rate of mRNA degradation, and is the site of many of the epigenetic changes that occur to spermatozoal mRNAs (Rassoulzadegan et al. 2006). The new method of global transcriptome discovery, RNA-Seq, will facilitate the identification of full-length transcripts, and can detect many of the modifications to known transcripts discussed above.

V. RNA-Seq: a global transcript discovery method

The identification of full-length transcripts in spermatozoa is essential to identify candidate transcripts for further analysis of function. As mentioned previously, studies to date have use methods that only detect a small portion of a transcript but do not reveal if full-length transcripts are present. For example, microarrays are a hybridization-based method that can detect the presence and amount of transcript but is limited to known transcripts. An incomplete transcriptome for sperm, oocytes, and embryos has been previously reported by microarrays in humans (Ostermeier et al. 2005; Chalmel et al. 2007; García-Herrero et al. 2010), rodents (Potireddy et al. 2006; Smirnova et al. 2006; Chalmel et al. 2007; Yao et al. 2010b; Liu et al. 2012), and bulls (Misirlioglu et al. 2006; Gilbert et al. 2007; Bissonnette et al. 2009; Thelie et al. 2009; Vigneault et al. 2009; Feugang et al. 2010).

RNA-sequencing (RNA-Seq) is a newer high-throughput method that sequences an entire transcriptome *in situ*, which removes the constraints of only searching for known transcripts and can sequence the entire length of a transcript. This allows for the discovery of novel transcripts (Wang et al. 2009; Mamo et al. 2011; Driver et al. 2012). Additionally, RNA-Seq provides quantitative expression levels by assessing the number of sequencing reads mapping to a specific transcript. In RNA-Seq, mRNA is fragmented, sequenced *in situ* into reads then aligns reads to a reference genome (Figure 3), or assembles them *de novo*.

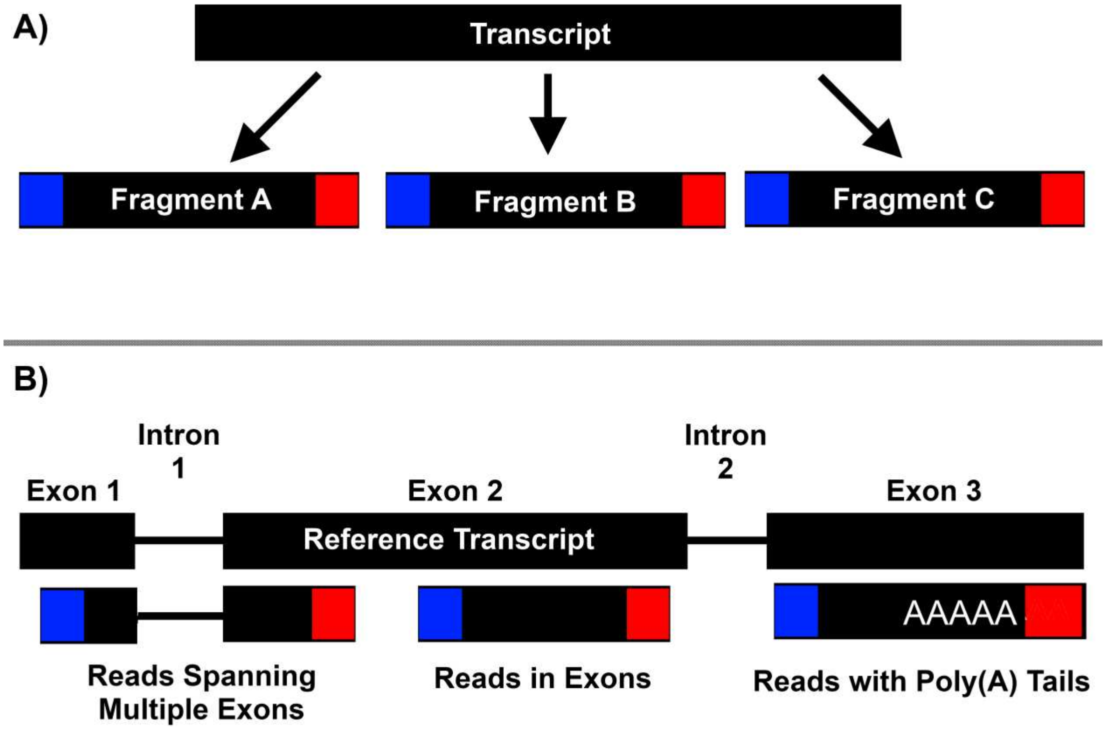


Figure 3: *RNA-Seq Methodology A) Transcripts are fragmented into many pieces, with adaptors attached to both ends, adaptors shown in blue and red. B) Example of RNA-Seq read alignment to a reference genome after sequencing. Exons depicted are categorized into three major categories, shown at bottom.*

Using a reference genome allows for improved annotation of the transcripts, and comparison of expression levels between tissues (Wang et al. 2009). Additionally, RNA-Seq is capable of identifying transcripts with novel exons, unusual isoforms, and alternatively polyadenylated ends (Cui et al. 2010; Driver et al. 2012). This allows for a thorough survey of the RNA population, although PCR verification is required to confirm all novel discoveries. Only a few studies have used RNA-Seq for oocytes, embryos, and testis (Ameur et al. 2011; Esteve-Codina et al. 2011), but the manuscript presented in this thesis (Card and Anderson et al., 2013) is the first to report the spermatozoal transcript profile using RNA-Seq for any species.

VI. Hypotheses

1. Full-length transcripts exist within the spermatozoal transcript profile.
2. If spermatozoal mRNA is functional after fertilization, then full-length transcripts will be found in the early embryo.

VII. Aims

The primary aim of this study is to identify and validate full-length mRNA primarily from RNA-Sequencing of bovine spermatozoa. The secondary aim is to determine if full-length spermatozoal transcripts are delivered to the oocyte at fertlilization, allowing for future studies to track their inheritance from paternal sources to the embryo.

VIII. References

Ameur, A., A. Zaghlool, J. Halvardson, A. Wetterbom, U. Gyllensten, L. Cavelier, and L. Feuk. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nature structural & molecular biology 18:1435–40.

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.

Avendaño, C., A. Franchi, E. Jones, and S. Oehninger. 2009. Pregnancy-specific {beta}-1-glycoprotein 1 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. Human reproduction (Oxford, England) 24:270–7.

Bissonnette, N., J.-P. Lévesque-Sergerie, C. Thibault, and G. Boissonneault. 2009. Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality. Reproduction (Cambridge, England) 138:65–80.

Boerke, a, S. J. Dieleman, and B. M. Gadella. 2007. A possible role for sperm RNA in early embryo development. Theriogenology 68 Suppl 1:S147–55.

Braun, R. E. 1998. Post-transcriptional control of gene expression during spermatogenesis. Seminars in cell & developmental biology 9:483–9.

Braun, R. E. 2000. Temporal control of protein synthesis during spermatogenesis. International journal of andrology 23 Suppl 2:92–4.

Card, C., E. J. Anderson, S. Zamberlan, K.-B. E. Krieger, M. Kaproth, and B. L. Sartini. 2013. Cryopreserved Bovine Spermatozoal Transcript Profile as Revealed by High-Throughput Ribonucleic Acid Sequencing. Biology of Reproduction .

Carrell, D. T., and S. S. Hammoud. 2010. The human sperm epigenome and its potential role in embryonic development. Molecular human reproduction 16:37–47.

Chalmel, F., A. D. Rolland, C. Niederhauser-Wiederkehr, S. S. W. Chung, P. Demougin, A. Gattiker, J. Moore, et al. 2007. The conserved transcriptome in human and rodent male gametogenesis. Proceedings of the National Academy of Sciences of the United States of America 104:8346–51.

Chiriva-Internati, M., N. Gagliano, E. Donetti, F. Costa, F. Grizzi, B. Franceschini, E. Albani, et al. 2009. Sperm protein 17 is expressed in the sperm fibrous sheath. Journal of translational medicine 7:61.

Cui, P., Q. Lin, F. Ding, C. Xin, W. Gong, L. Zhang, J. Geng, et al. 2010. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. Genomics 96:259–65.

Dadoune, J.-P., J.-P. Siffroi, and M.-F. Alfonsi. 2004. Transcription in haploid male germ cells. International Review Of Cytology 237:1–56.

Danckwardt, S., M. W. Hentze, and A. E. Kulozik. 2008. 3’ end mRNA processing: molecular mechanisms and implications for health and disease. The EMBO journal 27:482–98.

Das, P. J., N. Paria, A. Gustafson-Seabury, M. Vishnoi, S. P. Chaki, C. C. Love, D. D. Varner, et al. 2010. Total RNA isolation from stallion sperm and testis biopsies. Theriogenology 74:1099–1106, 1106e1–2.

Donovan, P. J., and S. I. Reed. 2003. Germline Exclusion of Cks1 in the Mouse Reveals a Metaphase I Role for Cks Proteins in Male and Female Meiosis nd io sc ie nc ot fo r d is t rib ut n . 275–276.

Driver, A. M., F. Penagaricano, W. Huang, K. R. Ahmad, K. S. Hackbart, M. C. Witlbank, and H. Khatib. 2012. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar in vivo- and in vitro-derived bovine blastocysts. BMC genomics 13:118.

Eddy, E. M. 2002. Male germ cell gene expression. Recent progress in hormone research 57:103–28.

Ellis, P. J. I., Y. Yu, and S. Zhang. 2011. for offspring sex-specific antigens in sperm.

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

Feugang, J. M., N. Rodriguez-Osorio, a Kaya, H. Wang, G. Page, G. C. Ostermeier, E. K. Topper, et al. 2010. Transcriptome analysis of bull spermatozoa: implications for male fertility. Reproductive biomedicine online 21:312–24.

Fischer, B. E., E. Wasbrough, L. a Meadows, O. Randlet, S. Dorus, T. L. Karr, and S. Russell. 2012. Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proceedings. Biological sciences / The Royal Society 279:2636–44.

Foote, R. H. 2003. Fertility estimation: a review of past experience and future prospects. Animal reproduction science 75:119–39.

Gallie, D. R. 1991. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes & Development 5:2108–2116.

Ganguly, I., G. K. Gaur, S. Kumar, D. K. Mandal, M. Kumar, U. Singh, S. Kumar, et al. 2012. Differential expression of protamine 1 and 2 genes in mature spermatozoa of normal and motility impaired semen producing crossbred Frieswal (HF×Sahiwal) bulls. Research in veterinary science 1.

García-Herrero, S., M. Meseguer, J. A. Martínez-Conejero, J. Remohí, A. Pellicer, and N. Garrido. 2010. The transcriptome of spermatozoa used in homologous intrauterine insemination varies considerably between samples that achieve pregnancy and those that do not. Fertility and sterility 94:1360–73.

Gilbert, I., N. Bissonnette, G. Boissonneault, M. Vallée, and C. Robert. 2007. A molecular analysis of the population of mRNA in bovine spermatozoa. Reproduction (Cambridge, England) 133:1073–86.

Gur, Y., and H. Breitbart. 2008. Protein synthesis in sperm: dialog between mitochondria and cytoplasm. Molecular and cellular endocrinology 282:45–55.

Hamatani, T. 2012. Human spermatozoal RNAs. Fertility and sterility 97:275–81.

Hayward, A., and J. F. Gillooly. 2011. The cost of sex: quantifying energetic investment in gamete production by males and females. PloS one 6:e16557.

Hecht, N., R. Behr, A. Hild, M. Bergmann, W. Weidner, and K. Steger. 2009. The common marmoset (Callithrix jacchus) as a model for histone and protamine expression during human spermatogenesis. Human reproduction (Oxford, England) 24:536–45.

Hecht, N., M. C. O. Cavalcanti, P. Nayudu, R. Behr, M. Reichenbach, W. Weidner, and K. Steger. 2010. Protamine-1 represents a sperm specific gene transcript: a study in Callithrix jacchus and Bos taurus. Andrologia 43:1–7.

Holstein, A. F., E. C. Roosen-Runge, and C. Schirren. 1988. Illustrated pathology of human spermatogenesis. Grosse, Berlin.

Holstein, A.-F., W. Schulze, and M. Davidoff. 2003. Understanding spermatogenesis is a prerequisite for treatment. Reproductive biology and endocrinology : RB&E 1:107.

Iguchi, N., J. W. Tobias, and N. B. Hecht. 2006. Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. Proceedings of the National Academy of Sciences of the United States of America 103:7712–7.

Isken, O., and L. E. Maquat. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes & development 21:1833–56.

Jenkins, T. G., and D. T. Carrell. 2012. The sperm epigenome and potential implications for the developing embryo. Reproduction (Cambridge, England) 143:727–34.

Jodar, M., S. Kalko, J. Castillo, J. L. Ballescà, and R. Oliva. 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. Human reproduction (Oxford, England) 27:1431–8.

Johnson, G. D., C. Lalancette, A. K. Linnemann, F. Leduc, G. Boissonneault, and S. a Krawetz. 2011a. The sperm nucleus: chromatin, RNA, and the nuclear matrix. Reproduction (Cambridge, England) 141:21–36.

Johnson, G. D., E. Sendler, C. Lalancette, R. Hauser, M. P. Diamond, and S. a Krawetz. 2011b. Cleavage of rRNA ensures translational cessation in sperm at fertilization. Molecular human reproduction 17:721–6.

Kempisty, B., P. Antosik, D. Bukowska, M. Jackowska, M. Lianeri, J. M. J. M. Jaśkowski, P. P. Jagodzi'nski, et al. 2008a. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reproduction, fertility, and development 20:513–8.

Kempisty, B., P. Antosik, D. Bukowska, M. Jackowska, M. Lianeri, J. M. Jaśkowski, and P. P. Jagodzi'nski. 2008b. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reproduction, Fertility and Development 20:513–518.

Killian, G. 2012. Fertility-associated proteins in male and female reproductive fluids of cattle 703–712.

Kleene, K. C. 2005. Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. Developmental biology 277:16–26.

Kocabas, A. M., J. Crosby, P. J. Ross, H. H. Otu, Z. Beyhan, H. Can, W.-L. Tam, et al. 2006. The transcriptome of human oocytes. Proceedings of the National Academy of Sciences of the United States of America 103:14027–32.

Kono, T., Y. Obata, Q. Wu, and K. Niwa. 2004. Birth of parthenogenetic mice that can develop to adulthood 428:3–7.

Lalancette, C., and D. Miller. 2008. Paternal contributions: new functional insights for spermatozoal RNA. Journal of cellular … 104:1570–1579.

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. Biology of reproduction 78:618–35.

Lee, C., and N. Kikyo. 2012. Strategies to identify long noncoding RNAs involved in gene regulation. Cell & bioscience 2:37.

Liu, W., R. T. K. Pang, P. C. N. Chiu, B. P. C. Wong, K. Lao, and K. Lee. 2012. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. PNAS 109:490–494.

Mali, P., A. Kaipia, M. Kangasniemi, J. Toppari, M. Sandberg, N. B. Hecht, and M. Parvinen. 1989. Stage-specific expression of nucleoprotein mRNAs during rat and mouse spermiogenesis. Reproduction, Fertility and Development 1:369–382.

Mamo, S., J. P. Mehta, P. McGettigan, T. Fair, T. E. Spencer, F. W. Bazer, and P. Lonergan. 2011. RNA sequencing reveals novel gene clusters in bovine conceptuses associated with maternal recognition of pregnancy and implantation. Biology of reproduction 85:1143–51.

Mao, B., W. Wu, G. Davidson, J. Marhold, M. Li, B. M. Mechler, H. Delius, et al. 2002. Kremen proteins are Dickkopf receptors that regulate Wnt/beta-catenin signalling. Nature 417:664–7.

Marshall Graves, J. a. 2000. Human Y chromosome, sex determination, and spermatogenesis- a feminist view. Biology of reproduction 63:667–76.

Memili, E., and N. L. First. 2000. Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. Zygote (Cambridge, England) 8:87–96.

Miller, D., and G. C. Ostermeier. 2006. Spermatozoal RNA: Why is it there and what does it do? Gynécologie, obstétrique & fertilité 34:840–6.

Miller, D., G. C. Ostermeier, and S. a Krawetz. 2005. The controversy, potential and roles of spermatozoal RNA. Trends in molecular medicine 11:156–63.

Misirlioglu, M., G. P. Page, H. Sagirkaya, a Kaya, J. J. Parrish, N. L. First, and E. Memili. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. Proceedings of the National Academy of Sciences of the United States of America 103:18905–10.

Ostermeier, G. C., D. J. Dix, D. Miller, P. Khatri, and S. A. Krawetz. 2002. Mechanisms of disease Spermatozoal RNA profiles of normal fertile men 772–777.

Ostermeier, G. C., R. J. Goodrich, J. S. Moldenhauer, M. P. Diamond, and S. a Krawetz. 2005. A suite of novel human spermatozoal RNAs. Journal of andrology 26:70–4.

P.L.Senger. 2005. Pathways to Pregnancy and Parturition (2nd ed., pp. 196–197). Current Conceptions, Inc., Pullman, WA.

Park, Y.-J., W.-S. Kwon, S.-A. Oh, and M.-G. Pang. 2012. Fertility-related proteomic profiling bull spermatozoa separated by percoll. Journal of proteome research 11:4162–8.

Parker, R., and U. Sheth. 2007. P bodies and the control of mRNA translation and degradation. Molecular cell 25:635–46.

Pessot, C. A., M. Brito, J. Figueroa, I. I. Concha, A. Yañez, and L. O. Burzio. 1989. Presence of RNA in the sperm nucleus. Biochemical and Biophysical Research Communications 158:272–278.

Petersen, C., and O. Soder. 2006. The sertoli cell--a hormonal target and “super” nurse for germ cells that determines testicular size. Hormone research 66:153–61.

Potireddy, S., R. Vassena, B. G. Patel, and K. E. Latham. 2006. Analysis of polysomal mRNA populations of mouse oocytes and zygotes: dynamic changes in maternal mRNA utilization and function. Developmental biology 298:155–66.

Rassoulzadegan, M., V. Grandjean, P. Gounon, S. Vincent, I. Gillot, and F. Cuzin. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 441:469–74.

Reik, W., and J. Walter. 2001. Genomic imprinting: parental influence on the genome. Nature reviews. Genetics 2:21–32.

Rogers, N. T., E. Hobson, S. Pickering, F. a Lai, P. Braude, and K. Swann. 2004. Phospholipase Czeta causes Ca2+ oscillations and parthenogenetic activation of human oocytes. Reproduction (Cambridge, England) 128:697–702.

Saunders, C. M., M. G. Larman, J. Parrington, L. J. Cox, J. Royse, L. M. Blayney, K. Swann, et al. 2002. PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development. Development (Cambridge, England) 129:3533–44.

Schlecht, U., and M. Primig. 2003. Review Mining meiosis and gametogenesis with DNA microarrays 447–456.

Sekiguchi, T., H. Iida, J. Fukumura, and T. Nishimoto. 2004. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. Experimental cell research 300:213–22.

Session, D. R., G. S. Lee, and D. J. Wolgemuth. 2001. Characterization of D1Pas1, a mouse autosomal homologue of the human AZFa region DBY, as a nuclear protein in spermatogenic cells. Fertility and sterility 76:804–11.

Smirnova, N. a, P. J. Romanienko, P. P. Khil, and R. D. Camerini-Otero. 2006. Gene expression profiles of Spo11-/- mouse testes with spermatocytes arrested in meiotic prophase I. Reproduction (Cambridge, England) 132:67–77.

Suri, A. 2004. Sperm specific proteins-potential candidate molecules for fertility control. Reproductive biology and endocrinology□: RB&E 2:10.

Sutovsky, P. 2003. Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: killing three birds with one stone. Microscopy research and technique 61:88–102.

Swann, K., C. M. Saunders, N. T. Rogers, and F. a Lai. 2006. PLCzeta(zeta): a sperm protein that triggers Ca2+ oscillations and egg activation in mammals. Seminars in cell & developmental biology 17:264–73.

Thelie, A., P. Papillier, C. Perreau, S. Uzbekova, C. Hennequet-Antier, and R. Dalbies-Tran. 2009. Regulation of bovine oocyte-specific transcripts during in vitro oocyte maturation and after maternal-embryonic transition analyzed using a transcriptomic approach. Molecular reproduction and development 76:773–82.

Thompson, W. E., J. Ramalho-Santos, and P. Sutovsky. 2003. Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. Biology of reproduction 69:254–60.

Vigneault, C., C. Gravel, M. Vallée, S. McGraw, and M.-A. Sirard. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. Reproduction (Cambridge, England) 137:245–57.

Vong, Q. P., Y. Li, Y.-F. C. Lau, M. Dym, O. M. Rennert, and W.-Y. Chan. 2006. Structural characterization and expression studies of Dby and its homologs in the mouse. Journal of andrology 27:653–61.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics 10:57–63.

Wykes, S. M., D. W. Visscher, and S. a Krawetz. 1997. Haploid transcripts persist in mature human spermatozoa. Molecular human reproduction 3:15–9.

Yamagata, K., T. Nakanishi, M. Ikawa, R. Yamaguchi, S. B. Moss, and M. Okabe. 2002. Sperm from the Calmegin-Deficient Mouse Have Normal Abilities for Binding and Fusion to the Egg Plasma Membrane. Developmental Biology 250:348–357.

Yao, C., Z. Wang, Y. Zhou, W. Xu, Q. Li, D. Ma, L. Wang, et al. 2010a. A study of Y chromosome gene mRNA in human ejaculated spermatozoa. Molecular reproduction and development 77:158–66.

Yao, C.-J. C. J., W.-J. W. J. Xu, X. L. X.-L. Gong, Y. Zhou, Z. Q. Z.-Q. Z. Q. Yan, Z.-J. Z. J. Zhu, Z. X. Z.-X. Wang, et al. 2010b. The role of Dby mRNA in early development of male mouse zygotes. Asian Journal of Andrology 12:567–577.

Zhang, Y., L. Zhong, B. Xu, Y. Yang, R. Ban, J. Zhu, H. J. Cooke, et al. 2012. SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. Nucleic Acids Research 1–8.

Ziyyat, a, B. Lassalle, J. Testart, P. Briot, E. Amar, C. Finaz, and a Lefèvre. 1999. Flow cytometry isolation and reverse transcriptase-polymerase chain reaction characterization of human round spermatids in infertile patients. Human reproduction (Oxford, England) 14:379–87.

CHAPTER 2: BIOLOGY OF REPRODUCTION MANUSCRIPT

This manuscript was co-authored with Christopher Card and published in January 2013 in the journal "Biology of Reproduction." Liz was a co-first author on this manuscript with a primary focus on identification of full-length transcripts in the bovine sperm transcript profile. She was also responsible for all gene ontology analysis. Liz was responsible for tables 3, 5, 6, and figure 3, and collaborated with Chris Card on figures 1, 4, and tables 1, 2, and 4. The writing and editing of the manuscript was shared equally between Chris and Liz.

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 02881, 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 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 Becky L. Sartini. This research is based in part upon work conducted using the Rhode Island Genomics and Sequencing Center that 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 the full-length amplification or 5’ and 3’ exons were confirmed for some 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 top 368 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 functional and fertility 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 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. Individual transcripts are stably regulated within and between individual males and perturbation of the ubiquitin-proteosome pathway during spermatogenesis can be detected in spermatozoal RNA making this a promising area for male fertility assay development [20-23].

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 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 from Holstein bulls with conception rate (CR) scores ranging from -2.9 to 3.5 were obtained from Genex Cooperative Inc. (Shawano, WI). Spermatozoa from nine bulls (-2.9 to 3.5 CR) was used to generate the amplified cDNA pool for RNA-Seq, for qPCR validation and PCR amplification of the 5' and 3' exons. Spermatozoal RNA from nine additional bulls was also converted to cDNA and amplified for PCR amplification of transcripts in individual bulls. Finally, a separate pool of spermatozoa RNA from three different bulls was not cDNA amplified 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). Spermatozoal 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 4°C. 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.

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). This method maintains gene representation in the original RNA pool. 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. Following ds-cDNA conversion and amplification, 5 µg of spermatozoal ds-cDNA was submitted for Illumina sequencing and the remaining spermatozoal ds-cDNA was used for qPCR validation post-sequencing. To verify the quality of ds-cDNA, an aliquot of the sample was run on the Agilent Bioanalyzer 2100 (Agilent Technologies; Santa Clara, CA).

PCR

PCR reactions validated the lack of genomic DNA and somatic cell RNA in the spermatozoal RNA as well as full-length and 5’ and 3’ exon amplification in the spermatozoal transcript profile. All PCR reactions were conducted with intron spanning primers (Table 1). Most 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). For amplification of full-length transcripts from unamplified ds-cDNA, spermatozoal RNA (1 µg) was reverse transcribed using the Superscript III Reverse Transcriptase Kit (Invitrogen, Carlsbad, CA). 1 µL OligoDT primers (IDT, Coralville, IA), 1 µL dNTPs, 2 µg mRNA from the isolation and 1 µL water are combined, incubated at 65^{O} C for 5 minutes, then chilled on ice for a minute. Following the chill, 4 µL 5X First Strand Buffer, 1 µL 0.1 mM DTT, 1 µL RNaseOut, and 1 µL Superscript III were added to the tube, incubated at 50^{O} C for 45 minutes, then 70^{O} C for 15 minutes. Then RNase H was added and the sample was incubated at 37^{O} C for 20 minutes. This procedure converts the mRNA to cDNA for use in PCR amplification. 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) and run with standard PCR conditions (94°C for 5 min, 35 cycles of 94°C for 30 sec), primer dependent annealing temperature for 30 sec then 72°C for 2 min followed by 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 then 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.

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]. Reads were converted with FASTQ groomer and then aligned to the bovine genome (Btau6/UMD_3.1) using Tophat [31]. 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 [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). 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 1B). Genomic DNA was also not detected in the isolated bovine spermatozoal RNA compared to a sample spiked with genomic DNA (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 GSTF1 was confirmed by PCR amplification of the intact transcript from the first to last exon in unamplified cDNA (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’ exons, potentially indicating that these transcripts are also full-length (Table 3).

A number of additional full-length bovine spermatozoal transcripts have not been previously reported in spermatozoal RNA, including *HMGB4*, *PSMA6*, *GTSF1*, and *CKS2* (Table 4). Variation in the amount of select transcripts among bulls was demonstrated in an independent population of nine bulls (not included in RNA-Seq population; Figure 4).

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 spermatozoa 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 the 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 spermatozoal 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 abundance values reported probably represent a sum of the full-length and fragmented exons for each transcript and the levels of intact transcripts is then 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.

While 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, biological replicates were not conducted and individual bulls were pooled for RNA-Seq and for some validation although individuals with a wide range of known fertility scores were used. The pool of mRNA from our 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 the presence of specific transcripts did vary in an independent population of 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] 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.

[22] 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.

[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.

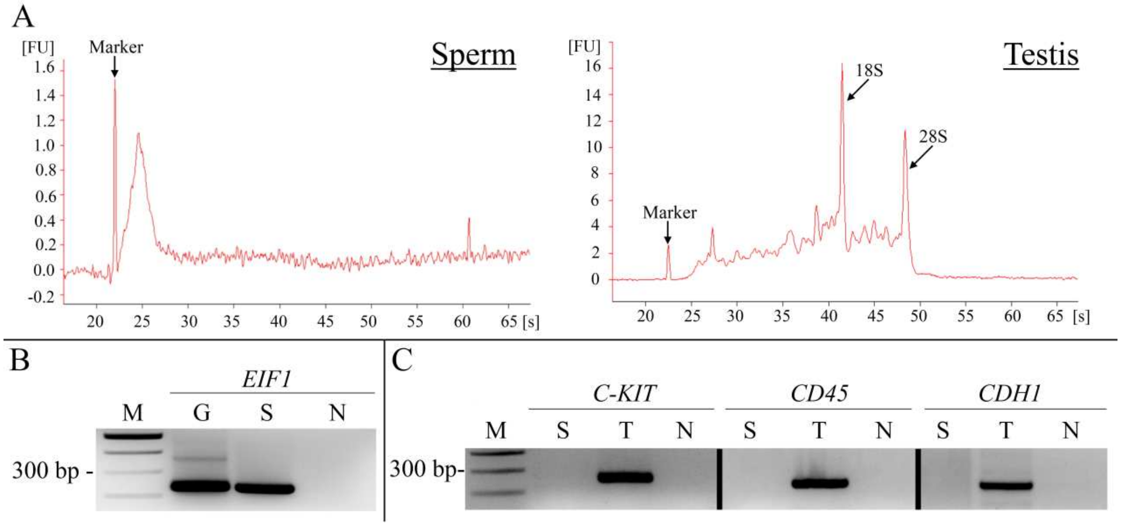


Figure 2. Correlation of qPCR transcript copy number and RNA-Seq FPKM based on nine transcripts. Axes are base 10 log scale.

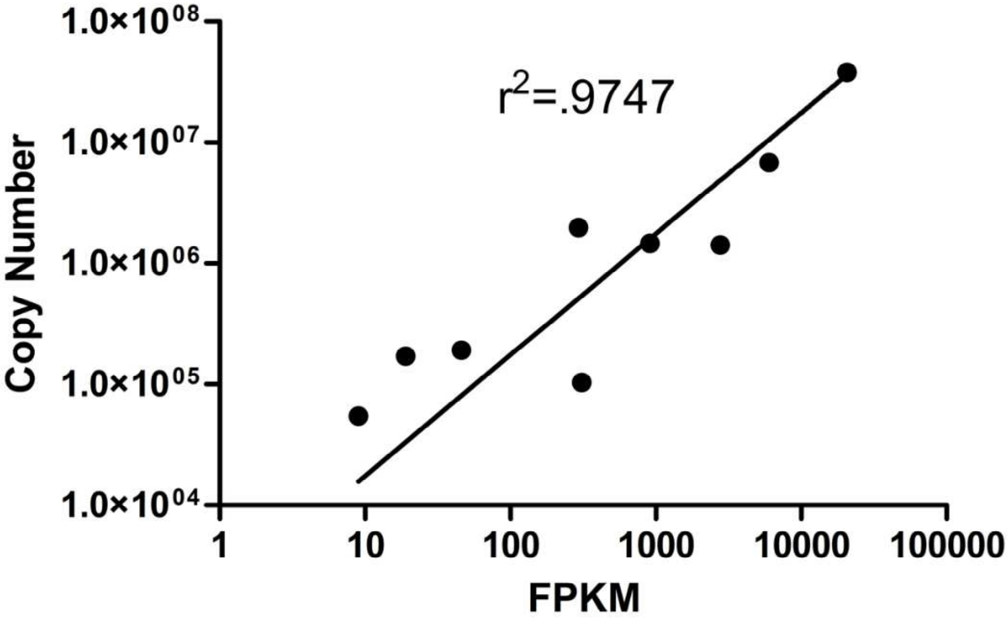


Figure 3. PCR amplification of (A) the 5’ and 3’ ends of *DDX3Y*, *PLCZ1*, *CRISP2* and *GSTM3* in amplified ds-cDNA. For 5’ end primers, all primers begin in the first exon, and for 3’ end primers, all primers end in the last exon. All primer sets are intron-spanning. N = negative control that did not 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)

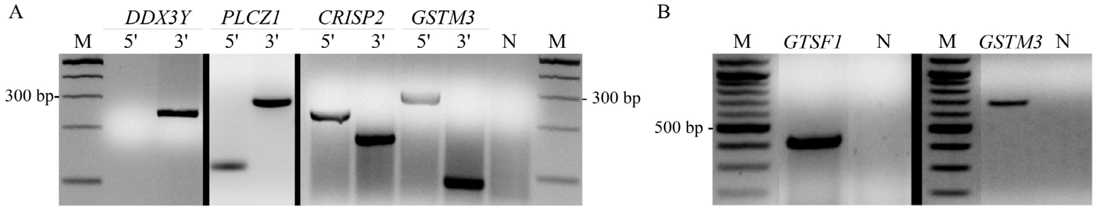
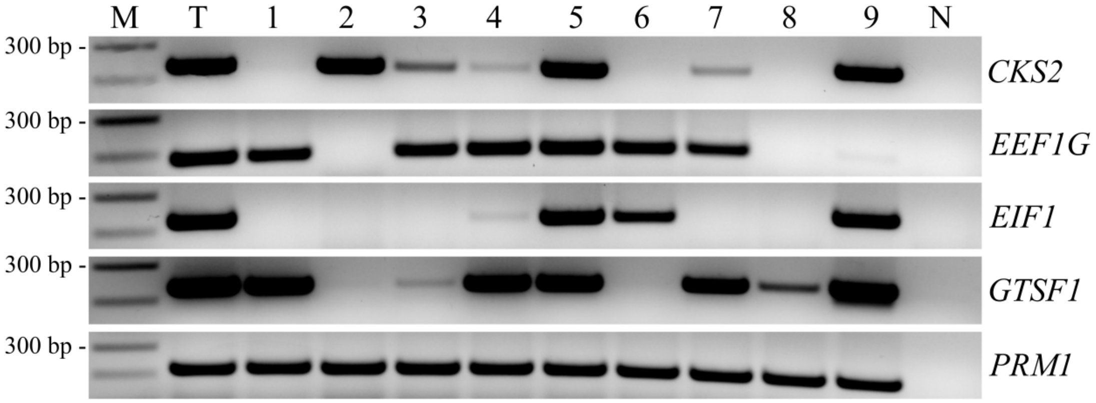


Figure 4. PCR amplification of select transcripts in individual bull spermatozoa amplified cDNA



Tables

Table 1. Bovine primer sequences. *Genomic DNA amplicon which includes 144 bp intron

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	**Individual Variation**	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
EIF1	BC103170	**qPCR**	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	**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**	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

Table 2. Top 10 bovine spermatozoal transcripts based on FPKM.

Million fragments mapped (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

Table 3. Comparing 5’ and 3’ exons in transcripts from previous literature

			End Exons Intact				
Transcript	Accession #	FPKM	5'	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

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
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
COX7A2	Cytochrome oxidase subunit 7a polypeptide 2	DQ347636	719
MLF1	Myeloid leukemia factor 1	BC109859	517
PFDN5	Prefoldin subunit 5	BC102252	405
GABARAP	GABA(A) receptor-associated protein	AJ297742	385

Table 5. Top 10 gene ontology categories for all spermatozoal transcripts FPKM>0 and for transcripts with FPKM>100. CP = Catabolic Process, Bi = Binding, Org = Organelle, NMBO = Non-Membrane Bounded Organelle, TTA = transmembrane transporter activity.

	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

Table 6. Full-Length Trasncripts 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 Protein 5	2778.08	N	N	N

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 Ortholog	887.309	Y	Y	Y

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 IMAGE:8063913	311.9	Y	Y	Y
BC103057	*UQCRB*	Ubiquinol-Cytochrome C	309.82	Y	Y	Y

		Reductase Binding Protein				
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 Rna, Similar To D-Loop	235.006	Y	Y	Y
BC110123	*C16H1orf49*	Chromosome 1 Open	234.333	Y	Y	Y

		Reading Frame 49 Ortholog				
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)	188.791	N	N	Y

		Box Polypeptide 25				
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	186.038	Y	Y	Y
AY911354	*LOC785691*	Similar To Mcg10725; 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 Similar To Beta 2-Microglobulin	151.332	N	N	Y

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	121.825	N	N	N

		Inhibitor 2				
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: MOLECULAR REPRODUCTION AND DEVELOPMENT MANUSCRIPT

Title: Are full-length mRNA in *Bos taurus* spermatozoa transferred to the oocyte during fertilization?

Authors: Anderson EJ[1], Card C[1], Sartini BL[1,3]

[1]Department of Fisheries, Animal and Veterinary Science, University of Rhode Island, Kingston, Rhode Island 02881

[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.

Institution at which work was performed: University of Rhode Island, Kingston, RI

Short Title: Full-Length mRNA in Bull Spermatozoa

Key Words (3-6): Spermatozoa, full-length, 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; PCR = polymerase chain reaction; UTR = untranslated region

Grant Support: This project was supported by an 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:

Spermatozoa are now known to contain a limited number of mRNAs and the contribution of these mRNAs to early embryonic development has been controversial. Although the spermatozoal transcript profile contains a predominance of degraded transcripts, only full-length mRNA transcripts will be capable of producing a protein after the spermatozoa fertilize the oocyte and the presence of full-length spermatozoal transcripts has not been investigated. This study sequenced 24 spermatozoal transcripts, chosen using four techniques: high FPKM within an RNA-Seq dataset, Gene Ontology of the RNA-Seq dataset, Y chromosome specificity, and transcripts discovered from previous studies, to determine if full-length transcripts exist in spermatozoa. Of these transcripts, 16 were full-length, while 8 are degraded. Additionally, four transcripts, *PSMA6, ATPase β*, *CHMP5*, and *DDX3Y*, have alternative polyadenylation sites, which the referenced RNA-Seq study failed to identify. To further characterize the potential functionality of these transcripts, the transfer of spermatozoal transcripts to oocytes and 2-cell stage embryos was examined. However, recurrent contamination and conflicting results from the oocyte and 2-cell embryos occurred. Despite this, the 16 transcripts that have been found to be full-length in spermatozoa do merit further investigation, and prove that while many transcripts are degraded remnants from spermatogenesis, full-length mRNAs, with the potential to be a functional protein, are found in spermatozoa.

Introduction:

Recent research has shown that spermatozoa contain not only genomic DNA, but also a subset of mRNA (Lalancette et al. 2008; Avendaño et al. 2009; Johnson et al. 2011; Card et al. 2013). Although spermatozoa are not transcriptionally or translationally active, they still contain mRNAs present from the final surge of transcriptional activity at the end of spermatogenesis (Miller and Ostermeier 2006; Boerke et al. 2007). Many of the transcripts found in spermatozoa are degraded (Gilbert et al. 2007), although functional roles of spermatozoa mRNAs have been proposed, including functions in early embryo development regulation, epigenetic modifications, paternal genome maintenance, and structural functions (Lalancette and Miller 2008). Spermatozoal mRNAs are likely most important prior to the transition from maternal to zygotic gene expression, since this is when they are most capable of affecting gene expression and epigenetic regulation. In cows, a transcriptional burst occurs at the eight to sixteen-cell embryonic stage, when zygotic control of mRNA expression begins (Vigneault et al. 2009). This activation occurs approximately 62 hours after fertilization, so any transcripts carried by the spermatozoa provide an advantage since there is little time for transcription and translation to occur before the zygote takes control of the mRNA expression (Memili and First 2000).

A role for sperm-derived RNAs after fertilization is supported by a functional role for microRNA-34c during the first cleavage of mouse embryos (Liu et al. 2012). Micro-RNAs may also have an epigenetic impact on the embryo (Carrell and Hammoud 2010). To date, *PLC*- ζ is the only sperm-derived mRNA with a known function in the oocyte after fertilization. *PLC*- ζ triggers the Ca^{2+} oscillations during oocyte activation (Swann et al. 2006; Hamatani 2012), and may also have a role in embryo cell-signaling (Boerke et al. 2007).

Limited studies have demonstrated that other spermatozoal transcripts can be detected in embryos, but are not present in oocytes, including *AKAP 4, CLGN*, *CLU, DDX3Y*, *PLC-ζ*, *PRM1*, *PRM2 SPAG9*, and *SRY* (Suri 2004; Boerke et al. 2007; Kempisty et al. 2008b). Of particular interest to this study are genes that are spermatozoa-specific, as they will be most useful for downstream protein identification in the embryo. The use of Y-chromosome specific transcripts allows for definitive identification of heritage of the transcript from the spermatozoa, an example transcript is *DDX3Y* (Sekiguchi et al. 2004; Vong et al. 2006; Bermejo-Alvarez et al. 2010; Yao et al. 2010a; Yao et al. 2010b). DDX3Y, a DEAD-box RNA helicase that can change RNA secondary structure, is sperm-derived but found in the embryo at fertilization (Yao et al. 2010a). A reduction in sperm-derived *DDX3Y* mRNA decreases embryonic development rates therefore a role for sperm-derived *DDX3Y* in early embryogenesis has been proposed (Abdelhaleem 2005; Yao et al. 2010b).

However, previous research has identified individual spermatozoal transcripts primarily using microarrays that detect only a small segment of a transcript and cannot determine if a full-length transcript is present. Although a large proportion of spermatozoal transcripts are degraded, but they may still be used as predictors of bull fertility (Gilbert et al. 2007; Bissonnette et al. 2009). No previous research has investigated if spermatozoal mRNAs are full-length, including the coding region and untranslated regions, despite the fact that incomplete mRNAs will be incapable of producing proteins used in embryo development. Identifying full-length transcripts may allow further studies to elucidate if these transcripts have a functional role in the embryo or are degraded in the embryo.

We hypothesize that full-length mRNAs are present in spermatozoa and some may be transferred to the oocyte and detected in the early embryo after fertilization. In this study, candidate spermatozoal transcripts were identified from the transcript profile of bull spermatozoa sequenced by RNA-Seq (Card et al. 2013). Transcripts were identified using four techniques: 1)if they had high FPKMs in the RNA-Seq dataset, 2) found in Gene Ontology (GO) categories of interest, 3) located on the Y chromosome, or 4) present in previous literature. With RNA-Seq, potential full-length transcripts can be identified by alignment of sequenced fragments to the bovine genome. Potentially functional spermatozoal transcripts were also identified by comparing microarray data from oocyte and embryo studies (Kocabas et al. 2006; Chalmel et al. 2007; Huang and Khatib 2010).

Additionally, this work highlights some specific examples of alternative 3’ ends for transcripts that are found in spermatozoa, and provides quantitative percent coverage of several transcripts. Identification of full-length spermatozoal transcripts that are transferred to the oocyte at fertilization could be used in further functional studies.

Materials and Methods:

Spermatozoal transcript selection

Four approaches were used to identify select candidate bovine spermatozoal transcripts for further analysis to determine if these transcripts are full-length in spermatozoa and if these transcripts are transferred to the oocyte at fertilization (Table 1). The first three approaches mined the bovine spermatozoa transcript profile (Card et al. 2013) and the same pool of bovine spermatozoa from nine bulls was used in both studies to validate full-length transcripts. In the first approach, we identified nine spermatozoal transcripts with high FPKMs (fragments per kilobase million mapped reads) in the bovine spermatozoal transcript profile (Card et al. 2013). In the second approach, we identified transcripts in the top Gene Ontology (GO) category, translation (Card et al. 2013). Third, Y chromosome specific transcripts in the bovine spermatozoal transcript profile were identified because it is a definitive way to prove that the transcript was sperm-derived in embryo. Finally, thirteen potentially functional spermatozoal transcripts from previous studies were confirmed present the bovine spermatozoa transcript profile.

To determine if any of the selected transcripts were potentially transferred to the oocyte by the spermatozoa, previous microarray studies were used to find transcripts that were previously present in spermatozoa (Chalmel et al. 2007) but absent in the oocyte (Kocabas et al. 2006).

Tissue Samples

Bovine testes were collected from a local abatoir for RNA isolation. Bovine cumulus-free oocytes and two-cell stage embryos were obtained from Sexing

Technologies (Navasota, TX). Cryopreserved bovine spermatozoa straws were obtained from Genex Cooperative Inc., (Shawano, WI). Spermatozoa from nine individual bulls was pooled after RNA isolation (see below) for cDNA amplification. These individuals had Conception Rate (CR) scores ranging from -2.9 to 3.5. Spermatozoa from an additional three bulls were used for mRNA Reverse Transcription (RT) for use in 5’ and 3’ Rapid Amplification of cDNA Ends (RACE) protocols. The three bull spermatozoa pool has CR scores of -0.3, 1.3, and -4.

Tissue RNA Isolations

RNA was isolated from bovine testis and spermatozoa using TRIzol isolation (Sigma-Aldritch; St. Louis, MO). Testis RNA isolations were performed according to manufacturer’s instructions, using 100 mg of tissue per isolation.

Spermatozoa TRIzol isolations were modified from the kit protocol as described in Card and Anderson et al. 2013. After washing spermatozoa with PBS (Phosphate Buffered Saline) to remove the cryoprotectant, 1 mL TRIzol with 15 µg/mL glycogen was added to the sperm. Samples were then lysed using a 26g needle and incubated at room temperature for thirty minutes. Then chloroform was added and incubated for 10 minutes at room temperature. Phase separation was done at 12000 *x g* for 15 minutes at 4° C. The aqueous layer was removed, combined with 500 µL ice-cold isopropanol, and chilled on ice for 10 minutes. After a 12000 *x g* centrifugation for 10 minutes, , the precipitated pellet washed with 1 mL 75% ethanol and centrifuged for an additional 10 minutes. Supernatant was removed, and pellet was air dried. The pellet of RNA was then resuspended in nuclease free water. RNA samples were DNAse treated using the RNeasy Mini Kit (Qiagen, Valencia, CA), and

RNA concentrations were measured using the NanoDrop UV/Vis spectrometer (Thermo Scientific; Wilmington, DE). Samples were then stored at -80°C.

Bovine oocyte and 2-cell stage embryo RNA isolations were performed according to manufacturer’s instructions, using the PicoPure RNA isolation kit (Arcturus, Mountain View, CA), with 40 oocytes or embryos used per isolation.

RACE Reverse Transcription and Polymerase Chain Reaction (RACE RT-PCR)

RACE RT-PCR was used for amplification of 5’ and 3’ ends of transcripts, using the three bull spermatozoa pool. For transcripts <1200 bp, the full-length transcripts was amplified using this technique. RT-PCR 5’ and 3’ RACE protocols were performed according to manufacturer’s instructions (Invitrogen, Carlsbad, CA), using a Gene Specific Primer (GSP) (Table 2) paired with a Generacer 5’ or 3’ universal primer for improved amplification of the poly(A) tail and 5’ Untranslated Region (UTR). Transcripts that were <1200 base pairs were amplified using primers spanning from the 5’ exon to the 3’ exon, with a 5’ RACE and a 3’RACE PCR run in conjunction to amplify all areas as necessary. For transcripts with >1200 base pairs, overlapping amplicons were used to cover the entirety of the transcript.

Primers were designed for use with 5’ and 3’ RACE RT-PCRs to amplify ends of transcripts. Additional primer sets for each transcript of interest were designed to provide overlapping amplicons for sequencing of the midsection of each transcript, and when transcripts were larger than 1200 base pairs, additional standard RT-PCRs (discussed below) were used to amplify regions spanning the exons from the 5’ end to the 3’ end. Each primer set was tested on testis RNA as a positive control. Additionally, a negative control lacking template was run in parallel.

cDNA Amplification & Standard RT-PCR

Due to low yield from spermatozoa RNA isolations, amplified ds-cDNA from a nine bull pool was used for PCR for transcripts with >1200 bp. RNA was converted to ds-cDNA and amplified before use in PCR reactions (SMARTer Pico PCR cDNA Synthesis Kit; Clontech, Mountain View, CA). The cDNA amplification enriches for full-length mRNA populations by using a modified oligo(dT) primer. All PCRs were run for 26 cycles to optimize them to reach the linear phase. Each of the nine bulls was equally represented in the final pooled sample. PCRs were performed under standard conditions as follows: 1X reaction buffer, 1.5 mM MgCl2, 10 mM dNTPs, 2.5 µM forward and reverse primers (Table 2), and 2.5 U Taq polymerase (NEB, Ipswich, MA). PCRs were run under standard temperatures of: 94°C for 5 min, 35 cycles of 94°C for 30 sec, primer dependent annealing temperature for 30 sec then 72°C for 2 min followed by a final extension at 72°C for 10 min. All PCRs were run with a positive Testis RNA control and a negative control without template RNA.

Oocytes and two cell stage embryos were amplified in two separate sets. The first set was cDNA amplified into ds-cDNA using the SMARTer Pico PCR cDNA Synthesis Kit (Clontech, Mountain View, CA). The second set was performed using the standard RT-PCR conditions described above. All RT-PCRs were run with a positive testis control and a negative no template RNA control.

Agarose Gels & Sequencing

PCR products were visualized on a 2.0% agarose gel stained with ethidium bromide, and run by gel electrophoresis. Bands excised from the gel were gel purified (Qiagen Gel Extraction kit; Valencia, CA), and submitted for sequencing at the URI

Genomics Center (Kingston, RI). Amplicon sequence identity was confirmed by submission to NCBI BLAST and by alignment to the original transcript accession the primers were designed from.

Results:

Twenty-four spermatozoal transcripts were chosen for this study from previous spermatozoal transcript studies, a spermatozoal RNA-Seq dataset, and for their male specificity (Table 1). Transcripts were chosen if visual inspection of the RNA-Seq read mappings using the UCSC browser showed that exons were present in the RNA-Seq analysis (Figure 1; Meyer et al. 2013). Only two transcripts were selected using the Gene Ontology selection method from the translation category, *EEF1G* and *EEF1A1*. Additional transcripts that did not show complete exon mapping were also chosen for analysis if previous studies had hypothesized a function for these transcripts including the Y chromosome transcript, *DDX3Y*. *DDX3Y* was the only Y chromosome spermatozoa transcript identified. This is not entirely unexpected, as the Y-chromosome contains only 33 genes in total (Yao et al. 2010b).

Of the selected candidate spermatozoal transcripts examined, the full-length of 16 spermatozoal transcripts were PCR amplified for the genes *CCT8, H2AFZ, COX7A2, CRISP2, EEF1A1, EEF1G, GSTM3, PLCZ1, PRM1, PSMA1, HMGB4, GTSF1, CKS2, PSMA6, SEC61G*, and *CMYC* (Figure 2). Transcripts were considered to be full-length when they contained greater than 90% sequencing coverage, although the transcripts *CCT8, COX7A2, CRISP2, EEF1A1, EEF1G, GSTM3, PRM1,* and *PSMA1* have been 100% sequenced (Table 3A).

Eight additional transcripts could not be sequenced in entirety, most likely because they are degraded in spermatozoa (Figure 1B). These eight transcripts found in the bovine spermatozoal transcript RNA-Seq data had similar missing regions in PCR amplifications (Figure 3; Table 3). An example of this is the transcript *PEBP1*, which had read mappings missing from the 5’ end exon entirely; corresponding to missing reads found using PCR amplification (Figure 1). Five transcripts, *PEBP1, ATPase B, CHMP5, UBE2N*, and *DDX3Y*, were found to be non-full-length in both the RNA-Seq data and in PCR identification.

A distinguishing characteristic that was found when sequencing full-length transcripts was evidence of alternative 3’ untranslated regions (UTRs) in novel locations to previous accessions. Although accessions for mRNA transcripts may be labeled for a single gene, there may be multiple isoforms or variants of a transcript that exist. Four transcripts, *PSMA6, ATPase β, CHMP5,* and *DDX3Y*, had alternative 3’ UTRs (Figures 1 and 2; Table 3). These transcripts were degraded, with the exception of *PSMA6*, which was shown to be full-length.

The transcripts chosen for amplification in oocytes and 2-cell stage embryos were identified as transferred to the embryo from spermatozoa in a different species or because these transcripts were not present in microarray studies of oocytes (Ostermeier et al. 2004; Kocabas et al. 2006; Chalmel et al. 2007; Kempisty et al. 2008a). Transcripts found to be present in oocytes from previous microarray data were eliminated for further analysis, since PCR will be incapable of distinguishing the origin of the mRNAs in the zygote (Table 4). This left nine transcripts to investigate further that were absent in the oocyte microarray data: *AKAP4, CLGN, CMYC, CRISP2, DDX3Y, HMGB4, PLCZ1, PRM1,* and *SPA17*.

Contamination and conflicting results from PCR amplification of spermatozoal transcripts in oocytes and 2-cell embryos have yielded inconclusive results. Two replicates of oocytes and 2-cell embryos pools, one cDNA amplified and one unamplified RNA, have demonstrated amplification of male-specific transcripts in female tissue, for example *DDX3Y* and *PRM1* (Figure 4).

Discussion:

Visual inspection of RNA-Seq reads mapped to the bovine genome from the bovine spermatozoa transcript profile was used to identify sixteen full-length transcripts and these were validated by PCR amplification and subsequent sequencing, proving that these transcripts have the basic potential for use downstream in the early stage embryo. These results validate RNA-Seq read mappings by directly sequencing the 5' and 3' UTRs in addition to the complete protein coding region of these transcripts. Importantly, PCR sequencing is capable of identifying transcripts that have been alternatively polyadenylated, which may impact the timing of translation, while RNA-Seq cannot. Additionally, if multiple isoforms of a transcript are present, they can be detected through PCR more accurately than with RNA-Seq. Studies of this nature that use both RNA-Seq and PCRs may prove useful as prerequisite identification of targets for determining if these transcripts are translated into protein in the embryo.

While other studies have looked at the mRNAs contained in spermatozoa, this is the first study to identify full-length transcripts. The importance of this is noted by the fact that this study found transcripts of interest from previous literature (Table 1) that are not full-length (Table 3), and therefore likely non-functional. An example of this is *DDX3Y*, which has been examined in previous studies. One study found that reduction in sperm-derived *DDX3Y* mRNA has a role in spermatogenesis (Abdelhaleem 2005). A second study found that *DDX3Y* was transferred from mouse spermatozoa to the oocyte and that it's absence decreases embryo development rates (Yao et al. 2010a). Although these previous studies point to a role for *DDX3Y*, it was degraded in the spermatozoa transcript population reported here and therefore could not be functional as a translational template in the embryo. The mRNA transcript *AKAP4* which was found to be degraded in this studies transcript profile correlates with its known function as a a regulator of spermatozoa motility (Miki et al. 2002). However, the use of these degraded transcripts may still serve as indicators of fertility, despite knowing that a particular degraded transcript will not have a functional use in the oocyte (Lalancette et al. 2008).

The transcripts found to be full-length in this study will be useful candidates for examination of protein function in future studies to determine if sperm-derived mRNA does have a functional role in early embryogenesis. Spermatozoal mRNAs are likely most important prior to the transition from maternal to zygotic gene expression, since this is when they are most capable of affecting gene expression and epigenetic regulation. In cows, a transcriptional burst occurs at the eight to sixteen-cell stage, when zygotic control of mRNA expression begins (Vigneault et al. 2009). This activation occurs around 62 hours after fertilization, so any transcripts carried by the spermatozoa provide an advantage since there is little time for transcription and translation to occur before the zygote takes control of the mRNA expression (Memili and First 2000). To date, the only full-length spermatozoal transcript known to have an embryonic function is *PLC-* ζ. *PLC-* ζ triggers the Ca^{2+} oscillations during mouse oocyte activation and may have a role in embryo cell-signaling (Swann et al. 2006; Boerke et al. 2007; Hamatani 2012). This study was able to demonstrate that *PLC-* ζ is full-length, although it is unable to confirm that *PLC-* ζ is delivered to the oocyte during bovine fertilization.

Other full-length spermatozoal transcripts that were identified in this study are known to rapidly degrade after entering the oocyte, including *PRM1*, indicating no further functionality within the embryo (Avendaño et al. 2009; Thelie et al. 2009). Furthermore, this study does not indicate the degree of degradation that might occur to the spermatozoal transcripts once they reach the oocyte. To further explore how much degradation occurs and when it occurs, future studies will need to explore oocytes and embryos to trace the inheritance of these spermatozoal mRNAs.

Full-length spermatozoal transcripts were sequenced that warrant further functional experiments. For example, *CRISP2* may assist with spermatozoa survival in the female tract and fertilization (Arangasamy et al. 2011). Preliminary reports of translation occurring in the spermatozoa mitochondria may indicate a need to replace proteins necessary for fertilization, as this translation occurs in the spermatozoa tailpiece right after the spermatozoa undergo capacitation (Gur and Breitbart 2008).

An additional hypothesized function of spermatozoal mRNA is an epigenomic effect on embryonic development. Although the degree of epigenetic influence that the spermatozoal mRNAs exhibit is unknown, genes such as *PRM1* and *H2AFZ* have been shown to have an influence on the epigenome (Jenkins and Carrell 2012). Specifically, sperm-specific *PRM1* proteins replace histones in spermatozoa chromatin, regulating how development proceeds (Carrell and Hammoud 2010). *PRM1* may act to selectively unwind specific regions of the paternal DNA to make it more accessible for transcription and translation in the early embryo before being degraded (Miller et al.

2005). The transcript *H2AFZ* is an effector of gene regulation as well by acting as a histone modifier (Misirlioglu et al. 2006). By demonstrating that *PRM1* and *H2AFZ* are full-length in spermatozoa, this study demonstrates the potential for spermatozoa to have an epigenetic effect on embryonic development.

However, it is important to highlight the inherent bias to these selection methods, with a clear focus on isolating genes with a high probability of being full-length. This study cannot be used to comment on the percentage of transcripts that may be full-length in the spermatozoa, only to prove that some do exist.

This study is incapable of determining if full-length spermatozoal transcripts are transferred to the oocyte at fertilization. The attempted oocyte and 2-cell embryo PCRs yielded inconclusive results. Several suspect amplifications indicated that the oocytes and 2-cell embryos were likely contaminated in both the cDNA amplified and unamplified RNA experiments (Figure 4). The presence of known sperm-specific transcripts and Y chromosome transcripts in these populations, as well as varying results from gel to gel (Figure 4), make it impossible to conclude more about the spermatozoa to oocyte inheritance of these transcripts at this time.

In conclusion, full-length transcripts have been definitively proven to exist within bull spermatozoa, lending credence to the potential function of these mRNAs within the oocyte and embryo. Global comparisons of oocyte and embryo transcripts to spermatozoal transcripts are also needed to further identify targets that may be translated into proteins and impact embryogenesis.

Acknowledgments:

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

References:

Abdelhaleem, M. 2005. RNA helicases: regulators of differentiation. Clinical biochemistry 38:499–503.

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.

Avendaño, C., A. Franchi, E. Jones, and S. Oehninger. 2009. Pregnancy-specific {beta}-1-glycoprotein 1 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. Human reproduction (Oxford, England) 24:270–7.

Bermejo-Alvarez, P., P. Lonergan, D. Rath, A. Gutiérrez-Adan, and D. Rizos. 2010. Developmental kinetics and gene expression in male and female bovine embryos produced in vitro with sex-sorted spermatozoa. Reproduction, fertility, and development 22:426–36.

Bissonnette, N., J.-P. Lévesque-Sergerie, C. Thibault, and G. Boissonneault. 2009. Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality. Reproduction (Cambridge, England) 138:65–80.

Boerke, a, S. J. Dieleman, and B. M. Gadella. 2007. A possible role for sperm RNA in early embryo development. Theriogenology 68 Suppl 1:S147–55.

Card, C., E. J. Anderson, S. Zamberlan, K.-B. E. Krieger, M. Kaproth, and B. L. Sartini. 2013. Cryopreserved Bovine Spermatozoal Transcript Profile as Revealed by High-Throughput Ribonucleic Acid Sequencing. Biology of Reproduction .

Carrell, D. T., and S. S. Hammoud. 2010. The human sperm epigenome and its potential role in embryonic development. Molecular human reproduction 16:37–47.

Chalmel, F., A. D. Rolland, C. Niederhauser-Wiederkehr, S. S. W. Chung, P. Demougin, A. Gattiker, J. Moore, et al. 2007. The conserved transcriptome in human and rodent male gametogenesis. Proceedings of the National Academy of Sciences of the United States of America 104:8346–51.

Gilbert, I., N. Bissonnette, G. Boissonneault, M. Vallée, and C. Robert. 2007. A molecular analysis of the population of mRNA in bovine spermatozoa. Reproduction (Cambridge, England) 133:1073–86.

Gur, Y., and H. Breitbart. 2008. Protein synthesis in sperm: dialog between mitochondria and cytoplasm. Molecular and cellular endocrinology 282:45–55.

Hamatani, T. 2012. Human spermatozoal RNAs. Fertility and sterility 97:275–81.

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

Jenkins, T. G., and D. T. Carrell. 2012. The sperm epigenome and potential implications for the developing embryo. Reproduction (Cambridge, England) 143:727–34.

Johnson, G. D., C. Lalancette, A. K. Linnemann, F. Leduc, G. Boissonneault, and S. a Krawetz. 2011. The sperm nucleus: chromatin, RNA, and the nuclear matrix. Reproduction (Cambridge, England) 141:21–36.

Kempisty, B., P. Antosik, D. Bukowska, M. Jackowska, M. Lianeri, J. M. J. M. Jaśkowski, P. P. Jagodzi’nski, et al. 2008a. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reproduction, fertility, and development 20:513–8.

Kempisty, B., P. Antosik, D. Bukowska, M. Jackowska, M. Lianeri, J. M. Jaśkowski, and P. P. Jagodzi’nski. 2008b. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. Reproduction, Fertility and Development 20:513–518.

Kocabas, A. M., J. Crosby, P. J. Ross, H. H. Otu, Z. Beyhan, H. Can, W.-L. Tam, et al. 2006. The transcriptome of human oocytes. Proceedings of the National Academy of Sciences of the United States of America 103:14027–32.

Lalancette, C., and D. Miller. 2008. Paternal contributions: new functional insights for spermatozoal RNA. Journal of cellular … 104:1570–1579.

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. Biology of reproduction 78:618–35.

Liu, W., R. T. K. Pang, P. C. N. Chiu, B. P. C. Wong, K. Lao, and K. Lee. 2012. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. PNAS 109:490–494.

Memili, E., and N. L. First. 2000. Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. Zygote (Cambridge, England) 8:87–96.

Meyer, L. R., A. S. Zweig, A. S. Hinrichs, D. Karolchik, R. M. Kuhn, M. Wong, C. a Sloan, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. Nucleic acids research 41:D64–9.

Miki, K., W. D. Willis, P. R. Brown, E. H. Goulding, K. D. Fulcher, and E. M. Eddy. 2002. Targeted Disruption of the Akap4 Gene Causes Defects in Sperm Flagellum and Motility. Developmental Biology 248:331–342.

Miller, D., and G. C. Ostermeier. 2006. Spermatozoal RNA: Why is it there and what does it do? Gynécologie, obstétrique & fertilité 34:840–6.

Miller, D., G. C. Ostermeier, and S. a Krawetz. 2005. The controversy, potential and roles of spermatozoal RNA. Trends in molecular medicine 11:156–63.

Misirlioglu, M., G. P. Page, H. Sagirkaya, a Kaya, J. J. Parrish, N. L. First, and E. Memili. 2006. Dynamics of global transcriptome in bovine matured oocytes and preimplantation embryos. Proceedings of the National Academy of Sciences of the United States of America 103:18905–10.

Ostermeier, G. C., D. Miller, J. D. Huntriss, M. P. Diamond, and S. A. Krawetz. 2004. Delivering spermatozoan RNA to the oocyte. Nature 429:2603.

Sekiguchi, T., H. Iida, J. Fukumura, and T. Nishimoto. 2004. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. Experimental cell research 300:213–22.

Suri, A. 2004. Sperm specific proteins-potential candidate molecules for fertility control. Reproductive biology and endocrinology□: RB&E 2:10.

Swann, K., C. M. Saunders, N. T. Rogers, and F. a Lai. 2006. PLCzeta(zeta): a sperm protein that triggers Ca2+ oscillations and egg activation in mammals. Seminars in cell & developmental biology 17:264–73.

Thelie, A., P. Papillier, C. Perreau, S. Uzbekova, C. Hennequet-Antier, and R. Dalbies-Tran. 2009. Regulation of bovine oocyte-specific transcripts during in vitro oocyte maturation and after maternal-embryonic transition analyzed using a transcriptomic approach. Molecular reproduction and development 76:773–82.

Vigneault, C., C. Gravel, M. Vallée, S. McGraw, and M.-A. Sirard. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. Reproduction (Cambridge, England) 137:245–57.

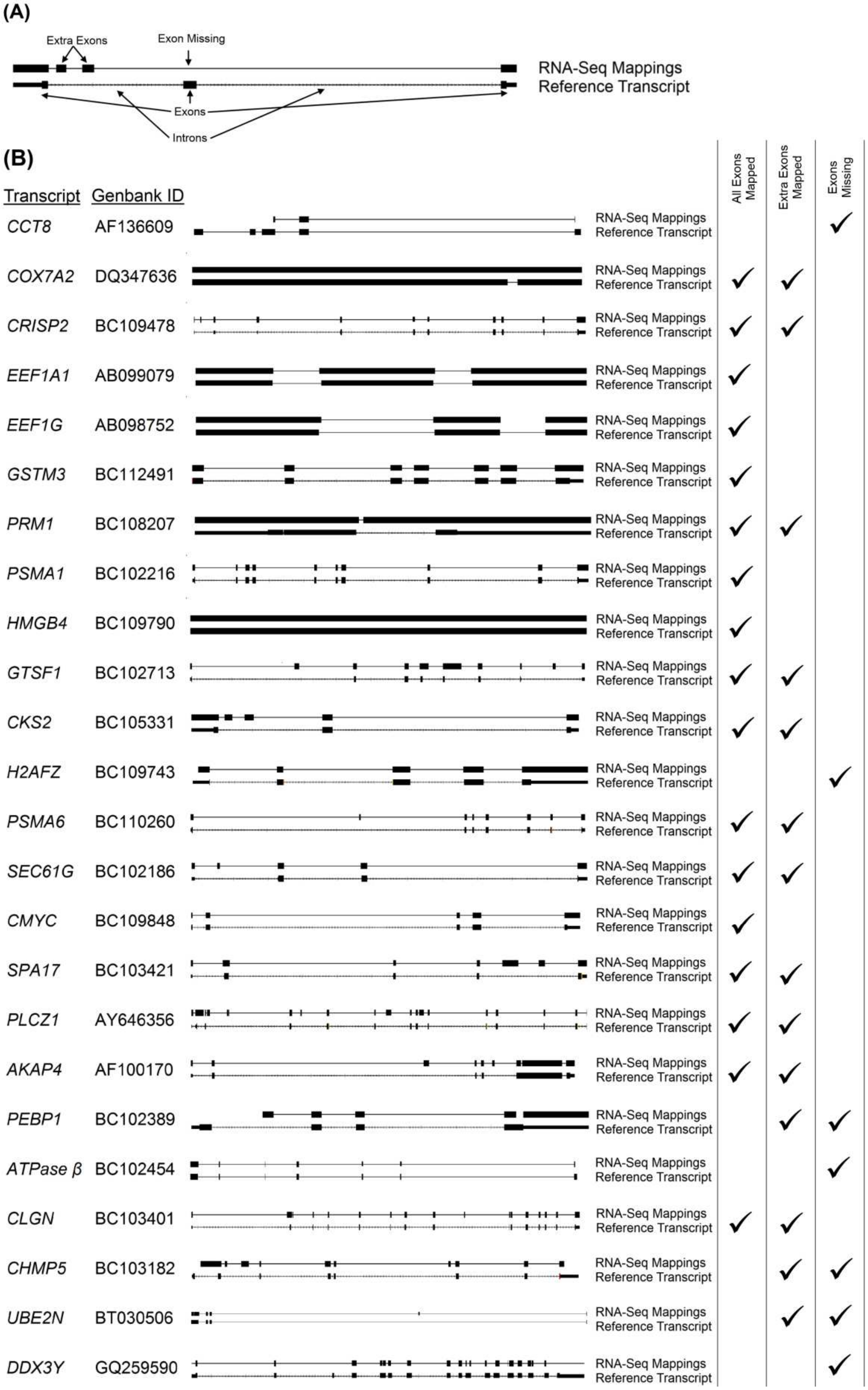
Vong, Q. P., Y. Li, Y.-F. C. Lau, M. Dym, O. M. Rennert, and W.-Y. Chan. 2006. Structural characterization and expression studies of Dby and its homologs in the mouse. Journal of andrology 27:653–61.

Yao, C. J., W. J. Xu, X. L. Gong, Y. Zhou, Z. Q. Yan, Z. J. Zhu, Z. X. Wang, et al. 2010a. The role of Dby mRNA in early development of male mouse zygotes. Asian Journal of Andrology 12:567–577.

Yao, C., Z. Wang, Y. Zhou, W. Xu, Q. Li, D. Ma, L. Wang, et al. 2010b. A study of Y chromosome gene mRNA in human ejaculated spermatozoa. Molecular reproduction and development 77:158–66.

Figures

Figure 1: Alignment of spermatozoa transcript RNA-Seq reads (bottom line) with bovine genome in UCSC Mappings (Meyer et al. 2013) using bovine spermatozoa RNA-Seq data from (Card et al. 2013) for bovine spermatozoa transcripts assayed in this study. Untranslated Regions (UTRs) are included in all transcripts shown here as thinner black lines at both ends of each transcript. RNA-Seq mappings don't show separate UTRs and have a continuous thick black line. Summary of read mappings indicating full-length (all exons mapped) or degraded transcripts (exons missing) is indicated on the right. Extra exons sequenced in RNA-Seq but not visualized in the bovine genome are also indicated.



Transcript	Genbank ID	All Exons Mapped	Extra Exons Mapped	Exons Missing
CCT8	AF136609			✓
COX7A2	DQ347636	✓	✓	
CRISP2	BC109478	✓	✓	
EEF1A1	AB099079	✓		
EEF1G	AB098752	✓		
GSTM3	BC112491	✓		
PRM1	BC108207	✓	✓	
PSMA1	BC102216	✓		
HMGB4	BC109790	✓		
GTSF1	BC102713	✓	✓	
CKS2	BC105331	✓	✓	
H2AFZ	BC109743			✓
PSMA6	BC110260	✓	✓	
SEC61G	BC102186	✓	✓	
CMYC	BC109848	✓		
SPA17	BC103421	✓	✓	
PLCZ1	AY646356	✓	✓	
AKAP4	AF100170	✓	✓	
PEBP1	BC102389		✓	✓
ATPase β	BC102454			✓
CLGN	BC103401	✓	✓	
CHMP5	BC103182		✓	✓
UBE2N	BT030506		✓	✓
DDX3Y	GQ259590			✓

Figure 2: Full-length bovine spermatozoal transcripts A) Transcripts sequenced from a single amplicon, spanning from 5' exon to the 3' exon. B) Transcripts sequenced from two overlapping amplicons, bands shown that cover 5' and 3' ends C) Transcript sequenced from three overlapping amplicons, from 5' to midsection to 3' ends. D) Full-length transcript found with an alternative 3'UTR. Negative control = no template.

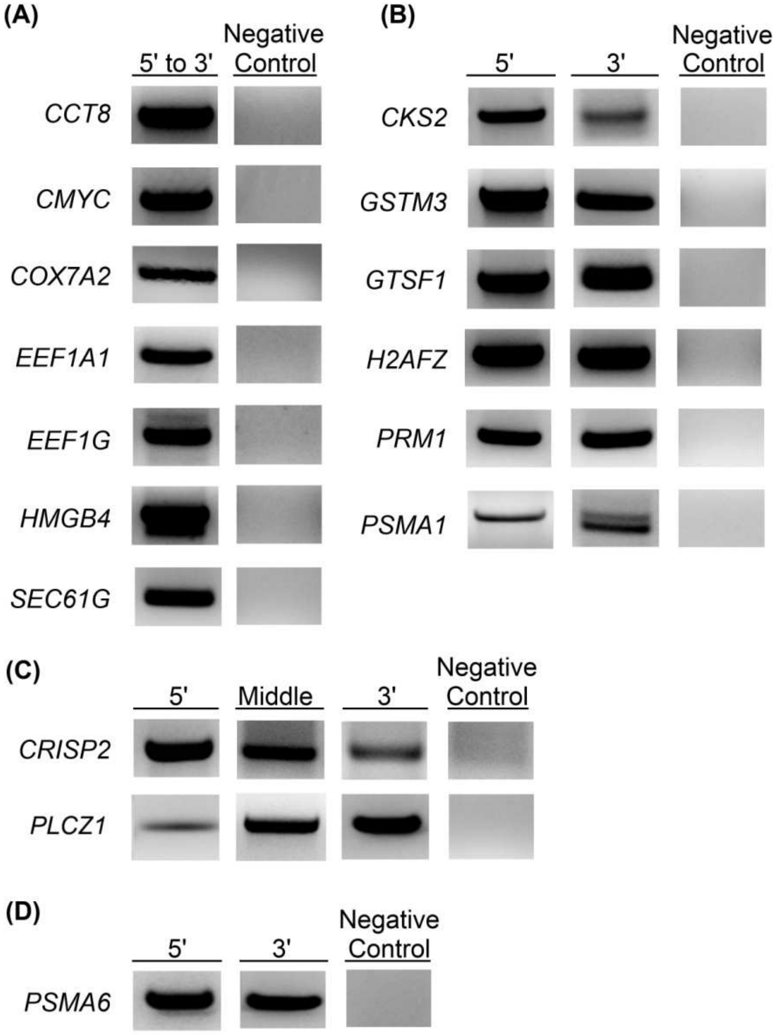


Figure 3: Degraded transcripts present in the bovine spermatozoa transcript profile. A) Transcripts with degradation, amplicons sorted by 5’ to 3’ end location. B) Transcripts degraded on the 5’ end, as well as having alternative 3’UTR. ATPase β and CHMP5 have only one band , but the band size and location of the poly(A) tail were different from the reference accession. Negative control = no template.

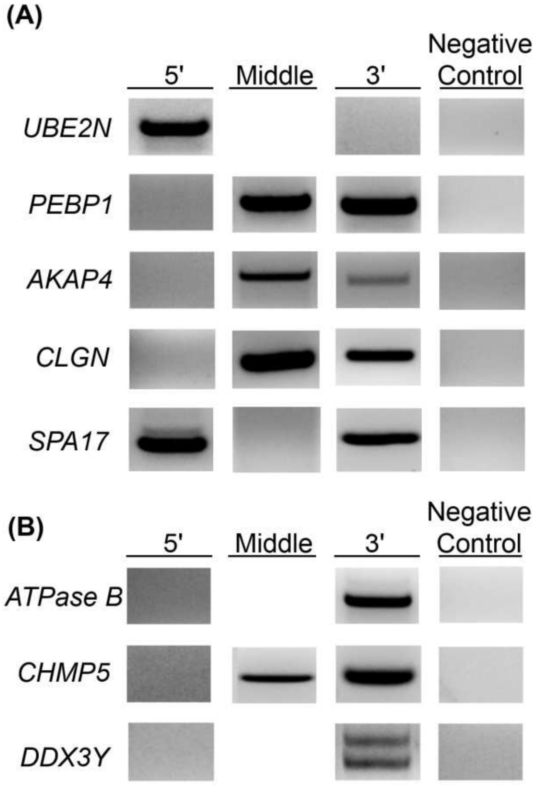
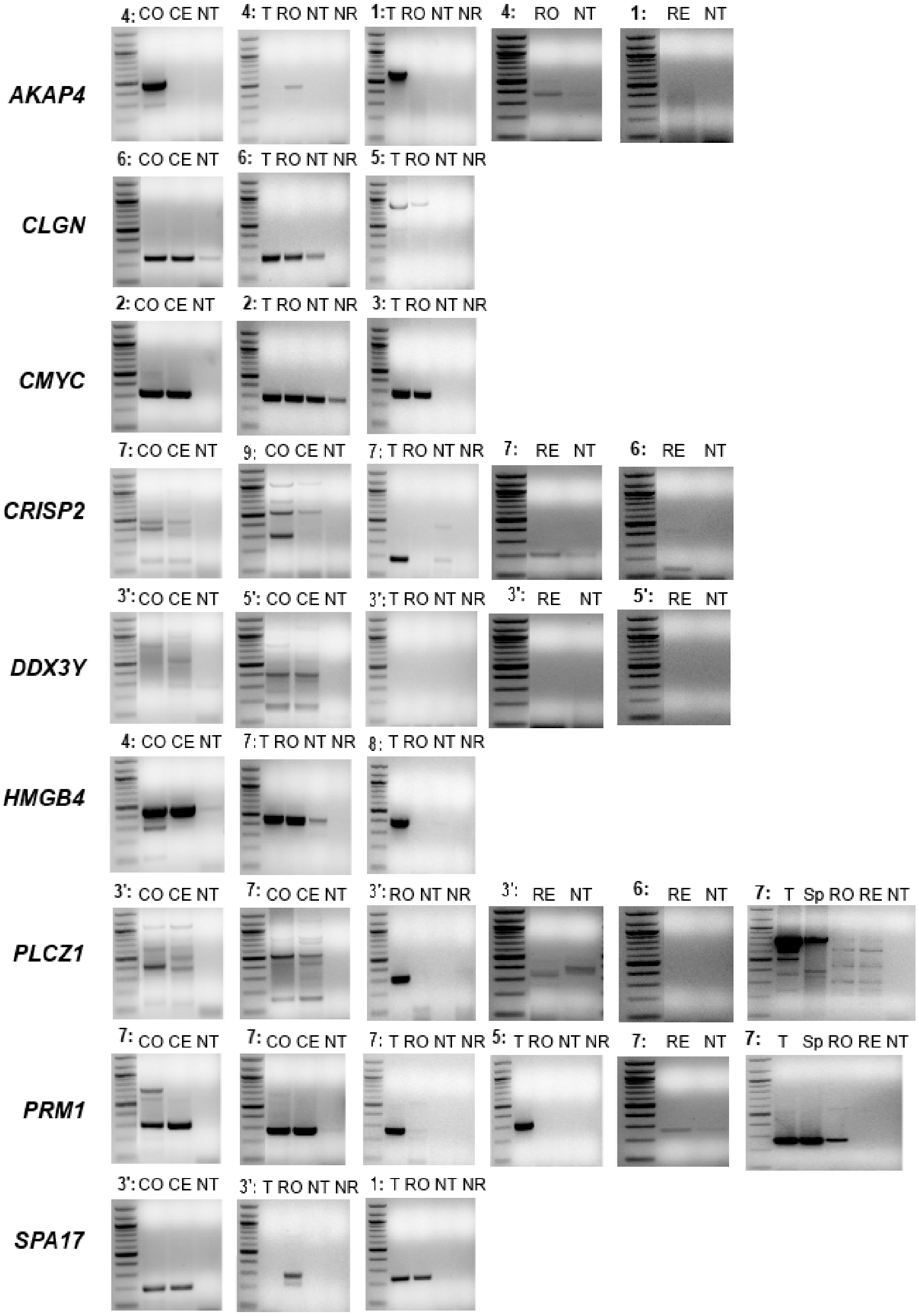


Figure 4: Oocyte and 2-Cell Embryo PCRs. All genes are arranged in order from left to right that the PCRs were performed. CO= cDNA oocyte, CE= cDNA 2-cell embryo, NT= No Template Control, NR= No RT Template Control, RO=RACE oocyte, RE=RACE 2-cell embryo, Sp= Spermatozoa cDNA from 9-bull pool.



Tables

Table 1: Transcript identification methods. Lit=Literature Searches, Y Csome= Y chromosome located, GO=found in gene ontology translation category

Gene	FPKM	Discovery Method	Reference
AKAP4	126.623	Lit	Ostermeier et al., 2005; Iguchi et al., 2006; Miller et al., 2006; Boerke et al., 2007; Gilbert et al., 2007; Peddinti et al., 2008
ATPase β	90.1644	High FPKM	
CCT8	17.9915	Lit	Arangasamy et al., 2011
CHMP5	2778.08	High FPKM	
CKS2	351.893	Lit	Donovan et al., 2003; Smirnova et al., 2006
CLGN	220.011	Lit	Iguchi et al., 2006; Hecht et al., 2010; Evans, 2012
CMYC	39.731	Lit	Miller et al., 2006; Gilbert et al., 2007; Gur et al., 2008; Mayr et al., 2009; Johnson et al., 2011; Hamatani, 2012; Liu et al., 2012
COX7A2	719.358	Lit	Misirlioglu et al., 2006; Bermejo-Alvarez et al., 2010; Alshagga et al., 2011
CRISP2	0.317044	Lit	Arangasamy et al., 2011
DDX3Y	10.9846	Lit, Y Csome	Sekiguchi et al., 2004; Vong et al., 2006; Yao et al., 2010; Paria et al., 2011
EEF1A1	431.873	GO	
EEF1G	116.059	GO	
GSTM3	2373.84	Lit	Misirlioglu et al., 2006; Bermejo-Alvarez et al., 2010; Alshagga et al., 2011
GTSF1	896.368	High FPKM	
H2AFZ	166.742	Lit	Misirlioglu et al., 2006; Gilbert et al., 2007; Vigneault et al., 2009; Bermejo-Alvarez et al., 2010
HMGB4	6021.96	High FPKM	
PEBP1	29.3446	Lit	Arangasamy et al., 2011
PLCZ1	41.3639	Lit	Miller et al., 2006; Swann et al., 2006; Boerke et al., 2007; Avendaño et al., 2009; Johnson et al., 2011; Fischer et al., 2012; Hamatani, 2012
PRM1	20667.2	Lit	Lassalle et al., 1999; Ostermeier et al., 2002; Miller et al., 2006; Gilbert et al., 2007; Lalancette et al., 2008; Avendaño et al., 2009; Bissonnette et al., 2009; Carrell et al., 2010; Feugang et al., 2010; Hecht et al., 2010; Johnson et al., 2011; Ganguly et al., 2012; Jenkins et al., 2012
PSMA1	84.5454	High FPKM	
PSMA6	913.21	High FPKM	
SEC61G	309.446	High FPKM	
SPA17	927.374	High FPKM	
UBE2N	245.596	High FPKM	

Table 2: Primers Used for PCR Amplification

Gene	**Genbank ID**	**Set**	**Forward (5' → 3')**	**Reverse (5' → 3')**	**BP Covered**	**Total BP**
AKAP4	AF100170	1	AGGGGTCAGTGTGCCTTTTC	TCGACCACCATCCCTACACT	385 - 998	2900
		2	AACATGCACTGAGCCCTTCA	TGGGCCATTTCAGAAGCGAT	481 - 931	
		3	TTTTACCAGACCAGTGGCCC	GACACCCTGTATTTGCACAGTC	2310 - 2840	
		4	CTACCAAGACTCTCACGGGC	ACACCCTGTATTTGCACAGTC	2363 - 2839	
		5	ACTTGGCACTGCCCACTTC	TCCAGACGTAGGCTCTGAGG	41 - 618	
		6	GCTTCTGAAATGGCCCATGAT	TCTTCACAGTTGAAGGGGCTC	915 - 2024	
		7	GAAAAGTGCGGAGGTAGCCA	GACACCCTGTATTTGCACAGTC	1857 - 2840	
ATPase β	BC102454	1	TGAGCGCACTCTGCTTGAG	CAGGAACTCTCCGGTTGTCC	1 - 253	1525
		2	GCACACACACCATGACGAAG	ATCCAACGCTGACACTGGTT	156 - 442	
		3	AATGTGGGCCATGCTTCAGA	GTTCAAAAGGCAGCAGAGGC	337 - 1210	
		4	AGATAATAGTGGTCCATGTCCTTCG	CACTGAATTTCTTGCAGCTATGA	1151 - 1175	
CCT8	AF136609	1	CGAGATGGTGCCATTTCTACC	TCTTTCTTCCCACTTGGAGGC	11 - 531	555
CHMP5	BC103182	1	GAGTGTTTAGGTTTTCCTAGCGG	GAAAGCCTCCTCCAAGCAA	1 - 869	1355
		2	TGGCACGGTGGACAGCAGAG	GAAAGCCTCCTCCAAGCAA	95 - 869	
		3	GGAGTCCTGGTGGATGAATTTGG	AGCCTCCTCCAAGCAACGAGT	645 - 866	
CKS2	BC105331	1	GAGTCGAGTCGTTGCCTTCA	GGACACCAAGTCTCCTCCAC	1 - 248	737
		2	CTTCACCTGACCCGGACGTT	AAAACACCTTACAGTAACCTACTTG	16 - 528	
CLGN	BC103401	1	TGGATTGAGCTGGGGGAGA	ATTCACCACACCCAATCCGA	26 - 1181	2450
		2	GGGCCCGCAAGACAGATAAT	TCAAGCCAGCCATCAGGTTT	98 - 1056	
		3	ATGGATGGAGAATGGGAGGC	GAAACTTTATTGCAATCAGCTCTGT	1118 - 2385	
CMYC	BC109848	1	GGCCGCTGTCACTATGGC	TCCTCCTGAGGTGGTTCATACT	1 - 309	528
		2	TGTCACTATGGCCCATTACAAA	TGGTGAGGTGGTTCATACTGAG	7 - 306	
		3	ACTTAGGAGCTGCTACCCCA	ACAGTTAACGTGTGATAGGTGA	165 - 463	
		4	CCACCTCAGGAGGAGAAACG	ACAGTTAACGTGTGATAGGTGAAT	296 - 463	
COX7A2	DQ347636	1	AACTGGCTGTGGCTTCGTTT	TGCTTTATTGGTGGCAGCTAA	1 - 204	206
CRISP2	BC109478	1	CGGCCGCTCTGCAACAGAAG	GTGCACTTGTTTGCCCACTT	20 - 437	1382
		2	AGTCTCTCCACCTGCCAGTA	TGCCTTCACACAGACAAGTCGCC	207 - 959	
		3	CACCTTGCGGCAGTTGCCCT	TGCCTTCACACAGACAAGTCGCC	761 - 922	
DDX3Y	GQ259590	1	TTGTTTCCGGTAGACCAACCTGTG	AGCGCCCTTTGCTAGCTGTACT	15 - 234	2790
		2	GGCCGTTCTAGGAGATTCAGTGG	CAACTGAATCTGCTTTCCAGCCAAG	1875 - 2120	
EEF1A1	AB099079	1	TCGTGTGGAGACTGGTGTTCT	TTAAAGACTGGGGTGGCAGTATTGG	4 - 634	636
EEF1G	AB098752	1	TCTTGCCCTGATTGAAGGCT	ATGGCTGGTCCCTGTGG	89 - 443	443
GSTM3	BC112491	1	GCGCTAAGGCACACAGGCGA	GGGCCATCTTGTTGTTGACAGGCAT	5 - 683	823
		2	CCGCATGCTCCTGGAGTTCACG	GTACAAGTCTGCCTCCTGCTC	94 - 728	
GTSF1	BC102713	1	GAGCACTTTGGATTTGGCTCC	CACGTGCTCTCAGCCATAGT	1 - 326	666
		2	AACTGGCAACTTGTCCCTTCA	GCAACTCACAGACAGTTTATCTT	173 - 650	
H2AFZ	BC109743	1	TGAGCGCAGTTTGAATCGC	CCACCACCAGCAATTGTAGC	1 - 436	884
		2	GTGGTGTCATTCCACACATCC	ATGACCTTTATTGAGCTTATCCACC	433 - 863	
HMGB4	BC109790	1	ACAGAAAATTTCACCGCCAGC	GACTCAGCTTGCTCGAACTCT	8 - 679	764
PEBP1	BC102389	1	TTTAACCTGGGTGGGTGTCAGC	CTCGTAAACCAGCCAGACATAGC	218 - 476	1476

		2	CCGATTATGTGGGCTCTGGG	ACCAACTCCAGAACAGTTTTCTTTT	134 409 - 1	
		3	AAGAGATTGACTGTCTCCGCC	ACAATATTCACCAACTCCAGAACA	100 135 9 - 0	
PLCZ1	AY646356	1	GCGTTTGGACCCAAAGGAAA	AGCAAGGCATCCCCAAATGT	108 6 - 9	2096
		2	TTTGGGGATGCCTTGCTGTC	AAGGCCACCATTTGACAACC	107 165 3 - 0	
		3	GGTCGGAATCCCACTCTTCA	AAAAGGGAAGCGGGCTCAA	134 209 8 - 6	
PRM1	BC108207	1	GACAGCCCACAAATTCCACC	GCAAGAGGGTCTTGAAGGCT	1 - 313	517
		2	GCCAGATACCGATGCTCCTCA	GTTAGCAGGCTCCTGTTCATGTC	98 - 364	
PSMA1	BC102216	1	TTCCACCCGCAGGTTTGAAG	GCTGATTGAGATCGGGCTCC	3 - 554	1172
		2	TTACGGGAAACCCTTCCTGC	ACAGTTGTCTTTAAAACCACAAAGA	113 634 - 2	
PSMA6	BC110260	1	GAGGGACGGCTCTACCAAGT	AGAGGCCGCATTTCAGCATT	118 - 464	984
		2	CGAAATTCCCGTGGACATGC	TGGCGTCACGGATTTGGTAA	384 - 838	
		3	TGGGTTTAAAGCAACTGCAGCAGGA	ATCGAGGGGCCCCCAAAATGT	546 - 889	
SEC61G	BC102186	1	GGCTCCTGTGCTACGTGTC	TTTCTGCTCCATCAGCTTCTCA	5 - 339	523
SPA17	BC103421	1	CCGGAACCATCGACTCCAGCTC	TCTTGCTCCTTGAATGCATGGTTGT	23 - 319	780
		2	GGGGCTAAGGTTGATGACCGC	GGCTAAGTGTCCCCGGAAGGC	267 - 464	
		3	TCAAGGAGCAAGAATCACCTG	GTGGGGGTAAAGCCAGTCTC	307 - 551	
UBE2N	BT030506	1	TGACAAGATGGCCGGGCTGC	GTGAGGGCTGTGATGTCTGT	106 27 - 8	2218
		2	ATCATCGGTGTCTTGCCACA	GAACAGCTTTGTGGTGGGGA	150 193 3 - 5	

Table 3: Sequencing of bovine spermatozoa transcripts where base pairs matched the predicted transcript using primers in Table 1. Y= Yes, N= No, n/a = no poly(A) present in genbank accession, * = Poly A tail is present but not where the original accession number indicates it. A) full-length transcripts B) degraded transcripts

A)

Gene	Genbank ID	Total BP	% Sequenced	BP sequenced	Poly A Tail
CCT8	AF136609	554	100	1 - 536	Y
COX7A2	BC102664	205	100	1 - 205	N
CRISP2	BC109478	1382	100	1 - 1382	Y
EEF1A1	AB099079	636	100	1 - 636	Y
EEF1G	AB098752	443	100	1 - 443	Y
GSTM3	BC112491	823	100	1 - 823	Y
PLCZ1	AY646356	2096	97	68 - 2096	Y
PRM1	BC108207	517	100	1 - 517	Y
PSMA1	BC102216	1172	100	1 - 1172	Y
HMGB4	BC109790	764	94	45 - 764	Y
GTSF1	BC102713	666	94	43 - 666	Y
CKS2	BC105331	737	93	49 - 737	Y
H2AFZ	BC109743	884	93	1 - 437 497 - 884	Y
PSMA6	BC110260	984	90	1 - 890	Y*
SEC61G	BC102186	523	89	57 - 523	
CMYC	BC109848	528	89	1 - 461	N

B)

Gene	Genbank ID	Total BP	% Sequenced	BP sequenced	Poly A Tail
SPA17	BC103421	780	60	1 - 320 370 - 514	N
AKAP4	AF100170	2900	50	440 - 1038 2056 - 2900	Y
PEBP1	BC102389	1476	50	370 - 684 1060 - 1476	N
ATPase β	BC102454	1525	31	1055 - 1525	Y*
CLGN	BC103401	2450	31	117 - 327 1017 - 1181 1656 - 1684 2105 - 2450	Y
CHMP5	BC103182	1355	22	66 - 269 647 - 740	Y*
UBE2N	BT030506	2218	21	160 - 442 878 - 1067	N
DDX3Y	GQ259590	2790	12	9 - 87 1948 - 2203	Y*

Table 4: Transcript presence in testis, sperm, oocyte and embryos from published microarray studies. Y=transcript present, N= transcript absent, M=results inconclusive. Oocyte & Embryo microarray data are from Kocabas et al. 2006. Testis and spermatozoa microarray data are from Chalmel et al., 2007.

Gene	**Genbank ID**	**Testis**	**Sperm**	**Oocyte**	**Embryo**
AKAP4	AF100170	Y	Y	N	N
ATPase β	BC102454	Y	Y	Y	N
CCT8	AF136609	Y	Y	Y	Y
CHMP5	BC103182	Y	Y	Y	N
CKS2	BC105331	Y	M	Y	Y
CLGN	BC103401	Y	M	N	N
CMYC	BC109848	Y	N	N	Y
CRISP2	BC109478	Y	Y	N	Y
DDX3Y	GQ259590	Y	Y	N	Y
EEF1A1	AB099079	Y	Y	Y	N
EEF1G	AB098752	Y	Y	Y	Y
GSTM3	BC112491	Y	Y	Y	N
GTSF1	BC102713	Y	Y	Y	N
H2AFZ	BC109743	Y	Y	M	Y
HMGB4	BC109790	Y	Y	N	Y
PEBP1	BC102389	Y	Y	Y	Y
PLCZ1	AY646356	Y	Y	N	N
PRM1	BC108207	Y	Y	N	N
PSMA1	BC102216	Y	Y	Y	Y
PSMA6	BC110260	Y	Y	M	Y
SEC61G	BC102186	Y	M	Y	Y
SPA17	BC103421	Y	Y	N	N
UBE2N	BT030506	Y	Y	Y	M

APPENDICES

APPENDIX I: PROTOCOLS

A) Testis RNA Isolations

B) Spermatozoa RNA Isolations

C) Oocyte & Embryo RNA Isolations

D) Reverse Transcription

E) Standard Polymerase Chain Reaction

F) cDNA Amplification

G) Rapid Amplification of cDNA Ends Polymerase Chain Reaction

H) Primer Dilution from IDT

I) Gel Extraction Protocol

J) Sequencing

K) RNA-Sequencing Protocols

A) Testis RNA Isolation Protocol

Tips & Techniques:

- Turn on microfuge and let cool to 4°C (20 minutes)
- Set up a ribonuclease free environment
    - Preparation: RNase/DNase free microfuge tubes, tips etc.
- KEEP SAMPLE ON ICE – except where noted

Protocol

1. Add 200 ul of TRI REAGENT to 100 mg tissue and homogenize with RNase-free blue pestle on ice. Add additional 800 ul of TRI REAGENT and mix.
2. Let sit at room temperature for 5 min *(Can store in –80°C at this point)*
3. Add 0.2 ml of chloroform per 1 ml of TRI REAGENT
4. Shake vigorously for 15 sec and let stand for 2-15 min
5. Centrifuge at 12,000 x g for 15 min at 4°C
6. Remove clear aqueous phase (top layer) and transfer to a new tube
7. Add 0.5 ml isopropanol/ml TRI REAGENT and let sit for 5-10 min at RT
8. Centrifuge at 12,000 g for 10 min at 4°C
9. Remove supernatant and wash the RNA pellet by adding 1 ml (minimum) of 75% ethanol (prepared with DEPC-treated water) per 1 ml of TRI REAGENT used in preparation.
    - *(Samples can be stored in ethanol at 4°C for 1 week and up to 1 year at -20°C).*

10. Vortex the sample and centrifuge at 7,500 for 5 min at 4°C

11. Briefly dry the RNA pellet for 5-10 min by air-drying on ice.

12. Add 10-50 ul of DEPC-water to RNA pellet. Mix at 55-60°C for 10-15 min.

RNA storage

Aqueous aliquots: After isolation, determine concentration using Nano drop in 10 mM Tris, pH 7.0. Freeze in 5 ug, 10 ug or 15 ug aliquots in DEPC-water. Store at -80°C

Ethanol aliquots: *most stable.* Precipitate by adding 1/10th volume 1M sodium acetate, pH 4.8 and 2.5 volumes of 95% ethanol. Store at -20°C. To retrieve: centrifuge at 7,5000 for 5 min at 4°C. Remove 95% ethanol from pellet and add 1 ml of 75% ethanol. Centrifuge at 7,5000 for 5 min at 4°C. Remove ALL ethanol, briefly air dry and re-suspend in DEPC-H_20.

B) Spermatozoa RNA Isolations

TRIzol Isolation with adaptations from Stallion Paper

1. Wash spermatozoa straws in 4 mL 1x PBS twice at 800x g for 10 minutes.
    - PBS at Room temp, water bath for spermatozoa straws at 37^{O}
2. Add 1 mL TRIzol reagent to spermatozoa pellet + 3 µL Glycogen.
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 reagent to sample.
    - Shake for 20 seconds then let sit at room temperature for 10 minutes.
5. Centrifuge at 12,000x g for 15 minutes at 4^{O} C.
6. Remove clear aqueous layer at top (contains RNA) and put in new tube.
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. Heat Tris for elution to 65^{O} C and add 60 µL to pellet.
    - Vortex until RNA pellet is dissolved in solution.
12. Nanodrop and store sample at -80^{O} C.

Total time expected: ~3 hours

C) Oocyte & Embryo RNA Isolations

(with PicoPure RNA isolation Kit)

Clean bench well with RNase out/Zap & wipe everything with 70% ethanol

** No need to dump waste between samples!**

1. Aliquot 20 µl of Extraction buffer (XB) into a 1.5 ml RNase-free tube
2. Collect oocytes/embryos into the extraction buffer
3. Parafilm tube and incubate in 42°C waterbath for 30 minutes (do not spin).
    - Stopping point: freeze on dry ice. (Our shipped samples are here)
4. Add 250 µl of Conditioning buffer (CB) onto the purification column.
5. Incubate at room temperature for 5 minutes.
6. Centrifuge column at 16.0 rcf for 1 minute
7. Add 20 µl of 70% ethanol into the cell extract from step number 3.
8. Mix well by pipetting (do not centrifuge).
9. Add the mixture onto the pre-conditioned column.
10. Centrifuge for 2 minutes at .1 rcf
11. Centrifuge for 30 seconds at 16.0 rcf
12. Add 100 µl of Wash Buffer 1 (W1)
13. Centrifuge for 1 minute at 8.0 rcf
14. Add 100 µl of Wash Buffer 2 (W2)
15. Centrifuge for 1 minute at 8.0 rcf
16. Add 100 µl of Wash Buffer 2 (W2)
17. Centrifuge for 3 minutes at 16.0 rcf
18. Add 20 µl of Elution buffer (EB) to the column.

19. Incubate for 1 minutes at room temperature

20. Centrifuge for 1 minute at 1.0 rcf

21. Centrifuge for 1 minute at 16.0 rcf

22. Nanodrop and store in -80°C

** Do samples separately! Get better results if samples are done separate and ** then combined and vacuumed.

D) Reverse Transcription

Superscript TM III

RT Reaction

1. Add the following to the 10ul of ligated RNA (want [ ] 2ug + xul dH20 = 10ul total)

GeneRacer OligoDT Primers	1ul
dNTP Mix	1ul
Sterile dH20	1ul

2. Incubate at 65°c for 5 min to remove and RNA secondary structure
3. Chill on ice for at least 1 min and centrifuge briefly
4. Add the following reagents to the 13ul ligated RNA and primer mixture:

5X First Strand Buffer	4ul
0.1 M DTT	1ul
RNaseOutTM (40U/ul)	1ul
SuperScriptTM III RT (200U/ul)	1ul
Total	20ul

5. Mix well by pipetting up and down gently
6. Centrifuge briefly and incubate at 50°c for 45 min
7. Inactivate the RT reaction at 70°c for 15 min
8. Chill on ice for 2 min and centrifuge briefly
9. Add 1ul of RNaseH (2U) to the reaction mix
10. Incubate at 37°c for 20 min

Centrifuge briefly and use immediately for amplification or store at -20°c

E) Standard Polymerase Chain Reaction

- *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

F) cDNA Amplification

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.

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.
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)*

* Add to the master mix just prior to use. Mix well by vortexing & spin down.

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 your downstream application requires long transcripts, *extend 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

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

shown:

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.

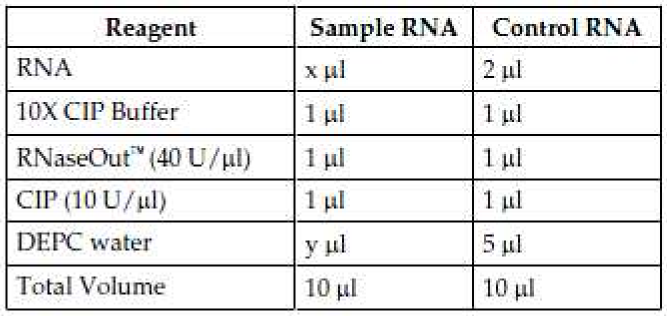
G) Rapid Amplification of cDNA Ends Polymerase Chain Reaction

Use 1-5 ug Total RNA

Dephosphorylating RNA

~ 1 Hour

1. Set up on ice the following 10 µl dephosphorylation reaction in a 1.5 ml sterile microcentrifuge tube using the reagents in the kit. Use 1-5 µg total RNA or 50-250 ng mRNA.



Reagent	Sample RNA	Control RNA
RNA	x µl	2 µl
10X CIP Buffer	1 µl	1 µl
RNaseOut™ (40 U/µl)	1 µl	1 µl
CIP (10 U/µl)	1 µl	1 µl
DEPC water	y µl	5 µl
Total Volume	10 µl	10 µl

2. Mix gently by pipetting and vortex briefly. Centrifuge to collect fluid.

3. Incubate at 50°C for 1 hour. After incubation, centrifuge briefly and place on ice.

Precipitating RNA

~45 minutes

1. To precipitate RNA, add 90 µl DEPC water and 100 µl phenol:chloroform and vortex vigorously for 30 seconds.

2. Centrifuge at maximum speed for 5 minutes at room temperature.
3. Transfer aqueous (top) phase to a new microcentrifuge tube (~100 µl).
4. Add 2 µl 10 mg/ml mussel glycogen, 10 µl 3 M sodium acetate, pH 5.2, and mix well. Add 220 µl 95% ethanol and vortex briefly.
5. Freeze on dry ice for 10 minutes. You may proceed to the next step or store at -20°C overnight.
   - **Note**: Do not store the RNA in DEPC water. Store RNA in ethanol at -20°C.
6. To pellet RNA, centrifuge at maximum speed in a microcentrifuge for 20 minutes at +4°C.
7. Note the position of the pellet and remove the supernatant by pipet. Be careful not to disturb pellet.
8. Add 500 µl 70% ethanol, invert several times, and vortex briefly.
9. Centrifuge at maximum speed in a microcentrifuge for 2 minutes at +4°C.
10. Note the position of the pellet and carefully remove the ethanol using a pipet. Centrifuge again to collect remaining ethanol.
11. Carefully remove the remaining ethanol by pipet and air-dry the pellet for 1-2 minutes at room temperature.
12. Resuspend the pellet in 7 µl DEPC water. If you want to check the stability of RNA after the CIP reaction, resuspend the pellet in 8 µl DEPC water and analyze 1 µl by agarose gel electrophoresis. Proceed to **Removing the mRNA Cap Structure**.

Decapping Reaction

~1 hour

1. Set up on ice the 10 µl decapping reaction in a 1.5 ml sterile microcentrifuge tube using the reagents in the kit.

Dephosphorylated RNA	7 µl
10X TAP Buffer	1 µl
RNaseOut™ (40 U/µl)	1 µl
TAP (0.5 U/µl)	1 µl
Total Volume	10 µl

2. Mix gently by pipetting and vortex briefly. Centrifuge briefly to collect fluid.
3. Incubate at 37°C for 1 hour.
4. After incubation, centrifuge briefly and place on ice.
5.

Precipitating RNA

~45 Minutes

1. To precipitate RNA, add 90 µl DEPC water and 100 µl phenol:chloroform and vortex vigorously for 30 seconds.
2. Centrifuge at maximum speed in a microcentrifuge for 5 minutes at room temperature.
3. Transfer aqueous (top) phase to a new microcentrifuge tube (~100 µl).
4. Add 2 µl 10 mg/ml mussel glycogen, 10 µl 3 M sodium acetate, pH 5.2, and mix well. Add 220 µl 95% ethanol and vortex briefly.

5. Freeze on dry ice for 10 minutes. You may proceed to the next step or store at -20°C overnight.
    - **Note**: Do not store the RNA in DEPC water. Store RNA in ethanol at -20°C.
6. To pellet RNA, centrifuge at maximum speed in a microcentrifuge for 20 minutes at +4°C.
7. Note the position of the pellet and remove the supernatant by pipet. Be careful not to disturb pellet.
8. Add 500 µl 70% ethanol, invert several times, and vortex briefly.
9. Centrifuge at maximum speed in a microcentrifuge for 2 minutes at +4°C.
10. Note the position of the pellet and carefully remove the ethanol using a pipet. Centrifuge again to collect remaining ethanol.
11. Carefully remove the remaining ethanol by pipet and air-dry the pellet for 1-2 minutes at room temperature.
12. Resuspend the pellet in 7 µl DEPC water. If you want to check the stability of RNA after the CIP reaction, resuspend the pellet in 8 µl DEPC water and analyze 1 µl by agarose gel electrophoresis. Proceed directly to **Ligating the RNA Oligo to Decapped mRNA.**
13.

Ligation Reaction

~ 1 hour 10 minutes

1. Add 7 µl of dephosphorylated, decapped RNA to the tube containing the pre-aliquoted, lyophilized GeneRacer™ RNA Oligo (0.25 µg). Pipet up and down

several times to mix and resuspend RNA Oligo. Centrifuge briefly to collect the fluid in the bottom of the tube.

2. Incubate at 65°C for 5 minutes to relax the RNA secondary structure.
    - **Note**: After the incubation, the total volume of this solution may decrease by 1 μl due to evaporation.
3. Place on ice to chill (~2 minutes) and centrifuge briefly.
4. Add the following reagents to the tube, mix gently by pipetting, and centrifuge briefly.

10X Ligase Buffer	1 μl
10 mM ATP	1 μl
RNaseOut™ (40 U/μl)	1 μl
T4 RNA ligase (5 U/μl)	1 μl
Total Volume	10 μl

5. Incubate at 37°C for 1 hour. Centrifuge briefly and place on ice. Precipitate the RNA.

Precipitating RNA

~45 minutes

1. To precipitate RNA, add 90 μl DEPC water and 100 μl phenol:chloroform and vortex vigorously for 30 seconds.
2. Centrifuge at maximum speed in a microcentrifuge for 5 minutes at room temperature.

3. Transfer aqueous (top) phase to a new microcentrifuge tube (~100 µl).
4. Add 2 µl 10 mg/ml mussel glycogen, 10 µl 3 M sodium acetate, pH 5.2, and mix well. Add 220 µl 95% ethanol and vortex briefly.
5. Freeze on dry ice for 10 minutes. You may proceed to the next step or store at -20°C overnight.
    - **Note**: Do not store the RNA in DEPC water. Store RNA in ethanol at -20°C.
6. To pellet RNA, centrifuge at maximum speed in a microcentrifuge for 20 minutes at +4°C.
7. Note the position of the pellet and remove the supernatant by pipet. Be careful not to disturb pellet.
8. Add 500 µl 70% ethanol, invert several times, and vortex briefly.
9. Centrifuge at maximum speed in a microcentrifuge for 2 minutes at +4°C.
10. Note the position of the pellet and carefully remove the ethanol using a pipet. Centrifuge again to collect remaining ethanol.
11. Carefully remove the remaining ethanol by pipet and air-dry the pellet for 1-2 minutes at room temperature.
12. Resuspend the pellet in 7 µl DEPC water. If you want to check the stability of RNA after the CIP reaction, resuspend the pellet in 8 µl DEPC water and analyze 1 µl by agarose gel electrophoresis. Proceed to **Reverse Transcribing mRNA**.

Reverse Transcribing mRNA (Superscript III RT Reaction)

~ 2 Hours

1. Add the following to the 10 µl of ligated RNA from Step 12:

Primers	1 µl
dNTP Mix	1 µl
Sterile, distilled water	1 µl

2. Incubate a**t 65°C for 5 minutes to rem**ove any RNA secondary structure.
3. Chill on ice for at least 1 minute and centrifuge briefly.
4. Add the following reagents to the 13-µl ligated RNA and primer mixture:

5X First Strand Buffer	4 µl
0.1 M DTT	1 µl
RNaseOut™ (40 U/µl)	1 µl
SuperScript™ III RT (200 U/µl)	1 µl
Total Volume	20 µl

5. Mix well by pipetting gently up and down.
    - **Note**: If you are using random primers, incubate the reaction mix at 25°C for 5 minutes prior to Step 6 to allow efficient binding of the random primers to the template.
6. Centrifuge briefly and incubate at 50°C for 30-60 minutes. If you are using gene-specific primers, increase the reaction temperature to 55°C.
7. Inactivate the RT reaction at 70°C for 15 minutes. Chill on ice for 2 minutes and centrifuge briefly at maximum speed in a microcentrifuge.

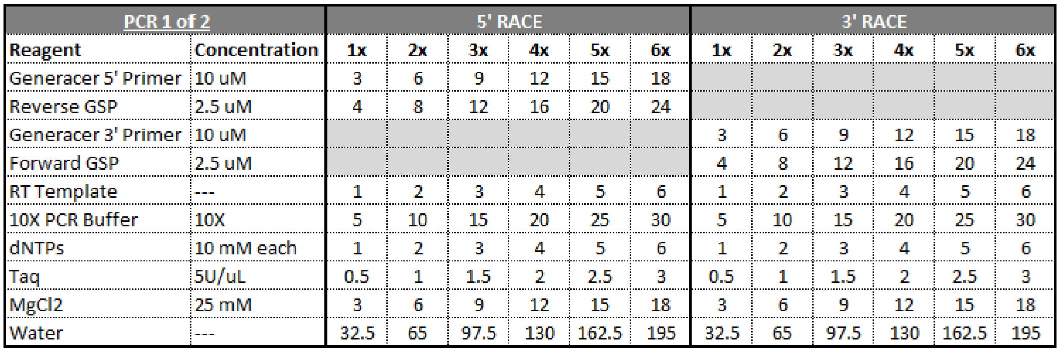
8. Add 1 µl of RNase H (2 U) to the reaction mix.

9. Incubate at 37°C for 20 minutes.

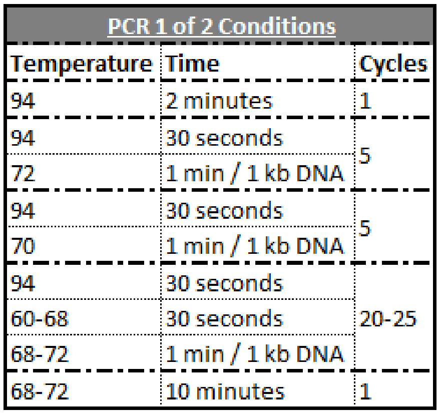
10. Centrifuge briefly and use immediately for amplification or store at -20°C.

You may use up to 2 µl of the RT reaction in each PCR reaction.

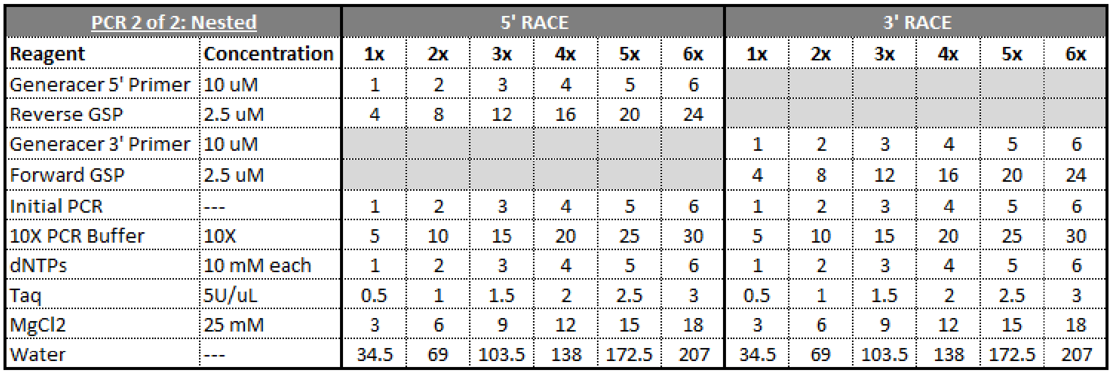
Nested PCRs with GSPs



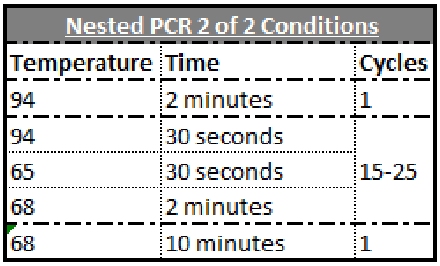
PCR 1 of 2		5' RACE						3' RACE					
Reagent	**Concentration**	**1x**	**2x**	**3x**	**4x**	**5x**	**6x**	**1x**	**2x**	**3x**	**4x**	**5x**	**6x**
Generacer 5' Primer	10 uM	3	6	9	12	15	18						
Reverse GSP	2.5 uM	4	8	12	16	20	24						
Generacer 3' Primer	10 uM							3	6	9	12	15	18
Forward GSP	2.5 uM							4	8	12	16	20	24
RT Template	---	1	2	3	4	5	6	1	2	3	4	5	6
10X PCR Buffer	10X	5	10	15	20	25	30	5	10	15	20	25	30
dNTPs	10 mM each	1	2	3	4	5	6	1	2	3	4	5	6
Taq	5U/uL	0.5	1	1.5	2	2.5	3	0.5	1	1.5	2	2.5	3
MgCl2	25 mM	3	6	9	12	15	18	3	6	9	12	15	18
Water	---	32.5	65	97.5	130	162.5	195	32.5	65	97.5	130	162.5	195



PCR 1 of 2 Conditions		
Temperature	**Time**	**Cycles**
94	2 minutes	1
94	30 seconds	5
72	1 min / 1 kb DNA	
94	30 seconds	5
70	1 min / 1 kb DNA	
94	30 seconds	20-25
60-68	30 seconds	
68-72	1 min / 1 kb DNA	
68-72	10 minutes	1



PCR 2 of 2: Nested		5' RACE						3' RACE					
Reagent	**Concentration**	**1x**	**2x**	**3x**	**4x**	**5x**	**6x**	**1x**	**2x**	**3x**	**4x**	**5x**	**6x**
Generacer 5' Primer	10 uM	1	2	3	4	5	6						
Reverse GSP	2.5 uM	4	8	12	16	20	24						
Generacer 3' Primer	10 uM							1	2	3	4	5	6
Forward GSP	2.5 uM							4	8	12	16	20	24
Initial PCR	---	1	2	3	4	5	6	1	2	3	4	5	6
10X PCR Buffer	10X	5	10	15	20	25	30	5	10	15	20	25	30
dNTPs	10 mM each	1	2	3	4	5	6	1	2	3	4	5	6
Taq	5U/uL	0.5	1	1.5	2	2.5	3	0.5	1	1.5	2	2.5	3
MgCl2	25 mM	3	6	9	12	15	18	3	6	9	12	15	18
Water	---	34.5	69	103.5	138	172.5	207	34.5	69	103.5	138	172.5	207



Nested PCR 2 of 2 Conditions

Temperature	Time	Cycles
94	2 minutes	1
94	30 seconds	15-25
65	30 seconds	
68	2 minutes	
68	10 minutes	1

H) Primer Dilution from IDT

To make freezer stock:

1. Spin down tubes

2. Add dH$_2$0 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℃*

5. Note primer location on Primer Inventory Sheet

To make 10 µM PCR Stock:

100 µl of 50 µM primer

400 µl of dH$_2$0

Store at -20℃

To make 2.5 µM working PCR stock

125 µl of 10 µM primer

375 µl of dH$_2$0

Store at -20℃

I) Gel Extraction 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 H_2O to the center of the QIAquick membrane and centrifuge the column for 1 min at maximum speed.

14.) Reapply the flow-through and centrifuge again for 1 min.

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

Store samples at -20℃

J) Sequencing

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).

Sample Analysis:

Sequencing on the ABI 3130xl genetic analyzer is routinely conducted using POP7 polymer, a 50 cm. 16-capillary array and the KB Basecaller software. These conditions normally produce high quality sequence that extends to 800-1,000 bases.

PCR products less than 900 bp in length will be analyzed in the 3130xl using an analytical protocol that looks for the end of the raw data. Please identify your PCR product and its size on the Submission Form so this protocol may be specified during instrument setup.

K) RNA-Sequencing Protocols

Protocol for Gene Ontology (GO) Analysis

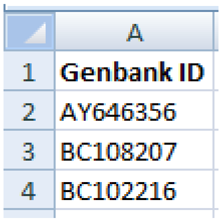
- Logon to the DAVID database (http://david.abcc.ncifcrf.gov/)
- Select “Start Analysis” from top menu
- Copy transcript list of interest into the “Paste a list” box
        - Use Genbank IDs from RNA-Seq study
    - For the “Select Identifier” box, choose appropriate type of sample accession submitted
        - 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

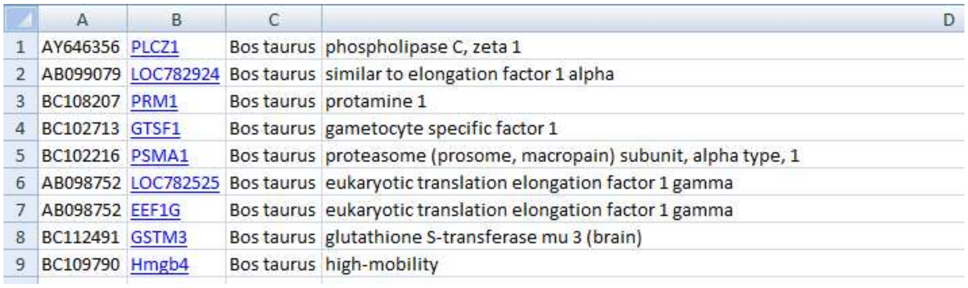
Pairing gene symbols & long names with known accessions

- Open file that you’re annotating these names onto, referred to here as “FILE 1”



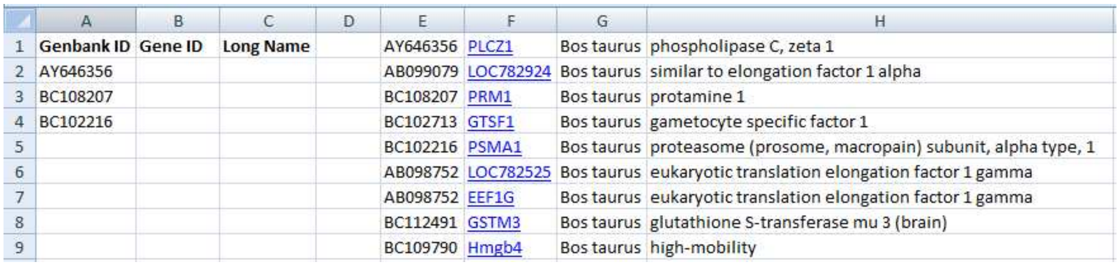
	A
1	Genbank ID
2	AY646356
3	BC108207
4	BC102216

- Open file with accessions/official gene IDs/long names in it, referred to here as "FILE 2"



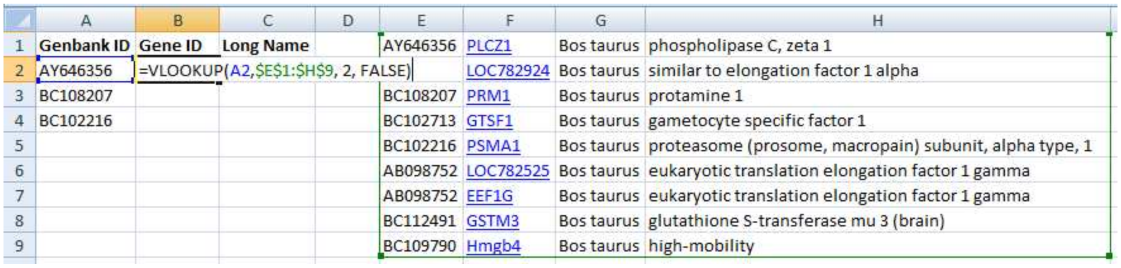
	A	B	C	D
1	AY646356	PLCZ1	Bos taurus	phospholipase C, zeta 1
2	AB099079	LOC782924	Bos taurus	similar to elongation factor 1 alpha
3	BC108207	PRM1	Bos taurus	protamine 1
4	BC102713	GTSF1	Bos taurus	gametocyte specific factor 1
5	BC102216	PSMA1	Bos taurus	proteasome (prosome, macropain) subunit, alpha type, 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)"

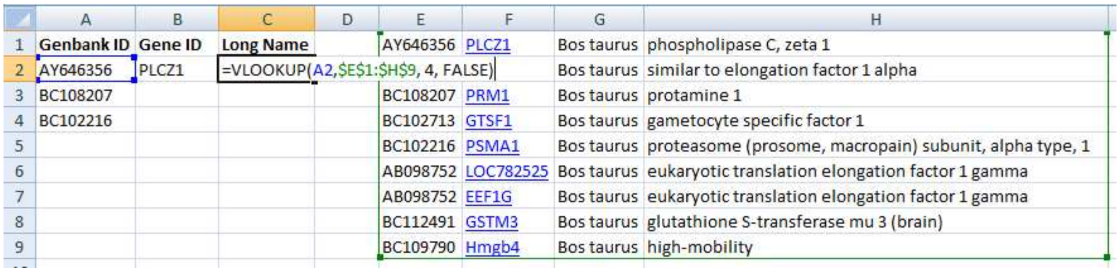


	A	B	C	D	E	F	G	H
1	**Genbank ID**	**Gene ID**	**Long Name**		AY646356	PLCZ1	Bos taurus	phospholipase C, zeta 1
2	AY646356	=VLOOKUP(A2,$E$1:$H$9, 2, FALSE)				LOC782924	Bos taurus	similar to elongation factor 1 alpha
3	BC108207				BC108207	PRM1	Bos taurus	protamine 1
4	BC102216				BC102713	GTSF1	Bos taurus	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

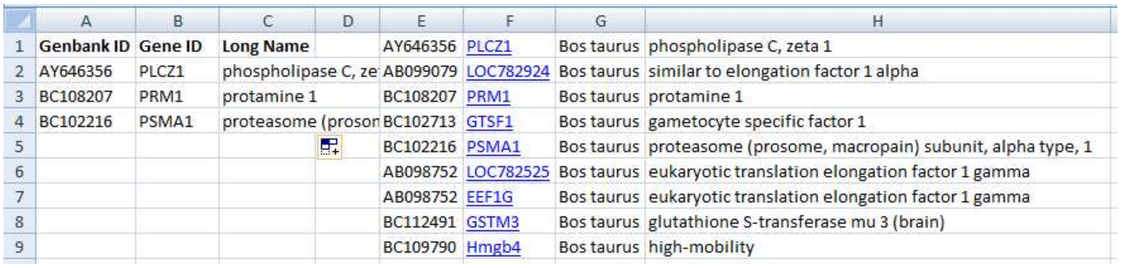
  - $E$1:$H$100 highlighting the entire table you're taking information from, with $ signs added to lock the entire thing in place for when you start dragging it
  - 2 column number <u>in the table you're searching </u>in containing the information you're looking for
  - FALSE to tell it to search for 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.



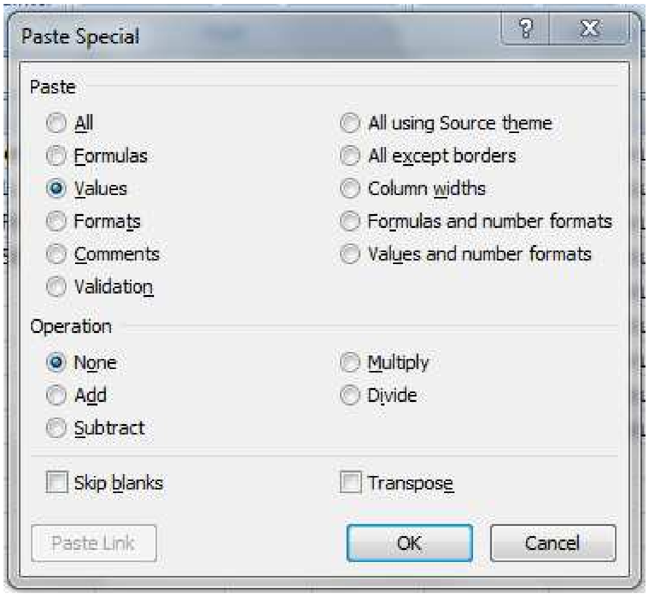
	A	B	C	D	E	F	G	H
1	**Genbank ID**	**Gene ID**	**Long Name**		AY646356	PLCZ1	Bos taurus	phospholipase C, zeta 1
2	AY646356	PLCZ1	=VLOOKUP(A2,$E$1:$H$9, 4, FALSE)				Bos taurus	similar to elongation factor 1 alpha
3	BC108207				BC108207	PRM1	Bos taurus	protamine 1
4	BC102216				BC102713	GTSF1	Bos taurus	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 (prosom		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


Paste Special

Paste

○ All
○ Formulas
● Values
○ Formats
○ Comments
○ Validation
○ All using Source theme
○ All except borders
○ Column widths
○ Formulas and number formats
○ Values and number formats

Operation

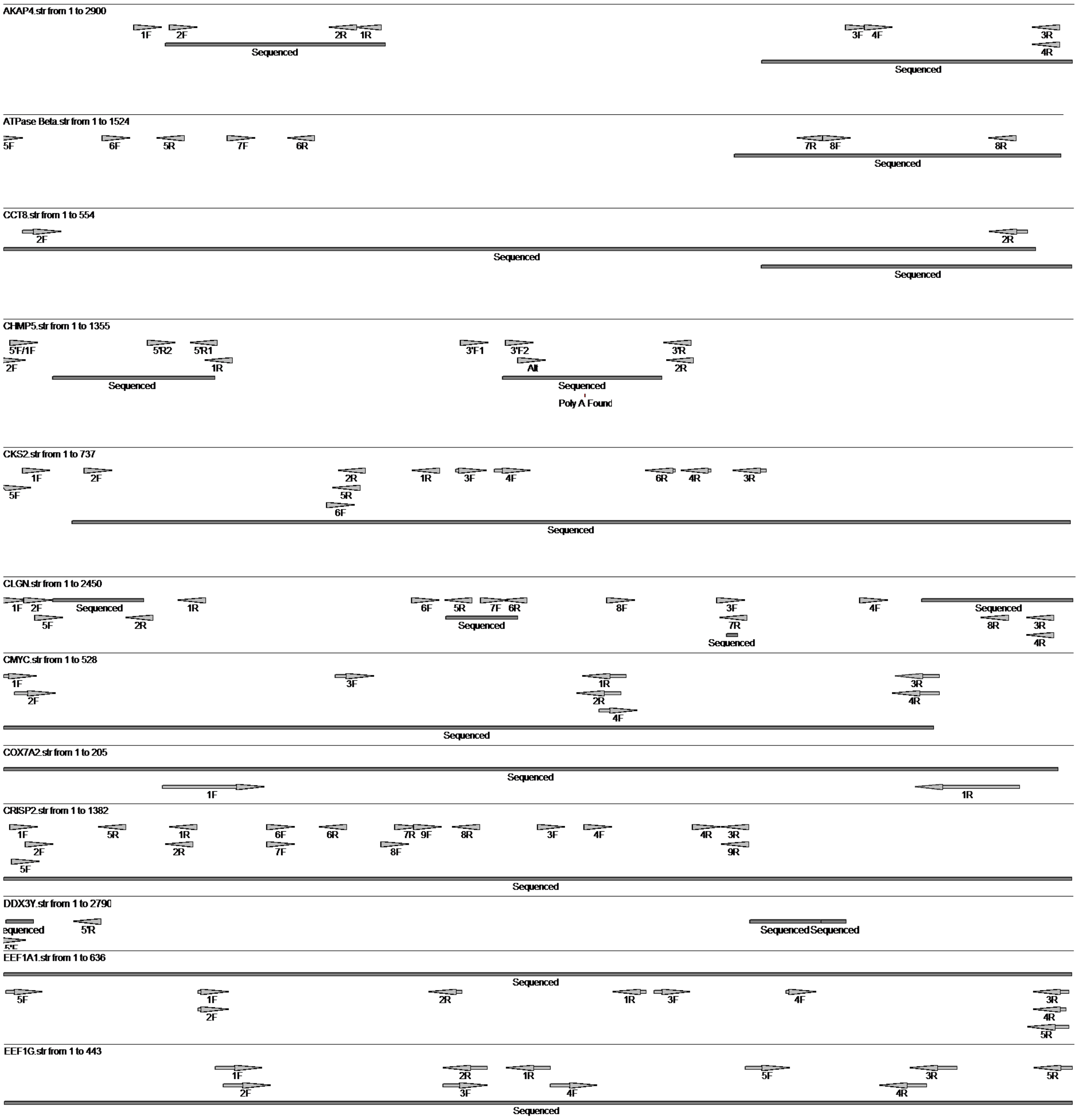
● None
○ Add
○ Subtract
○ Multiply
○ Divide

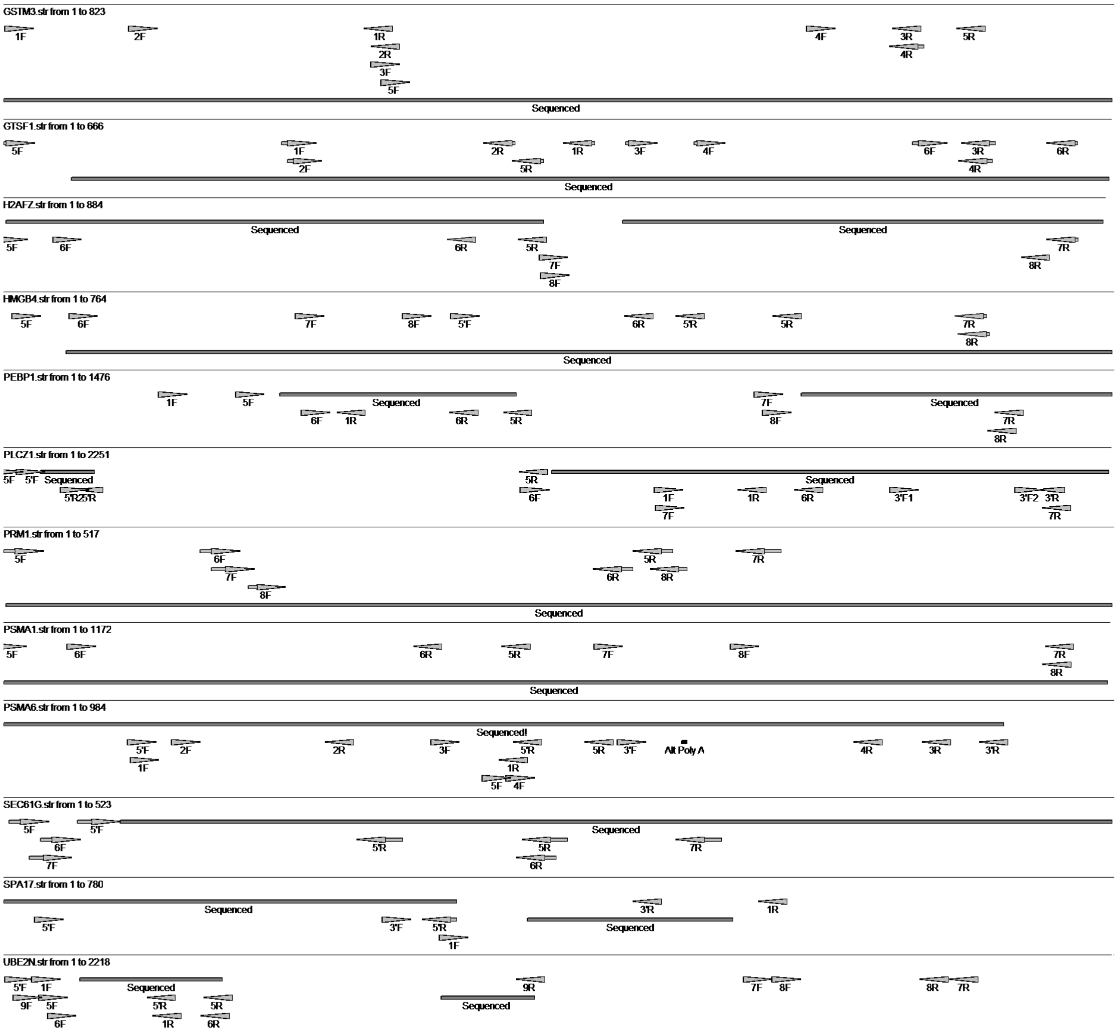
☐ Skip blanks
☐ Transpose

- Delete reference table, leaving just the annotated original file.

APPENDIX II: RAW APE FILES USED FOR CHAPTER 2

All primer sets here are labeled with the names used on the tubes as used in the laboratory. Primer titles were changed to a linear organization for chapter 2 to assist with cohesion of the work. No primer sequences were changed during this process. Below, the full genes with primer locations and sequencing locations are shown. The top bar on each represents the total length of the reference accession number.




GSTM3.str from 1 to 823
1F
2F
1R
2R
3F
5F
4F
3R
4R
5R
Sequenced
GTSF1.str from 1 to 666
5F
1F
2F
2R
5R
1R
3F
4F
6F
3R
4R
6R
Sequenced
H2AFZ.str from 1 to 884
Sequenced
Sequenced
5F
6F
6R
5R
7F
8F
7R
8R
HMGB4.str from 1 to 764
5F
6F
7F
8F
5'F
6R
5'R
5R
7R
8R
Sequenced
PEBP1.str from 1 to 1476
1F
5F
Sequenced
6F
1R
6R
5R
7F
8F
Sequenced
7R
8R
PLCZ1.str from 1 to 2251
5F
5'F
Sequenced
5'R2
5'R
5R
6F
Sequenced
1F
7F
1R
6R
3'F1
3'F2
3'R
7R
PRM1.str from 1 to 517
5F
6F
7F
8F
5R
6R
8R
7R
Sequenced
PSMA1.str from 1 to 1172
5F
6F
6R
5R
7F
8F
7R
8R
Sequenced
PSMA6.str from 1 to 984
Sequenced!
5'F
1F
2F
2R
3F
5'R
1R
5F
4F
5R
3'F
Alt Poly A
4R
3R
3'R
SEC61G.str from 1 to 523
5F
6F
7F
5'F
Sequenced
5'R
5R
6R
7R
SPA17.str from 1 to 780
Sequenced
5'F
3'F
5'R
1F
Sequenced
3'R
1R
UBE2N.str from 1 to 2218
5'F
1F
9F
5F
6F
Sequenced
5'R
1R
5R
6R
9R
Sequenced
7F
8F
8R
7R