1-1-2011

Systems Biology Modeling of Bovine Fertility using Proteomics

Divya swetha Peddinti

SYSTEMS BIOLOGY MODELING OF BOVINE FERTILITY USING PROTEOMICS

By

Divya swetha Peddinti

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Veterinary Medical Sciences
in the College of Veterinary Medicine

Mississippi State, Mississippi

April 2011

SYSTEMS BIOLOGY MODELING OF BOVINE FERTILITY USING PROTEOMICS

By

Divya swetha Peddinti

Approved:

Shane C. Burgess
Associate Dean and Professor of
CVM Basic Sciences
Director of Life Sciences and
Biotechnology Institute,
Co-director of Institute for Digital Biology
(Major Professor and Director of Dissertation)

Bindu Nanduri
Assistant Professor of CVM
Basic Sciences
(Co-major Professor)

Erdogan Memili
Associate Professor of Animal and
Dairy Sciences (Committee member)

Fiona M. McCarthy
Assistant Professor of CVM
Basic Sciences (Committee member)

Susan M. Bridges
Professor of Computer Science
and Engineering
Co-director of Institute for Digital
Biology (Committee member)

Larry A. Hanson
Professor of CVM
Basic Sciences
(Graduate Coordinator of CVM Basic
Sciences Department)

Kent H. Hoblet
Dean of the College of Veterinary Medicine
Mississippi State University

Name: Divya swetha Peddinti

Date of Degree: April 29, 2011

Institution: Mississippi State University

Major Field: Veterinary Medical Sciences

Major Professor: Dr. Shane C. Burgess

Title of the Study: SYSTEMS BIOLOGY MODELING OF BOVINE FERTILITY USING PROTEOMICS

Pages in Study: 105

Candidate for Degree of Doctor of Philosophy

Beef and milk production industries represent the largest agricultural industries in the United States with a retail equivalent value of approximately $112 billion (USDA, 2008). Infertility is the major problem for mammalian reproduction. In the United States approximately 66% of cows are bred by Artificial Insemination (AI), but only ~50% of these inseminations result in successful pregnancies. Infertility can occur either from male factor (spermatozoon) or female factor (oocyte) and male contributes approximately 40% of cases. Infertility costs the producer approximately $5 per exposed cow for every 1% reduction in pregnancy rate. In spite of its millions of dollars in economic impact, the precise molecular events/mechanisms that determine the fertilizing potential of an oocyte and spermatozoon are not well defined. The thesis of my doctoral dissertation is that proteomics-based “systems biology” modeling of bovine oocyte and spermatozoon can facilitate rapid understanding of fertility. To test this thesis, I needed to first identify the proteins associated with bovine oocyte and its associated cumulus cells, and

spermatozoon. The next step was functional annotation of the experimentally confirmed proteins to identify the major functions associated with the oocyte, cumulus cells and spermatozoon, and finally, generate a proteomics based systems biology model of bovine oocyte and cumulus cell communication and male fertility.

The results of my dissertation established the methods that provide afoundation for high-throughput proteomics approaches of bovine oocyte and cumuluscell biology and allowed me to model the intricate cross communication between oocyte and cumulus cells using systems biology approaches. Proteomics based systems biology modeling of oocytes and cumulus cells identified the signaling pathways and proteins associated with this communication that may have implications in oocyte maturation. In addition, systems biology modeling of differential spermatozoa proteomes from bulls of varying fertility rates enabled the identification of putative molecular markers and key pathways associated with male fertility. The ultimate positive impact of these results is to facilitate the field of biomedical research with useful information for comparative biology, better understanding of bovine oocyte and spermatozoon development, infertility, biomarker discovery, and eventually development of therapies to treat infertility in bovine as well as humans.

Keywords: Bovine fertility, bovine oocytes and spermatozoa, Gene Ontology, proteomics, GO annotation quality, systems biology

DEDICATION

I would like to dedicate this research to my parents Renuka and Satyanarayana Reddy Peddinti, my husband Pradeep kumar Reddy Dumpala, and my dear son Rohan Reddy Dumpala.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Dr. Shane Burgess, my mentor, for giving me the opportunity to pursue my PhD in his laboratory. Throughout my doctoral program and in the dissertation process, Dr. Burgess provided unflinching encouragement, intellectual freedom, undue credit for my hard work, and support. Pursuing a doctoral program under Dr. Burgess mentorship was truly a productive and rewarding experience.

In addition, I would like to extend my gratitude to my co-major advisor Dr. Bindu Nanduri for her professional guidance, encouragement, moral support, and critical review of my research. My deep gratitude is also extended to Dr. Erdogan Memilifor his professional guidance. My sincere appreciation is also due to the other members of my PhD committee, namely Dr. Susan M. Bridges, and Dr. Fiona M. McCarthy for introducing me to their area of expertise, their guidance, support, invaluable stimulating discussions, and their hard work in critical reviewing of this dissertation. Their valuable recommendations and suggestions have been very useful for the completion of research described in this dissertation.

My sincere gratitude is also extended to the Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University for providing me with all resources for this research.

I would like to thank Allen Shack for his support with proteomics experiments. I would also like to thank Tibor Pechan and Ken Pendarvis of the MSU Life Sciences and Biotechnology Institute for proteomics help. I would like to take this opportunity to acknowledge my fellow Agbase Biocurators and graduate students, Teresia, Lakshmi, Dusan, Bart, Ranjit, and Shyamesh for their friendship, support, and encouragements.

Lastly my endless gratitude belongs to my dear husband Pradeep Dumpala, my dear son Rohan Dumpala, my parents Satyanarayana Reddy Peddinti, Renuka Peddinti, and my sister Amrutha Ramya Peddinti. Their devoted support, care, love, patience, and encouragements have been essential for my personal development and helped me to complete my Ph.D. studies

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

GENERAL INTRODUCTION AND REVIEW OF PERTINENT LITERATURE

Cattle industry and infertility

The United States is the world's largest producer of beef and cow milk. Beef and milk production industries represent the largest agricultural industries in the United States with a retail equivalent value of approximately $112 billion (USDA, 2008). Artificial insemination (AI), a common breeding technique used in food animals, is one of the most effective genetic improvement tools available today. In the United States, approximately 66% of cows are bred by AI, but only 50% of these inseminations result in successful pregnancies (Killian, 1999). The economic viability of beef operations is greatly affected by the failure of breeding females to become pregnant. However, many producers fail to realize the impact of infertility, which costs approximately $5 per exposed cow for every 1% reduction in pregnancy rate (Cliff Lamb et al., 2008).

Previously the oocyte alone was thought to be responsible for regulation of the earliest events of embryo formation and the spermatozoon was considered to be the delivery vehicle of the paternal genetic complement to the ovum (Wassarman and Litscher, 2001; Zeng et al., 2004). However, sperm contribute more than just their DNA (Sutovsky and Schatten, 2000); they deliver their entire structure on fertilization, including RNAs and other sperm derived factors (Ostermeier et al., 2002; Krawetz,

2005). Therefore, both the oocyte and the spermatozoon are essential for early embryo development.

It is thus essential to have healthy oocytes and spermatozoa for artificial insemination (AI) to result in successful pregnancies. In addition, assisted reproductive technologies (ART) such as *in vitro* production of embryos, somatic cell nuclear transfer (SCNT), and embryonic stem cells require developmentally competent oocytes. In spite of millions of dollars in economic impact, the precise molecular events and mechanisms that determine the fertilizing potential of an oocyte and spermatozoa are not well defined. The thesis of my doctoral dissertation is that proteomics-based "systems biology" modeling of bovine oocyte and spermatozoon can facilitate rapid understanding of fertility

Male and Female fertility

Infertility is a major problem for mammalian reproduction, and can occur either from the male (spermatozoon) or female (oocyte) factors. The nature of sub fertility due to the male is as complex as that of the female and is responsible for approximately 40% of the infertility cases. Therefore it is important to understand factors that affect both male and female fertility.

Although most basic reproductive biology work is done in the mice (Eppig et al., 1993), significant species differences in oocyte and spermatozoa biology exist between human and mice (Sutton et al., 2003).The bovine model is not only economically-important worldwide, especially in many developing nations, but it is also a relevant model for human fertility studies because oocyte biology, many aspects of ovarian

follicular dynamics, and spermatozoa biology are more similar between humans and cows than they are between humans and mice (Bettegowda et al., 2008).

Female fertility is dependent in the oocyte's intrinsic developmental potential, which refers to the biochemical and molecular state of the oocyte that allows it to be matured and fertilized, and of the resulting zygote as it continues on through embryonic and fetal development until birth (Gilchrist et al., 2008). In accordance with this, poor oocyte quality results in polyspermy and/or arrested embryonic development or spontaneous abortion. An important practical reason to improve our understanding of the determinants of oocyte quality is to enhance the clinical implementation of oocyte *in vitro* maturation (IVM). IVM is an important ART that enables oocytes to be matured *in vitro* from ovaries that have received either no or low levels of gonadotrophin stimulation (Edwards, 1965; Smitz and Cortvrindt, 2004). There is potentially a great demand for IVM in clinical practice, as it has the potential to capture the vast supply of oocytes within a single ovary (Gilchrist et al., 2008). In domestic animals, embryo production from unstimulated ovaries using oocyte IVM is a routine practice and is an important platform technology for artificial breeding, cloning, and transgenic animal production. Despite the progress, the quality of oocytes matured *in vitro* is inferior to those produced *in vivo* (Sun and Nagai, 2003). Thus, understanding the mechanisms and factors that regulate oocyte maturation and the developmental potential or competence may help in improving the efficiency of clinical IVM and thereby provide new options for the treatment of infertility.

Male fertility is determined by the ability of the spermatozoon to fertilize the oocyte, and the resulting zygote continuing on through embryonic and fetal development

until birth (Eid et al., 1994; Saacke et al., 2000). AI utilizes semen from genetically superior sires to inseminate cows. In the United States more than ~70% of cows are bred by AI but only ~50% of these inseminations result in successful full term pregnancies (Killian, 1999). The underlying molecular events and mechanisms that determine the fertilizing potential of a semen sample are not well defined. A thorough understanding of these mechanisms is essential for obtaining consistently high reproductive efficiencyand to ensure lower cost and time-loss by the breeder.

The last decade's watershed developments of whole genome sequences for humans as well as the cow, biotechnologies for quantifying genes and their products, and the emergence of quantitative computational analysis of genetic networks (sometimes called "systems, network or integrative biology") enables high-throughput "post genomic" technologies to be applied to understand the molecular basis of reproduction and its pathologies. High-throughput DNA and mRNA measure genetic potential and transcription, respectively; protein expression, locations, isoforms (produced by alternative splicing and post-translational modifications), functional protein interactions with other proteins and other biomolecules and the resulting signaling cascades, are measured using proteomics technologies (Lee, 2001). Many reproductive diseases, or suboptimal economic performance, occur at the level of the proteome e.g. endometriosis, may result from aberrant protein expression and localization (Brewis, 1999).

Proteomics in reproductive research

This section of my introduction is extracted directly from: Proteomics in animal reproduction and breeding. D. Peddinti, E. Memilli, S.C. Burgess. Book chapter in Methods in Animal Proteomics, in press, 2010.

Developments in assisted reproduction, e.g. *in vitro* oocyte maturation (IVM), *in vitro* fertilization (IVF), intra cytoplasmic sperm injection (ICSI), animal cloning and embryo culture in cattle, pigs, sheep, and horses are continuing at considerable rate (Ellederova et al., 2004). The driving forces behind advanced reproductive technologies in traditional agricultural animals are genetic selection for improved production and reproductive traits as well as these animals's utility for biomedicine. Although understanding the molecular mechanisms of reproduction (e.g. for studying gametes and embryos and for allowing the manipulation of early development in agricultural animals) is considered paramount for rational advances in assisted reproduction techniques (Ellederova et al., 2004; Galli and Lazzari, 2008), technical achievements in assisted reproduction have often outpaced our understanding of these molecular mechanisms.

Proteomics is beginning to expand our understanding of the mammalian oocyte, spermatozoon, and embryo. Until a decade ago, lack of sensitivity was a major impediment to the general use of proteomics for understanding mammalian reproduction. Although still yet to become as democratized (and therefore as ubiquitous) as other functional genomics methods such as microarray analysis, advances in mass spectrometry now enable "omic" elucidation of underlying biological processes at a protein level (Katz-Jaffe and Gardner, 2007). There has been an increase in the number of proteomic studies associated with reproductive research in the last decade (Figure 1) and this was especially pronounced from 2005 to 2006 and is likely associated with technical advances in high-throughput non-electrophoretic proteomics. Most of these reproductive proteomics studies have been done in human and mouse (Figure 2).

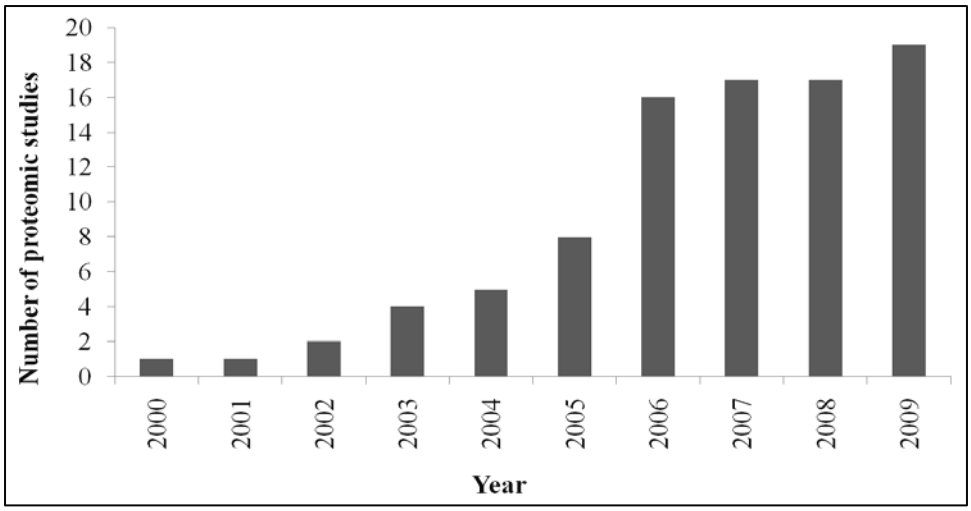

Figure 1.1 Growth of proteomics in reproductive research during the last decade.

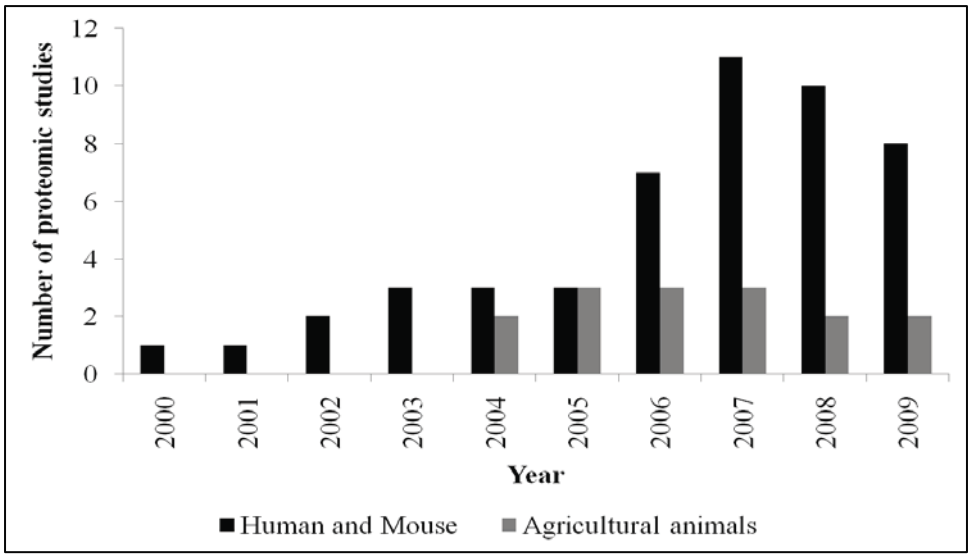

Figure 1.2 Comparison of number of proteomic studies in human and mouse vs. agricultural animal reproduction

The supply of human material for proteomics is limited and, though convenient, mice have small sex organs. Furthermore, significant species differences in oocyte and spermatozoon biology exist between humans and mice. Some of the limitations of mouse

models can be mitigated by using our well-studied agricultural species for reproductive proteomics and this is especially possible now that these species have sequenced genomes. Not only do agricultural animals have larger sex organs, but they have more similar reproductive biology (Bettegowda et al., 2008). One caveat though is that the functions of most agricultural animal proteins still remain unknown, and so functional annotation, both manually by biocuration from the literature and by using high-throughput bioinformatics methods is essential to understand how the proteome defines the phenotype in agricultural animal proteomics (McCarthy et al., 2006).In addition comparing conserved proteins and functional domains across different species, in addition to mice, provides evidence for essential evolutionarily conserved regions and functions of sperm and oocyte proteins (Oliva et al., 2009).

Regardless, understanding agricultural animal reproductive physiology is directly relevant to agricultural animal economics and also informative for human reproductive physiology. Here we review proteomics that has been applied to understanding the proteomes underlying male and female agricultural animal fertility.

Proteomic techniques

During the last decade, protein analysis and proteomics have become established tools (Reinders et al., 2004). One of the most widely used reductionist techniques for protein studies is Western blotting, which determines protein expression by using a specific antibody to bind to the protein of interest. Western blotting, however, is limited by access to, and the cost of, monoclonal antibodies and affinity purified sera, and only a few proteins can be tested in a single experiment. Regardless, although western blotting is considered by some to represent a “gold standard”, more often than not the

fundamental mechanisms of antibody affinity and avidity are ignored and good antibody specificity is assumed when, in the context of a whole genome's proteome, it may not exist (Michaud et al., 2003). Another fundamental limitation is the combination of western blotting's targeted nature and limitation to hypothesis testing: such experiments can find (or not find) only what they are designed to look for and unpostulated effects in different cellular compartments or cellular processes remain undiscovered. A proteomic approach conceived from traditional western blotting and utilizing the technologies available for nucleic acid hybridization is protein microarrays (Stoevesandt et al., 2009).

The past decade has seen mass spectrometry (MS) based-methods enable rapid and facile identification and quantification of proteins at a global level (Hachey and Chaurand, 2004; Guerrera and Kleiner, 2005). These methods have been reviewed extensively in the past decade, even for agricultural animal research (Sirard et al., 2003; Hachey and Chaurand, 2004; Fadiel et al., 2005) and will be introduced only briefly here. MS-based proteomics couples high-performance mass spectrometry instruments with multidimensional–electrophoretic (gel-based)or multidimensional-chromatographic non-electrophoretic (gel-free) methods for separation, followed by qualitative and quantitative analysis of complex proteome samples (Ferguson and Smith, 2003). These methods complement antibody-based methods and have the advantage that they are far less epitope-limited. They do, however, suffer from analogous issues with specificity. In this case, however, the specificity limitations of MS-based protein identification are due to mass spectral quality combined with the "search space" when searching databases with tandem mass spectra (Sevinsky et al., 2008).

2-D PAGE and 2-D DIGE can show not only basic information for individual protein identification and quantification but also have the advantage of monitoring post translational modifications. 2-D PAGE and 2-D DIGE are of limited use for highly hydrophobic, acidic and alkaline proteins due to the finite limits to the hydrophobicity, isoelectric point (pI), and molecular weight range (Sato et al., 2002; Ferguson and Smith, 2003) and abundant proteins will mask less-abundant proteins. Gel-based approaches require efficient recovery of the proteins or digested peptides from gels for MS identification but recovery may be less than 1% (Sato et al., 2002). Multidimensional-chromatographic non-electrophoretic proteomics approaches rely on generating proteome coverage using proteolytic cleavage (usually using trypsin) of protein samples, followed by multidimensional (i.e., ion exchange, size exclusion, reversed phase, and affinity) chromatography coupled to multidimensional mass spectrometry. Most 2D chromatographic separation techniques include strong cation exchange (SCX) coupled to reversed phase liquid chromatography (RPLC) which separate peptides by charge and then hydrophobicity. Advances in chemical isotope labeling techniques (e.g. isotope coded affinity tag (ICAT (Haqqani et al., 2008; Schmidt and Urlaub, 2009)) and isobaric tags for relative and absolute quantitation (iTRAQ (Schmidt and Urlaub, 2009)) enable the relative quantification between two or more samples. However, such labeling reagents do restrict the numbers of samples that can be compared, the numbers of peptides (and thus proteins) that can be identified as well as complicating inter-experimental comparisons. Label-free relative quantification methods also exist and are based on the relationship between protein abundance and sampling statistics such as the peptide count (Gao et al., 2003), spectral count (Liu et al., 2004), sequence coverage

(Florens et al., 2002), measurement of mass spectral peak intensities (Old et al., 2005) and sum of the mass spectra and theoretical peptide cross correlation scores (ΣXCorr) (Bridges et al., 2007). Notably, label-free relative quantitative approaches do either just as well, or often better, than stable isotope labeling approaches (Turck et al., 2007) and have the added advantages of better protein identification at lower costs.

Proteomics and female fertility

The oocyte is the female gamete and uniquely establishes the program of life after fertilization. Oogenesis (the production of oocytes) starts during the female's own early fetal development and in most mammals is complete by parturition (Yang et al., 1998). Oocytes undergo a series of developmentally-regulated changes before ovulation and fertilization and enter the meiotic cell cycle and arrest at the prophase I in mammalian fetal ovaries. The first change is the 'growth' phase during their arrested period at prophase I (Sirard et al., 1998; Telfer, 1998). During the growth phase, oocytes develop from primordial, primary, secondary, and tertiary follicles in the ovary and acquire the capacity to resume meiosis ('acquisition of meiotic competence') (Sirard et al., 1998). Fully grown immature oocytes enter the maturational phase, undergo germinal vesicle breakdown (GVBD), condense their chromosomes, realization of first meiosis (MI), and another arrest of development at the metaphase of the second meiosis (MII). This process is collectively called nuclear maturation (figure 3). In addition to these nuclear maturation events, significant cytoplasmic changes occur which are essential for future fertilization and initiation of zygote development, including structural changes of organelles and major translational activity when many new proteins are synthesized and synthesis of others is terminated, (Moor et al., 1990; Coenen et al., 2004). At ovulation,

oocytes are at the MII stage; upon stimulation by sperm, an oocyte resumes meiosis II, completes maturation and emits the second polar body (Sun and Nagai, 2003).

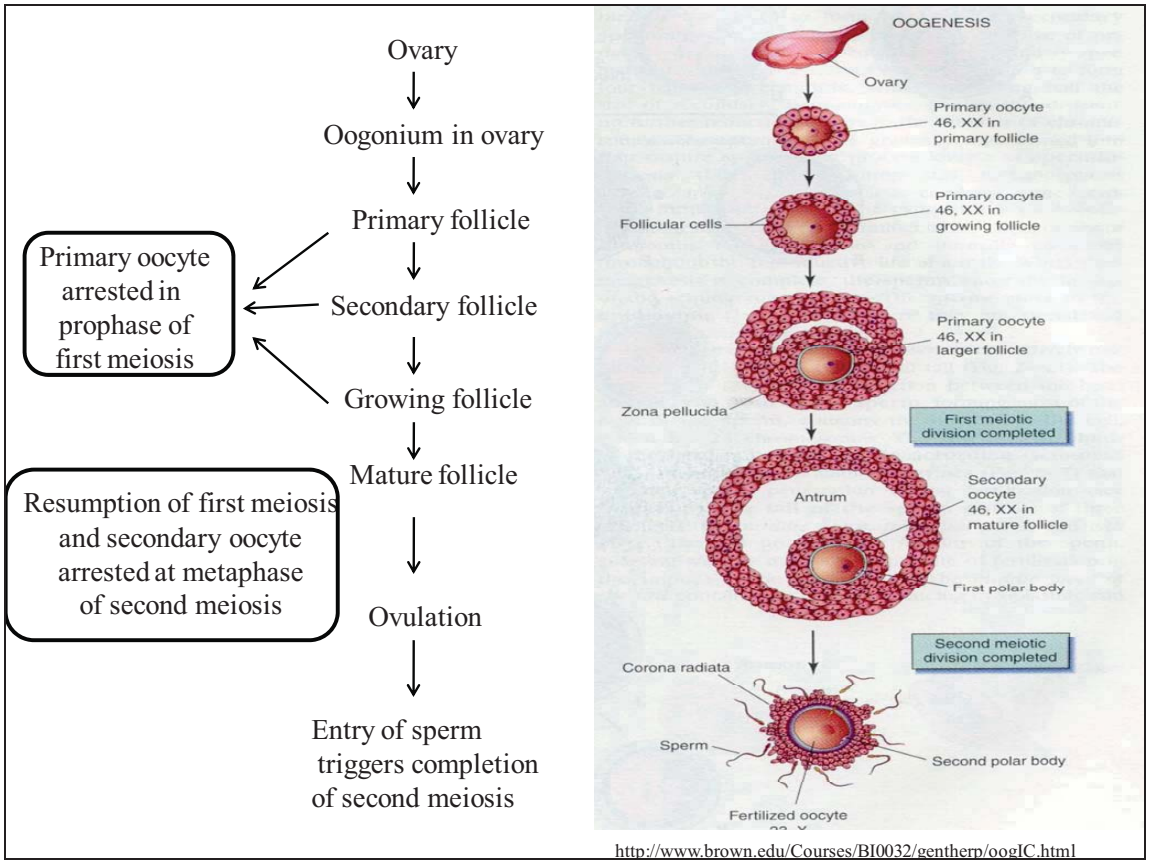

Figure 1.3 Flow chart describing the stages of oogenesis.

Oocyte transcriptional activity rapidly decreases during maturation (De La Fuente and Eppig, 2001; Tomek et al., 2002; De La Fuente et al., 2004; Ellederova et al., 2004; Bhojwani et al., 2006); however, translation of the mRNA pool continues throughout the final stages of meiosis (Heikinheimo and Gibbons, 1998). During the growing phase, oocytes accumulate a stockpile of mRNAs and ribosomes which will be used later, for example, to sustain the completion of meiosis or to promote early embryonic

development prior to the maternal-to-embryonic transition (MET) (De La Fuente et al., 2004). In many species, initial protein synthesis utilizes transcripts that are synthesized and stored during oogenesis (Richter, 1991). There are several mechanisms that determine which transcripts are translated during oogenesis and which are not. The principle mechanisms involved in the time-specific translation of these stored mRNAs is cytoplasmic polyadenylation/deadenylation (Richter, 1999). Cytoplasmic polyadenylation is believed to be mediated by cytoplasm polyadenylation element binding protein binding to the cytoplasm polyadenylation element (CPE), which is a U-rich *cis*-element present in the 3'-untranslated region (3' UTR) of a number of mRNAs such as *c-mos*, cdk2, and cyclin-B (Richter, 1999; Uzbekova et al., 2008). In addition to polyadenylation, translation of some transcripts depends on poly (A) tail length: increased poly (A) tail length generally correlates with translational activation, whereas decreased poly (A) tail length generally correlates with repression (Richter, 1991; Wickens et al., 1997).

Previous studies have lead to our understanding that oocyte maturation involves the coordinated action of several protein kinases. Mitogen-activated protein kinases (MAPKs)and maturation-promoting factor (MPF) (a complex of cyclin B1 and cell-cycle controller p34 kinase cyclin-dependent kinase-2) constitute the main signaling pathways of oocyte maturation (Motlik and Kubelka, 1990; Motlik et al., 1998). Two isoforms of MAPKs, ERK1 (p44) and ERK2 (p42), express widely in mammalian oocytes and play a pivotal role in meiosis (Motlik et al., 1998; Sun et al., 1999). MAPK exists in an inactive form in the GV stage oocytes and is activated around GVBD (Inoue et al., 1995; Fissore et al., 1996; Lee et al., 2000; Sun et al., 2001; Sun et al., 2001; Wehrend and Meinecke,

2001). After GVBD, MAPK is involved in the regulation of microtubule organization and meiotic spindle assembly. Inhibition of MAPK activity during MI to MII transition results in the failure of first polar body emission and formation of MII spindle (Fan and Sun, 2004). In all mammalian oocytes, MAPK remains phosphorylated (activation) till metaphase II and this activation is essential for the maintenance of metaphase II arrest (Fan and Sun, 2004). The MAPK cascade or MOS/MEK1/MAPK/p90rsk signaling pathway regulates the oocyte maturation through a cascade of protein kinase phosphorylation (Sobajima et al., 1993; Harrouk and Clarke, 1995). MAPK cascade interacts extensively with other proteins involved in oocyte meiotic cell cycle regulation like MPF, protein kinase A, protein kinase C, and calmodulin-dependent protein kinase II, as well as with protein phosphatases (Fan and Sun, 2004). Activation of MPF is usually induced by specific dephoshphorylation of p34cdc2 at threonine-14 and tyrosine-15 by cdc25 phosphatase. Translational arrest of cdc25, a protein phosphatase upstream to MPF, inhibits plasma membrane disassembly and spindle formation during meiotic progression (Dai et al., 2000). *In vitro* maturation studies in porcine and bovine oocytes showed that MPF activation begins at the time of germinal vesicle break down (GVBD), sharply rises at MI stage, declines at anaphase-telophase transition and reaches a high level again at the MII stage (Mattioli et al., 1991; Wu et al., 1997; Sun and Nagai, 2003). High level of MPF and MAP kinase activity in MII stage oocytes works as a cytostactic factor (CSF) and cause MII arrest in mammalian oocytes. In addition to MAPK, a drop in intracellular cAMP levels followed by the decrease of the cAMP-dependent PKA activity is also associated with resumption of meiosis via dephosphorylation of p34cdc2 by CDC25 phosphatase (Mehlmann, 2005). The matured oocyte not only contributes half of

the embryonic genome, but it supports fertilization and early embryonic development until the MET occurs (Tomek et al., 2002). In particular the oocyte role is critical during the interval between fertilization and the so-called MET when the transcriptional activity of the embryonic genome becomes fully functional. Recently, two maternal effect genes that encode the oocyte-specific proteins Zar-1 (zygote arrest 1) and Mater (maternal antigen that embryos require) have been identified in mouse, human and bovine models during oocyte to embryo transition (Tong et al., 2000; Tong et al., 2002; Wu et al., 2003; Brevini et al., 2004; Pennetier et al., 2006). Zar-1 has been shown to be required for progression through the one-cell stage, whereas Mater is required for development beyond the two-cell stage (Tong et al., 2000; Wu et al., 2003).

Coenen *et al* (Coenen et al., 2004) used radio labeling 2-D PAGE and demonstrated three major patterns of translational activity during bovine oocyte maturation *in-vitro* (one at the initiation of maturation, 0–4 h; one in the middle during the transition to GVBD and MI, 4–16 h; and one after completion of MI, 16–28 h) suggesting a developmentally regulated series stage-specific protein synthesis. Patterns of protein synthesis at the beginning of maturation (0–4 hr) and after maturation (24–28 hr) are very different. Only half of proteins synthesized after maturation was already synthesized at the beginning of the process indicating that GV stage oocytes produced a set of proteins whose synthesis either stopped, or was substantially reduced, after resumption of meiosis. Furthermore, more new proteins are observed at 4–8 hr and 16–20 hr of maturation which correlated with the breakdown of GV and appearance of MII respectively, but some could be synthesized also in preparation for the fertilization. Furthermore, 15% of proteins were synthesized throughout maturation (0-28 hrs). These

authors proposed that these proteins could represent "house-keeping" proteins required for cell structure maintenance and function of bovine oocytes. In addition, proteins that are present in immature oocytes and expressed throughout oocyte maturation and embryonic development until MET may be required for successful activation of the embryonic genome and be considered "maternal house-keeping proteins" (MHKP). Massicotte *et al* (Massicotte et al., 2006) used radio labeling 2-D PAGE coupled to MALDI-TOF mass spectrometry to identify the protein synthesis patterns during bovine oocyte maturation (GV, MI, and MII) and early embryo development (2-cell, 4-cell, and 8-cell). They identified 10 proteins as MHKPs that are expressed throughout oocyte maturation and early embryo development up to the MET. These proteins are four chaperones (HSC71, HSP70-2, CCTε and CypA), two proteins related to degradation pathway (UCH-L1 and E2D3), the glycolytic pathway protein (2, 3-BPGM), cytoskeleton proteins (β-actin and/or γ–actin), anti oxidant protein (GSTM5) and one protein (E-FABP) involved in lipid transport. Compared to reductionist approaches, proteomics based approach has the unique advantage of identifying changes in protein patterns between different stages of oocyte maturation and early embryo development. Consequently, the resulting differentially expressed proteins between various stages of oocyte and early embryo development have diagnostic value as biomarkers for oocyte maturation and quality and indicate directions for more specific investigations.

Oocytes do not develop in isolation; they are intimately involved with cumulus cells (CC). Bidirectional communication between oocytes and surrounding CCs is essential for both cell types survival (Fukui and Sakuma, 1980; Buccione et al., 1990; Eppig, 1991; Hussein et al., 2005). The oocyte is dependent on CCs for nutrients and

regulatory signals which are necessary to promote nuclear and cytoplasmic maturation of oocyte and hence the acquisition of developmental competence (Chian and Sirard, 1995; Ka et al., 1997; Hussein et al., 2005). CCs bind to the oocyte zona pellucida (ZP) and connect to the oocyte cytoplasmic membrane to form a cumulus-oocyte complex (COC) through transzonal cytoplasmic process (Eppig, 1991). Gilcrist *et al* and other groups revealed 'gap junctions' in the regions of association between oocytes and CCs that allow nutrient and paracrine factor transport between oocytes and CCs (Grazul-Bilska et al., 1997; Tanghe et al., 2002; Gilchrist et al., 2004). Molecules that pass through gap junctions include ions, metabolites, and amino acids necessary for oocyte growth, as well as small regulatory molecules such as cAMP that control oocyte maturation. Gap junctional signaling is a key means of disseminating local and endocrine signals to the oocyte via CCs. CC removal before maturation, or the obstruction gap junctions, suppresses oocyte maturation (Mori et al., 2000; Vozzi et al., 2001; Fatehi et al., 2002; Wongsrikeao et al., 2005; Assidi et al., 2008). Furthermore, CCs protect oocytes by preventing oxidative stress-induced apoptosis and DNA damage by enhancing the glutathione content in oocytes of swine (Tatemoto et al., 2000) and cow (de Matos et al., 1997) thus functionally influence its competence. Conversely, oocytes and their secreted growth factors play active roles in granulosa cell differentiation into the two phenotypically and functionally distinct sub-types of CCs and mural granulosa cells (MGCs) (Eppig et al., 1997; Li et al., 2000). Oocyte-secreted factors (OSFs) regulate key CC functions, including cell growth promotion and preventing death as well as luteinization by regulating steriodogenesis, inhibin synthesis and suppressing luteinizing hormone receptor expression. Conversely, mural granulosa cells in antral follicles, which

have no direct physical contact with the oocyte and, presumably, experience a more diffuse concentration of OSFs, proceed to a different phenotype. Removing oocyte paracrine signaling causes CCs to lose their distinctive phenotype and display characteristics more typical of MGCs (Eppig et al., 1997; Li et al., 2000; Gilchrist et al., 2006; Dragovic et al., 2007). In the ovulating follicle, OSF's are also vital for facilitating ovulation by enabling cumulus cell expansion and regulating extracellular matrix stability (Gilchrist et al., 2004). Although this bidirectional communication and paracrine signaling between CCs and oocytes is critical for oocyte growth and regulation of meiotic maturation of the oocyte (Eppig et al., 1993; De La Fuente and Eppig, 2001; Gilchrist et al., 2003; Sugiura and Eppig, 2005), the molecular details underlying these mechanisms are poorly defined.

Reductionist approaches have again been used for identifying proteins important in oocyte – CC communication. Gap junctions, composed of connexin proteins, allow transfer of small molecules between the two cell types (Albertini et al., 2001). Simon *et al* noted the failure of oocyte maturation in connexin-37 knock-out mice, which emphasizes the importance of this interconnection for proper oocyte maturation (Simon et al., 1997). Sasseville *et al* showed that *PDE8A* and *PDE8B* are the predominant cAMP-phosphodiesterases expressed in the bovine cumulus cells and play an important role in cAMP modulation during oocyte maturation. Inhibition of PDE8 using dipyridamole caused the dose dependent increase in cAMP levels in the COC and delayed oocyte nuclear maturation (Sasseville et al., 2009). Conversely, oocytes through the secretion of OSFs, regulate a multitude of key CC functions, which may in turn produce positive regulatory factors that pass back to the oocyte, improving subsequent

development. Growth differentiation factor 9 (GDF9) and bone morphogenetic protein 15 (BMP15) are two crucial OSFs essentially unique to gametes (McNatty et al., 2004). Diffusible paracrine factors produced by oocytes promote the expression of an array of CC genes and regulate a broad range of CC functions. For example, GDF9 up regulates CC gene expression of hyaluronic acid synthase 2 (HAS2), cyclooxygenase 2 (COX2; PTGS2) and gremlin (GREM1), that have been associated with oocyte maturation and subsequent embryo developmental potential (van Montfoort et al., 2008). Recently it has een demonstrated that OSFs also regulate CC amino acid and energy substrate uptake and transport of these substances to the oocyte. Thus, CCs nurture the oocyte towards developmental competence by supplying various small regulatory molecules such as cAMP and energy substrates like pyruvate (Albertini et al., 2001). Exposure of COCs during oocyte maturation to OSFs, whether in their native form as an uncharacterized mix of growth factors secreted by the oocyte or as exogenous recombinant BMP15 or GDF9 leads to an ~50% increase in oocyte developmental competence suggesting that OSFs improve developmental potential of oocytes (Gilchrist et al., 2008).

Proteomics and male fertility

Spermatogenesis is the creation of spermatozoa (sperm cells) from spermatogonia (germ cells) over a period of time within the seminiferous tubules of the testis. During the process of differentiation into spermatozoa, type A and type B spermatogonia undergo repeated mitotic divisions to yield primary spermatocytes. Primary spermatocytes, through two successive meiotic divisions, form haploid spermatids followed by terminal differentiation of spermatids into the spermatozoa. Spermatozoa, which are non-motile at this stage, continue to mature during their transit through epididymis and acquire

motility. Protein addition, loss, and modifications (as well as lipid membrane changes) occur as the spermatozoa pass through the caput, corpus, and cauda epididymis and during ejaculation (Nikolopoulou et al., 1985; Topfer-Petersen et al., 2005; Vadnais et al., 2007). Transcriptionally inactive spermatozoa rely completely on post-translational modifications in order to acquire functional competence (Blaquier et al., 1988; Blaquier et al., 1988). Again we will summarize reductionist work done to describe the molecular details underlying sperm development and maturation and this will be followed by a discussion of proteomic approaches.

Following ejaculation, molecules present in the seminal plasma form a coating on the entire surface of the sperm cell, known as decapacitation factors (DF), keeping the cell in a non capacitated state. DFs may be of seminal vesicle, prostate, or epididymal origin. Several molecules present in the seminal plasma from different mammalian species have been determined as possible DFs, including a high-molecular-weight protein named acrosome stabilizing factor(Eng and Oliphant, 1978), spermine (Rubinstein and Breitbart, 1991), free cholesterol (Cross, 1996), and spermadhesins (Dostalova et al., 1994; Strzezek et al., 2005). In porcine, two epidydimal proteins, antiagglutinin and CRISP1, associate with the spermatozoon membrane during transit through the epididymis (Harayama et al., 1999; Vadnais and Roberts, 2007). The proposed roles for these epididymal proteins include the prevention of premature sperm capacitation and acrosomal exocytosis and thus modulate the spermatozoon fertilizing ability (Vadnais et al., 2007). An additional decapacitation factor identified from seminal plasma of domestic animals is the ubiquitous enzyme, platelet-activating factor-acetyl hydrolase (PAF-AH) (Hough and Parks, 1994). PAF-AH is responsible for the inactivation of PAF

which is thought to be involved in autocrine activation of capacitation. Spermatozoa are unable to fertilize an egg immediately after ejaculation. A spermatozoon must reach the site of fertilization and undergo a process of maturation known as capacitation for successful fertilization to occur (Braundmeier and Miller, 2001). Spermatozoa undergo a series of biochemical and biophysical modifications in the female genital tract, including destabilization of the plasma membrane, cholesterol efflux and increase in membrane fluidity, alterations in intracellular ions concentrations, protein phosphorylation, changes in swimming patterns, and chemotactic motility (Breitbart and Spungin, 1997; Baldi et al., 2000). During capacitation, DF's are progressively released from sperm surface giving spermatozoa the maximal fertilizing ability at the site of fertilization (Yanagimachi, 1994). Albumin, high density lipoproteins (HDL), heparin, glycosaminoglycans (GAG) and β-cyclodextrins have been shown to promote sperm capacitation by acting as cholesterol acceptors that mediate cholesterol efflux from the plasma membrane (Vadnais et al., 2007). In a bovine model, bovine seminal plasma proteins (BSP-A1, BSP-A2, BSP-A3 and BSP-30-kDa) form a coating on the sperm surface by binding to the phosphocholine lipids on the plasma membrane at the time of ejaculation. These proteins have been shown to interact with heparin like GAG and HDL in the female reproductive tract, stimulating cholesterol and phospholipid efflux from the sperm plasma membrane and resulting in capacitation (Manjunath and Therien, 2002).

Previous research identified that sperm capacitation is associated with a phosphorylation cascade which involves a PKA-regulated increase in protein tyrosine phosphorylation (Salicioni et al., 2007) and appears to be a necessary prerequisite for a spermatozoon to fertilize an egg (Visconti et al., 1995; Urner et al., 2001; Urner and

Sakkas, 2003). cAMP produced during capacitation and subsequent activation of protein kinase A (PKA) appears to play an essential role in the increase of tyrosine phosphorylation during capacitation (Visconti et al., 1999). As PKA is able to phosphorylate various protein substances, its sequestration in specific cellular compartments is necessary to spatially restrict its action. A-kinase anchor proteins (AKAPs) bind to the regulatory subunit of PKA, thereby recruiting the kinase to specific sperm components (Rubin, 1994; Dell'Acqua and Scott, 1997). AKAPs have been recently characterized in spermatozoa of different species (Carrera et al., 1996; Moss et al., 1999; Vijayaraghavan et al., 1999; Moss and Gerton, 2001). In addition to participating in the regulation of protein phosphorylation in sperm, AKAPs are themselves phosphorylated on tyrosine residues (AKAP82, FSP95; (Carrera et al., 1996; Johnson et al., 1997; Mandal et al., 1999)) during capacitation, indicating that their functions are modified by their phosphorylation status. Galantino-Homer *et al* identified the major fibrous sheath AKAP, two flagellar outer dense fiber proteins, tubulins, and several mitochondrial proteins as substrates of tyrosine phosphorylation during bovine spermatozoa capacitation (Galantino-Homer and Kopf, 2000). The endpoint of *in vitro*/*in vivo* capacitation is the ability of carbohydrate binding proteins on sperm plasma membrane (receptors) to bind to their complementary glycan molecules (ligands) on zona pellucida (ZP) of an egg and initiate acrosome reaction (AR) (Tulsiani et al., 2007). The AR is characterized by fusion of a spermatozoon outer acrosomal membrane with overlying plasma membrane at multiple sites followed by acrosomal exocytosis. The exocytosis of hydrolytic enzymes (proteinases, glycohydrolases) released at the surface of ZP enables the acrosome-reacted (and hyperactive) spermatozoon to penetrate the ZP

and fertilize the egg. There are three candidate enzymes involved in sperm penetration through the ZP: acrosomal serine protease acrosin (Baba et al., 1989), testicular serine protease 5 (TESP5) (Honda et al., 2002), and multi subunit proteolytic holoenzyme proteasome in the acrosome (Sawada et al., 2002). The physiological role of sperm entry into and penetration through the ZP is most likely to remove the acrosome from the fertilizing sperm. The molecular details and signal transduction pathways mediating the sperm capacitation and acrosome reaction have been extensively reviewed in the recently published articles (Baldi et al., 2000; Salicioni et al., 2007; Tulsiani et al., 2007; Vadnais et al., 2007; Abou-Haila and Tulsiani, 2009) and will not be repeated here.

Compared to reductionist approaches, and in contrast to the oocyte, at the time of writing there were thirty-nine proteomic studies (Olson and Hinton, 1985; Ross et al., 1990; Druart et al., 1994; Syntin et al., 1996; Fouchecourt et al., 2000; Chaurand et al., 2003; Lefievre et al., 2003; Tan et al., 2004; Baker et al., 2005; Gatti et al., 2005; Strzezek et al., 2005; Cao et al., 2006; Frenette et al., 2006; Lalancette et al., 2006; Martinez-Heredia et al., 2006; Moura et al., 2006; Moura et al., 2006; Pilch and Mann, 2006; Stein et al., 2006; Bassols et al., 2007; de Mateo et al., 2007; Ding et al., 2007; Domagala et al., 2007; Hassan et al., 2007; Moura et al., 2007; Oliva et al., 2007; Aitken and Baker, 2008; Baker et al., 2008; Choi et al., 2008; Luo et al., 2008; Martinez-Heredia et al., 2008; Oliva et al., 2008; Puri et al., 2008; Thimon et al., 2008; Zou et al., 2008; Khan et al., 2009; Oliva et al., 2009; Secciani et al., 2009) of spermatozoa and their surrounding milieu in total and thirteen studies (Druart et al., 1994; Syntin et al., 1996; Fouchecourt et al., 2000; Gatti et al., 2005; Strzezek et al., 2005; Frenette et al., 2006; Lalancette et al., 2006; Moura et al., 2006; Moura et al., 2006; Bassols et al., 2007;

Moura et al., 2007; Choi et al., 2008; Puri et al., 2008) in animals other than mice and humans. Syntin *et al*(Syntin et al., 1996) analyzed the secretory activity of epithelial cells throughout the boar epididymis using one and 2-D gel electrophoresis. A total of 146 proteins were found to be secreted by the epididymis and most are secreted from distal caput when compared to proximal caput and cauda of epididymis. Qualitative and quantitative differences in secretory activityof the epididymal epithelium have been shown to be present throughout the epididymis in the human (Ross et al., 1990), rat (Brooks and Tiver, 1983), mouse (Bendahmane and Abou-Haila, 1994) and in the stallion (Fouchecourt et al., 2000) and ram (Druart et al., 1994). Clusterin, glutathione peroxidase, retinol-binding protein, lactoferrin, EP4, p-N-acetyl-hexosaminidase, alpha-mannosidase, and procathepsin L were shown to be highly secreted (Syntin et al., 1996). Fouchécourt *et al*(Fouchecourt et al., 2000) carried out the proteomic analysis of stallion epididymal fluid using 2-D electrophoresis and identified 117 proteins secreted by the epididymal epithelium in various parts of the organ. They also found that lactoferrin, clusterin, procathepsin D, cholesteroltransfer protein (CTP/HE1), glutathione peroxidase (GPX),hexosaminidase, and prostaglandin D2 synthase (PGDS) were among the most abundant proteins representingaround 80% of the total epididymal secretory activity in the stallion. Although it is accepted that sperm acquires the capacity to fertilize oocytes *in vivo* while in the epididymis (Horan and Bedford, 1972; Amann and Griel, 1974), there is also evidence that secretions of the accessory sex glands (AG) influence sperm physiology and fertilization (Henault et al., 1995; Ying et al., 1998). For example, Moura *et al* (Moura et al., 2006) carried out the proteomic analysis of accessory sex gland fluid of Holstein bulls by 2-D gel electrophoresis and capillary liquid chromatography–nano

electrospray ionization mass spectrometry (CapLC–MS/MS). They identified the presence of proteins potentially involved in capacitation (bovine seminal plasma protein—BSP-A1/A2 and A3, BSP 30 kDa, albumin); acrosome reaction and sperm-oocyte interaction (phospholipase A2 (PLA2), osteopontin); sperm membrane protection, prevention of oxidative stress, complement-mediated sperm destruction and anti-microbial activity (albumin, clusterin, acidic seminal fluid protein (aSFP), 5_-nucleotidase (5_-NT), PLA2); interaction with the extracellular matrix (tissue inhibitor of metalloproteinase 2, clusterin) and sperm motility (aSFP, spermadhesin Z13, 5_-NT) (Moura et al., 2006). So far, many proteomic studies have been focused on the surrounding mileau of sperm to determine their influence on sperm physiology and fertilization. Lalancette *et al*(Lalancette et al., 2006) carried out proteomic analysis of bull sperm cytosolic fraction enriched in tyrosine kinases using affinity chromatography and mass spectrometry. They identified three families of tyrosine kinases: Src (lyn), csk and Tek (Bmx, Btk) in bull sperm cytosolic fraction which are essential for tyrosine phosphorylation of specific sperm proteins during capacitation. In addition to tyrosine kinases, they have also identified proteins implicated in different cellular events (actin polymerization, vesicle transport/membrane fusion, metabolism, sperm–egg interaction) associated with sperm capacitation and AR. Until recently, sperm proteomic analyses were largely conducted using 2-D PAGE (Pixton et al., 2004; Tan et al., 2004; Martinez-Heredia et al., 2006).

Sperm-oocyte plasma membrane binding and fusion

Successful sperm-oocyte interactions are necessary for fertilization *in vivo* and the propagation of sexually-reproducing organisms (Stein et al., 2006). The sperm interacts

with the oocyte on three separate levels: the cumulus cell layer surrounding the oocyte, the egg extracellular matrix (the zona pellucid (ZP)), and the oocyte plasma membrane. Sperm penetration through the cumulus matrix is facilitated by a membrane-bound hyaluronidase (Talbot et al., 2003). Binding of spermatozoa with the egg's ZP initiates the AR and allows the sperm to penetrate the zona. AR induced changes in the sperm are important for fertilization process, as only acrosome reacted sperm can bind and fuse with the oolemma (Yanagimachi, 1994). Recent research in conjunction with genomic and proteomic techniques has uncovered a number of cell surface molecules with putative roles in the sperm-oolemmal interaction, but the picture is still far from complete (Nixon et al., 2007).

Using cell biology approaches and gene deletion studies in mice, two plasma membrane proteins IZUMO and CD9 on the male and female gametes respectively, were found to be essential for sperm-oolemma fusion (Sutovsky, 2009). IZUMO, a novel member of the immunoglobulin superfamily, plays key role in sperm-oocyte fusion**.** It islocated on the inner acrosomal membrane and equatorial segment of the sperm.*In vivo* knock out studies in mouse have shown that sperm from IZUMO -/- males were capable of binding to and penetrating the ZP and subsequently binding to but not fusing with the oolemma (Inoue et al., 2005). CD9, a memberof the tetraspanin family, spans the plasma membrane four times, having two extracellular loops (one small, one large) and short cytoplasmic NH_2-terminal and COOH-terminal tails (Hemler, 2001). Female mice carrying a gene knockout for CD9 exhibited severely reduced fertility attributable to a lack of fusibility of their CD9-deficient oocytes, whereas, no significant differences were observed for sperm-binding ability between control and CD9-deficient eggs (Le Naour et

al., 2000; Miyado et al., 2000). Glycophosphatidylinositol (GPI)-anchored proteins oolemma-associated tetraspanin CD81 and egg surface proteins, also appear to play an essential role in fusion (Coonrod et al., 1999; Hemler, 2001). Additional proteins that may have nonessential yet still facilitating roles in sperm-oolemma adhesion and fusion include sperm disintegrins (ADAMs) and epididymal-derived cysteine-rich secretory (CRISP) proteins such as CRISP1 and CRISP2. To date, there are no proteomic studies on sperm regions that mediate sperm–egg interactions in animals.

Biomarkers in reproduction

Biomarkers are biological molecules that are indicators of normal biological process, pathogenic processes, or pharmacologic responses to a therapeutic invention (http://www.biomarkersconsortium.org). Biomarkers play an important role in disease diagnosis and treatment, especially for the early detection of the disease, and to enable screening of asymptomatic populations (Srinivas et al., 2002). As proteins define a cells phenotype, measurement of differential expression of proteins could indicate disease specific changes in cells/tissues that correlates with the risk or progression of a disease or with the susceptibility of disease to a given treatment (Venugopal et al., 2009). Biomarkers are also seen as the key to personalized medicine, where they are used to index individual differences in disease predisposition, progression, response to therapeutics, and then to design treatments individually to specific patients for highly efficient intervention in disease processes. Two approaches are used to discover biomarkers. Again, the first is "reductionist" based on more traditional candidate-gene assay for a particular protein of interest, selected based on either a biological rationale or from analysis of candidates derived from some other sources. The second is a "de novo"

discovery approach, using different proteomic techniques (Matt et al., 2008). Both approaches are complementary to each other and have their own advantages and disadvantages, and may be performed in parallel (Guo et al., 2007). In reproductive medicine, there are numerous biomarkers currently being used to detect estrus (estrogen), pregnancy diagnosis (progesterone), ovarian tumor diagnosis (Inhibins, leptin, prolactin, osteopontin and insulin like growth factor II (Robertson and McNeilage, 2004; Mor et al., 2005)), to predict the freezability of spermatozoa (apoptoptic markers such as active capsases (Pena et al., 2007; Ortega-Ferrusola et al., 2008)), and to detect fetal anomalies (alpha – feto protein (Janzen et al., 1982; De Mees et al., 2006)). At present, there are very few biomarkers to distinguish poor fertility samples from those of high fertility. Here, we briefly discuss the identification of biomarkers associated with fertility using reductionist approaches followed by proteomic approaches.

Reductionist approaches have shown to be successful in identifying biomarkers associated with fertility. For example, using the bovine model, Jeon *et al*(Jeon et al., 2008) examined the relationship between survivin protein expression and the quality of cumulus-oocyte complexes (COCS) or the quality of pre-implantation embryo using immunoflourscence. Survivin levels were higher in good quality COCS with significantly better developmental competence than in poor quality COCS. In male fertility, Erikson *et al* (Erikson et al., 2007) identified 55 KDa osteopontin (OPN) as a high fertility biomarker in Holstein bull seminal plasma. Another protein, clusterin, has been used as a biomarker in male fertility because semen samples with many clusterin- positive spermatozoa have lower fertility, as determined by nonreturn-to-estrus rates and had many morphological abnormalities (Ibrahim et al., 2000).

Proteomic approaches, especially those involving mass spectrometry, provide data on protein expression as well as post-translational modifications in different conditions, which could lead to the discovery of biomarkers. There have been very few proteomic studies in animals describing the biomarkers of male and female fertility**.** Bhojwani *et al.,* 2006conducted similar studies in bovine oocyte and identified that the expression of four proteins was altered during IVM, namely tubulin- β-chain (TBB), cyclin E2 (CCNE2), protein disulfide isomerase (PDIA3) and one of two different forms of peroxiredoxin 2 (PDX2). Expression of three proteins (CCNE2, TBB, and PDX2) has been significantly decreased and one protein (PDIA3) has been significantly increased inMI and MII stages compared with GV oocyte. They suggested that the proteins that are differentially regulated during bovine oocyte IVM can serve as potential molecular markers for oocyte maturation and quality (Bhojwani et al., 2006). Bidirectional communication between oocytes and surrounding cumulus cells (CC) is essential for maturation of a competent oocyte. The surrounding cumulus cell population of granulosa cells represents an attractive potential target for identification of objective molecular markers predictive of oocyte competence. Several studies have been focused on gene expression in cumulus cells in order to find specific molecular markers associated with successful oocyte maturation (Assidi et al., 2008; Bettegowda et al., 2008; Tesfaye et al., 2009). So far no proteomic studies have focused on cumulus cells in animals to identify molecular markers predictive of oocyte competence.

Few studies have been conducted on the identification of biomarkers associated with male fertility. Killian *et al*(Killian et al., 1993) carried out proteomic analysis of bull seminal plasma of varying fertility rates using 2-D gel electrophoresis and showed that an

abundance of specific proteins in semen was related to fertility: two seminal plasma proteins (26 kDa, pi 6.2; 55 kDa, pI 4.5) were more abundant in semen of higher fertility bulls and two other proteins (16 kDa pI 6.7 and 16 kDa pI 4.1) were more abundant in lower fertility bulls. Protein characterization of two proteins (26 and 55 kDa) abundant in high fertility bulls revealed those proteins as prostaglandin D synthase and osteopontin respectively (Cancel et al., 1997; Gerena et al., 1998). Studies on bovine seminal plasma revealed the presence of two heparin binding proteins namely fertilization associated antigen (FAA) and tissue inhibitor of metalloproteinases-2 (TIMP-2) associated with increased pregnancy rates in cows and heifers (McCauley et al., 1999; McCauley et al., 2001). Comparative proteomic analysis of cauda epididymal fluid (CEF) of bulls of varying fertility rates revealed the differential expression of three proteins namely α-L-fucosidase, cathepsin D, and prostaglandin D synthase between high and low fertility bull CEF which may serve as biomarkers associated with male fertility (Moura et al., 2006). So far proteomic studies have focused on the surrounding mileau of spermatozoa (epididymis, accessory sex gland fluids, and seminal vesicles) for identification of biomarkers associated with male fertility.

Proteomics identifies panels of biomarkers at one time, which would enhance the positive predictive value of a test and minimize false positives or false negatives. The inherent advantage afforded to proteomics is that it enables the identification of the protein changes responsible for the development and/or pathological outcome of disease processes in a relatively accurate manner. Moreover, mass spectrometry based quantitative proteomics offers a promising platform to understand molecular pathogenesis of infertility and for discovery of potential biomarkers of oocyte and

spermatozoon quality. Proteomics should not only be able to inform our attempts to identify markers for male or female fertility regulation but also help us to resolve the molecular basis of defective oocyte or spermatozoon function with implication of our ability to both diagnose and treat this condition.

Systems biology modeling

Biological systems are extremely complex and have an emergent behavior that is not explicitly described by studying the behavior of the individual components of the system. Therefore it is difficult to predict the functional phenotype (Van Regenmortel, 2004). The components of a complex system interact in many ways, including negative feed-back and feed-forward control, which lead to dynamic features that cannot be predicted by simple linear models that disregard cooperative and non-additive effects (Van Regenmortel, 2004). Systems biology is a scientific approach which focuses on system level understanding of biological systems. It characterizes the behavior of a system more fully and it uses all information in the data while estimating the parameters of interest. Modeling in terms of interaction networks can identify key elements in genetic networks (regulatory hubs) that, when altered, change the fundamental properties of the networks (Nanduri et al., 2008). Recent developments in high throughput 'omic' studies and bioinformatics are providing data that molecular biologists need to simulate behavior of complex biological networks and systems (Kitano, 2002; Van Regenmortel, 2004). Systems biology modeling of entire proteomes reveals how proteins interact with each other, identify their functional role, what kinds of metabolic and signaling networks they form, and compare variations in their expression levels under different conditions (Cannataro et al., 2005). Gene Ontologies that provide a standardized vocabulary for data

analysis (Blake, 2004), and pathway analysis software which creates visual representation of the biological systems (Aderem and Smith, 2004), make it possible to analyze large amounts of data and build realistic models of complex systems. The promise of systems biology is to acquire a deeper understanding of biological systems which could then be utilized in practical real – world applications such as predictive and preventive medicine. In the context of this thesis, application of systems biology modeling can result in a better understanding of infertility which arises from perturbations of multileveled dynamic processes. It also can be used for identification of novel therapeutic targets for preventing or treating infertility in humans as well farm animals.

Systems modeling using pathway analysis tools

Today, advances in high-throughput proteomic data analysis have resulted in the generation of large volumes of protein function data published in journal articles. Public databases like NCBI PubMed archive these articles and allow users to view either the full-text version or the abstract. With the accumulation of biological information in literature, there is a need for an accurate and efficient way of extracting information from text. Using various text mining tools and manual extraction, pathway analysis tools build their databases based on the curated data from literature. Many pathway analysis tools are available such as Ingenuity Pathway analysis (IPA) (Ingenuity Systems, 2006), Pathway Studio (PS) (Yuryev et al., 2006), integrative visual analysis tool for biological networks and pathways (VisANT) (Hu et al., 2008), Genome Expression Pathway Analysis Tool (GEPAT) (Weniger et al., 2007), Pathway Explorer (Mlecnik et al., 2005), String (von Mering et al., 2007) and GenMapp (Dahlquist et al., 2002). The goal of all these tools is

to make connections between biological entities and expand the network for a wider biological picture.

These modeling methods have the potential to be applied in reproductive proteomics. Using one of our bovine spermatozoa proteomic datasets, we recently compared commercially-available pathway analysis tools: Ingenuity Pathways Analysis (Ingenuity Systems Inc, Redwood City, CA) and Pathway Studio (Ariadne Genomics, Rockville, MD). Both tools map related genes and gene products upstream and downstream of the user-specified list of genes/proteins and then can be used to draw direct and indirect connections between interacting genes/proteins, as well as report a list of functions and diseases with which the genes/proteins are associated. During comparison, we focused on the product features and functionality, database management, user-friendliness, and customer support of both tools(Table 2.1). On one hand IPA's manually curated database is more accurate than text-mining done by a search engine. On the other hand, MedScan Reader's natural language processing (NLP) text-mining is able to cover a broader list of relationships and is more up-to-date with the exponential increase in published scientific literature due to automated text-mining. Results from our comparison showed that IPA is finding relatively more relationships than Pathway Studio, yet both tools are finding comparable amounts of supporting references for all relationships in common. In addition IPA has other advantages like simplifying the pathway display for better visualization (by showing multiple relations between two edges as one edge) and creating networks and pathways of manageable size (allows users to select the number of entities to be mapped in a pathway).

The main disadvantage of IPA is it doesn't have interaction data directly from the bovine studies. It converts the bovine proteins into human, mouse and rat orthologues and transfers the interaction data from human, mouse and rat to bovine proteins. Another disadvantage is only 70-80% of the proteins from our datasets mapped to corresponding gene objects in the Ingenuity Pathways Knowledgebase (IPKB) and are known to interact with other genes/proteins based on published, peer reviewed content in the IPKB. Approximately 20-30% of our identified spermatozoa proteins were not included in the analysis. And also there is no specific function terms associated with sperm, oocyte, and early embryonic development. Apart from these disadvantages, this tool saved us time as opposed to reading every single article manually in order to find interesting relations connecting our proteins of interest.

Systems modeling using Gene Ontology (GO)

To capture species specific information, I did GO to my proteomic datasets. The Gene Ontology (GO) is the *de facto* standard for functional annotation of gene products (Lewis, 2005). GO describes gene products in terms of cellular component, biological process and molecular function. Previous studies associated with animal reproduction have been successful in using GO to analyze their functional genomic datasets in order to find important biological processes associated with their datasets(Fair et al., 2007; Vallee et al., 2008; Katz-Jaffe et al., 2009; Smith et al., 2009). GO is also useful in generating hypotheses to the functional genomic datasets either by GO enrichment analysis (GO terms that are statistically over- or under-represented in a set of genes) or annotation clustering (group all gene products under a similar GO category). Despite its critical importance in analyzing proteomic and other functional genomics data, relatively few

gene products associated with reproductive cells and tissues of farm animals have detailed GO annotations (more depth). This is due to the fact that most proteins identified in newly sequenced animal genomes are predicted with no functional literature associated with them. Although literature is available for some known proteins, there are very few proteins are annotated from literature in domestic animals and sometimes there are no GO terms that are transferable from the literature. So, for providing finer and more precise functional annotations (i.e. improved GO *depth*), we need to have more localization and functional studies associated with reproductive cells and tissues to confirm the predicted proteins. Additionally, for known proteins, we need to search and curate the available literature and annotate proteins with more specific terms and experimental evidence codes. AgBase (www.agbase.msstate.edu)(McCarthy et al., 2006), a curated, open web-accessible resource is working on providing more detailed and precise GO annotations (manual GO curation from the literature) to the agriculturally important animal gene products. AgBase also provides annotations to the predicted proteins with no functional literature by sequence similarity and orthologue search. AgBase integrates structural and functional annotations and provides tools in an easy-to-use pipeline, allowing agricultural and biomedical researchers to rapidly and effectively model and derive biological significance from proteomic and other functional genomics datasets.

Research goals

The overall objective of my doctoral research dissertation is to understand the precise protein events and mechanisms that determine the oocyte and cumulus

communication and male fertility using a bovine model. My *central hypothesis* is that proteomics-based "computational systems biology" modeling of bovine oocyte and its associated cumulus cells, as well as spermatozoa will provide a comprehensive model of pathways and networks related to the fertility. The rationale for undertaking this project, and using the approaches that I have, is that understanding the molecular mechanisms underlying the oocyte and cumulus cell communication, and spermatozoon development will aid in development of rational methods optimized to produce more developmentally-competent oocytes and spermatozoa. This research on the bovine oocyte and the spermatozoon may provide the fundamental research platform for developing reproductive techniques and for studying reproductive diseases in human and other farm animals. The ultimate positive impact of this research is to facilitate the field of biomedical research with useful information for comparative biology, better understanding of bovine oocyte and spermatozoon development, infertility, biomarker discovery, and eventually development of therapies to treat infertility in bovine as well as humans. My dissertation has three aims:

Aim 1.Proteomic analysis of bovine germinal vesicle oocyte and cumulus cells. Germinal vesicle (GV) breakdown is fundamental for maturation of fully grown, developmentally competent, mammalian oocytes. Bidirectional communication between oocytes and surrounding cumulus cells (CCs) at GV stage is essential for maturation of a competent oocyte. Underlying mechanism(s) of these events are not known, however, a number of different gene products are expected to play an important role. My *hypothesis* for this aim is proteomic analysis of GV oocyte and its surrounding CC will aid in further understanding of molecular networks and pathways that define oocyte maturation,

fertilization, and embryonic development. The rationale for undertaking this aim is that this study will not only provide a foundation for signaling and cell physiology at the GV stage of oocyte development, but are also valuable for comparative studies of other stages of oocyte development at the molecular level. Furthermore, some of these proteins may represent molecular biomarkers for developmental potential of oocytes.

Aim 2.Proteomics-based systems biology modeling of bovine germinal vesicle stage oocyte and cumulus cell interaction. Oocytes are the female gametes which establish the program of life after fertilization. Interactions between oocyte and the surrounding cumulus cells at germinal vesicle (GV) stage are considered essential for proper maturation or 'programming' of oocytes, which is crucial for normal fertilization and embryonic development. However, neither the factors involved in this communication nor the mechanisms of their actions are well defined. My *hypothesis* for this aim is proteins and signaling pathways involved in the bidirectional communication between oocyte and cumulus cells will have implications in oocyte growth and survival, oocyte development and maturation, and fertilization. The rationale for undertaking this aim is that this study will provide the basis for further hypothesis-driven research on oocyte and cumulus cell communication and related reproductive abnormalities.

Aim 3.Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. Currently available fertility assays assess the defects that affect functional competence of spermatozoa (i.e.capacitation, acrosome reaction, sperm-oocyte interaction); however these cannot definitively predict fertility. At present, the molecular nature of sperm defects or biomarkers for accurate fertility prediction is not known. My *hypothesis* for

this aim is proteins and pathways that were differentially expressed between high and low fertility bulls will have a role in male fertility prediction. The rationale for undertaking this aim is that this study will identify putative molecular markers and pathways that could be tested in the future for male fertility.

References

Abou-Haila A, Tulsiani DR (2009) Signal transduction pathways that regulate sperm capacitation and the acrosome reaction. Arch Biochem Biophys

Aderem A, Smith KD (2004) A systems approach to dissecting immunity and inflammation. Semin Immunol 16: 55-67

Aitken RJ, Baker MA (2008) The role of proteomics in understanding sperm cell biology. Int J Androl 31: 295-302

Albertini DF, Combelles CM, Benecchi E, Carabatsos MJ (2001) Cellular basis for paracrine regulation of ovarian follicle development. Reproduction 121: 647-653

Amann RP, Griel LC, Jr. (1974) Fertility of bovine spermatozoa from rete testis, cauda epididymidis, and ejaculated semen. J Dairy Sci 57: 212-219

Assidi M, Dufort I, Ali A, Hamel M, Algriany O, Dielemann S, Sirard MA (2008) Identification of Potential Markers of Oocyte Competence Expressed in Bovine Cumulus Cells Matured with Follicle-Stimulating Hormone and/or Phorbol Myristate Acetate In Vitro. Biol Reprod

Assidi M, Dufort I, Ali A, Hamel M, Algriany O, Dielemann S, Sirard MA (2008) Identification of potential markers of oocyte competence expressed in bovine cumulus cells matured with follicle-stimulating hormone and/or phorbol myristate acetate in vitro. Biol Reprod 79: 209-222

Baba T, Kashiwabara S, Watanabe K, Itoh H, Michikawa Y, Kimura K, Takada M, Fukamizu A, Arai Y (1989) Activation and maturation mechanisms of boar acrosin zymogen based on the deduced primary structure. J Biol Chem 264: 11920-11927

Baker MA, Hetherington L, Reeves GM, Aitken RJ (2008) The mouse sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics 8: 1720-1730

Baker MA, Witherdin R, Hetherington L, Cunningham-Smith K, Aitken RJ (2005) Identification of post-translational modifications that occur during sperm maturation using difference in two-dimensional gel electrophoresis. Proteomics 5: 1003-1012

Baldi E, Luconi M, Bonaccorsi L, Muratori M, Forti G (2000) Intracellular events and signaling pathways involved in sperm acquisition of fertilizing capacity and acrosome reaction. Front Biosci 5: E110-123

Bassols J, Bonet S, Belghazi M, Dacheux F, Dacheux JL (2007) Proteomic study of the establishment of boar epididymal cell cultures. Theriogenology 68: 76-86.

Bendahmane M, Abou-Haila A (1994) Synthesis, characterization and hormonal regulation of epididymal proteins during postnatal development of the mouse. Differentiation 55: 119-125

Bettegowda A, Patel OV, Lee KB, Park KE, Salem M, Yao J, Ireland JJ, Smith GW (2008) Identification of novel bovine cumulus cell molecular markers predictive of oocyte competence: functional and diagnostic implications. Biol Reprod 79: 301-309

Bhojwani M, Rudolph E, Kanitz W, Zuehlke H, Schneider F, Tomek W (2006) Molecular analysis of maturation processes by protein and phosphoprotein profiling during in vitro maturation of bovine oocytes: a proteomic approach. Cloning Stem Cells 8: 259-274

Blake J (2004) Bio-ontologies-fast and furious. Nat Biotechnol 22: 773-774

Blaquier JA, Cameo MS, Cuasnicu PS, Gonzalez Echeverria MF, Pineiro L, Tezon JG (1988) The role of epididymal factors in human sperm fertilizing ability. Ann N Y Acad Sci 541: 292-296

Blaquier JA, Cameo MS, Cuasnicu PS, Gonzalez Echeverria MF, Pineiro L, Tezon JG, Vazquez MH (1988) On the role of epididymal factors in sperm fertility. Reprod Nutr Dev 28: 1209-1216

Braundmeier AG, Miller DJ (2001) The search is on: finding accurate molecular markers of male fertility. J Dairy Sci 84: 1915-1925

Breitbart H, Spungin B (1997) The biochemistry of the acrosome reaction. Mol Hum Reprod 3: 195-202

Brevini TA, Cillo F, Colleoni S, Lazzari G, Galli C, Gandolfi F (2004) Expression pattern of the maternal factor zygote arrest 1 (Zar1) in bovine tissues, oocytes, and embryos. Mol Reprod Dev 69: 375-380

Brewis IA (1999) Proteomics in reproductive research: the potential importance of proteomics to research in reproduction. Hum Reprod 14: 2927-2929

Bridges SM, Magee GB, Wang N, Williams WP, Burgess SC, Nanduri B (2007) ProtQuant: a tool for the label-free quantification of MudPIT proteomics data. BMC Bioinformatics 8 Suppl 7: S24

Brooks DE, Tiver K (1983) Localization of epididymal secretory proteins on rat spermatozoa. J Reprod Fertil 69: 651-657

Buccione R, Schroeder AC, Eppig JJ (1990) Interactions between somatic cells and germ cells throughout mammalian oogenesis. Biol Reprod 43: 543-547.

Cancel AM, Chapman DA, Killian GJ (1997) Osteopontin is the 55-kilodalton fertility-associated protein in Holstein bull seminal plasma. Biol Reprod 57: 1293-1301

Cannataro M, Guzzi PH, Mazza T, Tradigo G, Veltri P (2005) Using ontologies in PROTEUS for modeling proteomics data mining applications. Stud Health Technol Inform 112: 17-26

Cao W, Gerton GL, Moss SB (2006) Proteomic profiling of accessory structures from the mouse sperm flagellum. Mol Cell Proteomics 5: 801-810

Carrera A, Moos J, Ning XP, Gerton GL, Tesarik J, Kopf GS, Moss SB (1996) Regulation of protein tyrosine phosphorylation in human sperm by a calcium/calmodulin-dependent mechanism: identification of A kinase anchor proteins as major substrates for tyrosine phosphorylation. Dev Biol 180: 284-296

Chaurand P, Fouchecourt S, DaGue BB, Xu BJ, Reyzer ML, Orgebin-Crist MC, Caprioli RM (2003) Profiling and imaging proteins in the mouse epididymis by imaging mass spectrometry. Proteomics 3: 2221-2239

Chian RC, Sirard MA (1995) Effects of cumulus cells and follicle-stimulating hormone during in vitro maturation on parthenogenetic activation of bovine oocytes. Mol Reprod Dev 42: 425-431

Choi YJ, Uhm SJ, Song SJ, Song H, Park JK, Kim T, Park C, Kim JH (2008) Cytochrome c upregulation during capacitation and spontaneous acrosome reaction determines the fate of pig sperm cells: linking proteome analysis. J Reprod Dev 54: 68-83

Cliff Lamb G, Dahlen C, Maddox M (2008) What is the Economic Impact of Infertility in Beef Cattle? : 1-3

Coenen K, Massicotte L, Sirard MA (2004) Study of newly synthesized proteins during bovine oocyte maturation in vitro using image analysis of two-dimensional gel electrophoresis. Mol Reprod Dev 67: 313-322

Coonrod SA, Naaby-Hansen S, Shetty J, Shibahara H, Chen M, White JM, Herr JC (1999) Treatment of mouse oocytes with PI-PLC releases 70-kDa (pI 5) and 35- to 45-kDa (pI 5.5) protein clusters from the egg surface and inhibits sperm-oolemma binding and fusion. Dev Biol 207: 334-349

Cross NL (1996) Human seminal plasma prevents sperm from becoming acrosomally responsive to the agonist, progesterone: cholesterol is the major inhibitor. Biol Reprod 54: 138-145

Dai Y, Lee C, Hutchings A, Sun Y, Moor R (2000) Selective requirement for Cdc25C protein synthesis during meiotic progression in porcine oocytes. Biol Reprod 62: 519-532

Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR(2002).GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. Nat Genet.31(1):19-20.

De La Fuente R, Eppig JJ (2001) Transcriptional activity of the mouse oocyte genome: companion granulosa cells modulate transcription and chromatin remodeling. Dev Biol 229: 224-236

De La Fuente R, Viveiros MM, Burns KH, Adashi EY, Matzuk MM, Eppig JJ (2004) Major chromatin remodeling in the germinal vesicle (GV) of mammalian oocytes is dispensable for global transcriptional silencing but required for centromeric heterochromatin function. Dev Biol 275: 447-458

De Mateo S, Martinez-Heredia J, Estanyol JM, Dominguez-Fandos D, Vidal-Taboada JM, Ballesca JL, Oliva R (2007) Marked correlations in protein expression identified by proteomic analysis of human spermatozoa. Proteomics 7: 4264-4277

De Matos DG, Furnus CC, Moses DF (1997) Glutathione synthesis during in vitro maturation of bovine oocytes: role of cumulus cells. Biol Reprod 57: 1420-1425

De Mees C, Laes JF, Bakker J, Smitz J, Hennuy B, Van Vooren P, Gabant P, Szpirer J, Szpirer C (2006) Alpha-fetoprotein controls female fertility and prenatal development of the gonadotropin-releasing hormone pathway through an antiestrogenic action. Mol Cell Biol 26: 2012-2018

Dell'Acqua ML, Scott JD (1997) Protein kinase A anchoring. J Biol Chem 272: 12881-12884

Ding Z, Qu F, Guo W, Ying X, Wu M, Zhang Y (2007) Identification of sperm forward motility-related proteins in human seminal plasma. Mol Reprod Dev 74: 1124-1131

Domagala A, Pulido S, Kurpisz M, Herr JC (2007) Application of proteomic methods for identification of sperm immunogenic antigens. Mol Hum Reprod 13: 437-444

Dostalova Z, Calvete JJ, Sanz L, Topfer-Petersen E (1994) Quantitation of boar spermadhesins in accessory sex gland fluids and on the surface of epididymal, ejaculated and capacitated spermatozoa. Biochim Biophys Acta 1200: 48-54

Dragovic RA, Ritter LJ, Schulz SJ, Amato F, Thompson JG, Armstrong DT, Gilchrist RB (2007) Oocyte-secreted factor activation of SMAD 2/3 signaling enables initiation of mouse cumulus cell expansion. Biol Reprod 76: 848-857

Druart X, Gatti JL, Dacheux F, Dacheux JL (1994) Analysis by two-dimensional gel electrophoresis of ram epididymal secreted proteins. . Cell Mol Biol (Noisy-le-grand) 40: 91-93

Edwards RG (1965) Maturation in vitro of mouse, sheep, cow, pig, rhesus monkey and human ovarian oocytes. Nature 208: 349-351

Eid LN, Lorton SP, Parrish JJ (1994) Paternal influence on S-phase in the first cell cycle of the bovine embryo. Biol Reprod 51: 1232-1237

Ellederova Z, Halada P, Man P, Kubelka M, Motlik J, Kovarova H (2004) Protein patterns of pig oocytes during in vitro maturation. Biol Reprod 71: 1533-1539

Eng LA, Oliphant G (1978) Rabbit sperm reversible decapacitation by membrane stabilization with a highly purified glycoprotein from seminal plasma. Biol Reprod 19: 1083-1094

Eppig JJ (1991) Intercommunication between mammalian oocytes and companion somatic cells. Bioessays 13: 569-574

Eppig JJ, Chesnel F, Hirao Y, O'Brien MJ, Pendola FL, Watanabe S, Wigglesworth K (1997) Oocyte control of granulosa cell development: how and why. Hum Reprod 12: 127-132

Eppig JJ, Wigglesworth K, Chesnel F (1993) Secretion of cumulus expansion enabling factor by mouse oocytes: relationship to oocyte growth and competence to resume meiosis. Dev Biol 158: 400-409

Erikson DW, Way AL, Chapman DA, Killian GJ (2007) Detection of osteopontin on Holstein bull spermatozoa, in cauda epididymal fluid and testis homogenates, and its potential role in bovine fertilization. Reproduction 133: 909-917

Fadiel A, Anidi I, Eichenbaum KD (2005) Farm animal genomics and informatics: an update. Nucleic Acids Res 33: 6308-6318

Fair T, Carter F, Park S, Evans AC, Lonergan P (2007) Global gene expression analysis during bovine oocyte in vitro maturation. Theriogenology 68 Suppl 1: S91-97

Fan HY, Sun QY (2004) Involvement of mitogen-activated protein kinase cascade during oocyte maturation and fertilization in mammals. Biol Reprod 70: 535-547

Fatehi AN, Zeinstra EC, Kooij RV, Colenbrander B, Bevers MM (2002) Effect of cumulus cell removal of in vitro matured bovine oocytes prior to in vitro fertilization on subsequent cleavage rate. Theriogenology 57: 1347-1355

Ferguson PL, Smith RD (2003) Proteome analysis by mass spectrometry. Annu Rev Biophys Biomol Struct 32: 399-424

Fissore RA, He CL, Vande Woude GF (1996) Potential role of mitogen-activated protein kinase during meiosis resumption in bovine oocytes. Biol Reprod 55: 1261-1270

Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE, Yates JR, Carucci DJ (2002) A proteomic view of the Plasmodium falciparum life cycle. Nature 419: 520-526

Flowers WL (1998) Boar fertility and artificial insemination. 45-52

Fouchecourt S, Metayer S, Locatelli A, Dacheux F, Dacheux JL (2000) Stallion epididymal fluid proteome: qualitative and quantitative characterization; secretion and dynamic changes of major proteins. Biol Reprod 62: 1790-1803

Frenette G, Girouard J, Sullivan R (2006) Comparison between epididymosomes collected in the intraluminal compartment of the bovine caput and cauda epididymidis. Biol Reprod 75: 885-890

Fukui Y, Sakuma Y (1980) Maturation of bovine oocytes cultured in vitro: relation to ovarian activity, follicular size and the presence or absence of cumulus cells. Biol Reprod 22: 669-673

Galantino-Homer HL, Kopf GS (2000) Purification and identificationof major bovine sperm proteins undergoing cAMP-dependentprotein tyrosine phosphorylation during in vitro capacitation. Biology of reproduction (Suppl) 62: 78

Galli C, Lazzari G (2008) The manipulation of gametes and embryos in farm animals. Reprod Domest Anim 43 Suppl 2: 1-7

Gao J, Opiteck GJ, Friedrichs MS, Dongre AR, Hefta SA (2003) Changes in the protein expression of yeast as a function of carbon source. J Proteome Res 2: 643-649

Gatti JL, Metayer S, Belghazi M, Dacheux F, Dacheux JL (2005) Identification, proteomic profiling, and origin of ram epididymal fluid exosome-like vesicles. Biol Reprod 72: 1452-1465

Gerena RL, Irikura D, Urade Y, Eguchi N, Chapman DA, Killian GJ (1998) Identification of a fertility-associated protein in bull seminal plasma as lipocalin-type prostaglandin D synthase. Biol Reprod 58: 826-833

Gilchrist RB, Lane M, Thompson JG (2008) Oocyte-secreted factors: regulators of cumulus cell function and oocyte quality. Hum Reprod Update 14: 159-177

Gilchrist RB, Morrissey MP, Ritter LJ, Armstrong DT (2003) Comparison of oocyte factors and transforming growth factor-beta in the regulation of DNA synthesis in bovine granulosa cells. Mol Cell Endocrinol 201: 87-95

Gilchrist RB, Ritter LJ, Armstrong DT (2004) Oocyte-somatic cell interactions during follicle development in mammals. Anim Reprod Sci 82-83: 431-446

Gilchrist RB, Ritter LJ, Myllymaa S, Kaivo-Oja N, Dragovic RA, Hickey TE, Ritvos O, Mottershead DG (2006) Molecular basis of oocyte-paracrine signalling that promotes granulosa cell proliferation. J Cell Sci 119: 3811-3821

Grazul-Bilska AT, Reynolds LP, Redmer DA (1997) Gap junctions in the ovaries. Biol Reprod 57: 947-957

Guerrera IC, Kleiner O (2005) Application of mass spectrometry in proteomics. Biosci Rep 25: 71-93

Guo Y, Fu Z, Van Eyk JE (2007) A proteomic primer for the clinician. Proc Am Thorac Soc 4: 9-17

Hachey DL, Chaurand P (2004) Proteomics in reproductive medicine: the technology for separation and identification of proteins. J Reprod Immunol 63: 61-73

Haqqani AS, Kelly JF, Stanimirovic DB (2008) Quantitative protein profiling by mass spectrometry using isotope-coded affinity tags. Methods Mol Biol 439: 225-240

Harayama H, Magargee SF, Kunze E, Shidara O, Iwamoto E, Arikawa S, Miyake M, Kato S, Hammerstedt RH (1999) Changes in epididymal protein anti-agglutinin on ejaculated boar spermatozoa during capacitation in vitro. Reprod Fertil Dev 11: 193-199

Harrouk W, Clarke HJ (1995) Mitogen-activated protein (MAP) kinase during the acquisition of meiotic competence by growing oocytes of the mouse. Mol Reprod Dev 41: 29-36

Hassan MI, Kumar V, Kashav T, Alam N, Singh TP, Yadav S (2007) Proteomic approach for purification of seminal plasma proteins involved in tumor proliferation. J Sep Sci 30: 1979-1988

Heikinheimo O, Gibbons WE (1998) The molecular mechanisms of oocyte maturation and early embryonic development are unveiling new insights into reproductive medicine. Mol Hum Reprod 4: 745-756

Hemler ME (2001) Specific tetraspanin functions. J Cell Biol 155: 1103-1107

Henault MA, Killian GJ, Kavanaugh JF, Griel LC, Jr. (1995) Effect of accessory sex gland fluid from bulls of differing fertilities on the ability of cauda epididymal sperm to penetrate zona-free bovine oocytes. Biol Reprod 52: 390-397

Honda A, Yamagata K, Sugiura S, Watanabe K, Baba T (2002) A mouse serine protease TESP5 is selectively included into lipid rafts of sperm membrane presumably as a glycosylphosphatidylinositol-anchored protein. J Biol Chem 277: 16976-16984

Horan AH, Bedford JM (1972) Development of the fertilizing ability of spermatozoa in the epididymis of the Syrian hamster. J Reprod Fertil 30: 417-423

Hough SR, Parks JE (1994) Platelet-activating factor acetylhydrolase activity in seminal plasma from the bull, stallion, rabbit, and rooster. Biol Reprod 50: 912-916

http://www.biomarkersconsortium.org

Hu, Z. Mellor, Snitkin, S. E., & DeLisi, C. (2008). VisANT: an integrative framework for networks in systems biology, Briefings in Bioinformatics 9:317-325

Hussein TS, Froiland DA, Amato F, Thompson JG, Gilchrist RB (2005) Oocytes prevent cumulus cell apoptosis by maintaining a morphogenic paracrine gradient of bone morphogenetic proteins. J Cell Sci 118: 5257-5268

Ibrahim NM, Gilbert GR, Loseth KJ, Crabo BG (2000) Correlation between clusterin-positive spermatozoa determined by flow cytometry in bull semen and fertility. J Androl 21: 887-894

Ingenuity Systems.(2006). IPA 4.0 Help Manual v.3.

Inoue M, Naito K, Aoki F, Toyoda Y, Sato E (1995) Activation of mitogen-activated protein kinase during meiotic maturation in porcine oocytes. Zygote 3: 265-271

Inoue N, Ikawa M, Isotani A, Okabe M (2005) The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. Nature 434: 234-238

Janzen RG, Mably ER, Tamaoki T, Church RB, Lorscheider FL (1982) Synthesis of alpha-fetoprotein by the pre-implantation and post-implantation bovine embryo. J Reprod Fertil 65: 105-110

Jeon K, Kim EY, Tae JC, Lee CH, Lee KS, Kim YO, Jeong DK, Cho SK, Kim JH, Lee HY, Riu KZ, Cho SG, Park SP (2008) Survivin protein expression in bovine follicular oocytes and their in vitro developmental competence. Anim Reprod Sci 108: 319-333

Johnson LR, Foster JA, Haig-Ladewig L, VanScoy H, Rubin CS, Moss SB, Gerton GL (1997) Assembly of AKAP82, a protein kinase A anchor protein, into the fibrous sheath of mouse sperm. Dev Biol 192: 340-350

Ka HH, Sawai K, Wang WH, Im KS, Niwa K (1997) Amino acids in maturation medium and presence of cumulus cells at fertilization promote male pronuclear formation in porcine oocytes matured and penetrated in vitro. Biol Reprod 57: 1478-1483

Katz-Jaffe MG, Gardner DK (2007) Embryology in the era of proteomics. Theriogenology 68 Suppl 1: S125-130

Katz-Jaffe MG, McCallie BR, Preis KA, Filipovits J, Gardner DK (2009) Transcriptome analysis of in vivo and in vitro matured bovine MII oocytes. Theriogenology 71: 939-946

Khan SA, Suryawanshi AR, Ranpura SA, Jadhav SV, Khole VV (2009) Identification of novel immunodominant epididymal sperm proteins using combinatorial approach. Reproduction 138: 81-93

Killian GJ (1999) High - Fertility Proteins Enhance Reproduction Rates in Dairy Cattle.

Killian GJ, Chapman DA, Rogowski LA (1993) Fertility-associated proteins in Holstein bull seminal plasma. Biol Reprod 49: 1202-1207

Kitano H (2002) Systems biology: a brief overview. Science 295: 1662-1664

Krawetz SA (2005) Paternal contribution: new insights and future challenges. Nat Rev Genet 6: 633-642

Lalancette C, Faure RL, Leclerc P (2006) Identification of the proteins present in the bull sperm cytosolic fraction enriched in tyrosine kinase activity: a proteomic approach. Proteomics 6: 4523-4540

Le Naour F, Rubinstein E, Jasmin C, Prenant M, Boucheix C (2000) Severely reduced female fertility in CD9-deficient mice. Science 287: 319-321

Lee J, Miyano T, Moor RM (2000) Localisation of phosphorylated MAP kinase during the transition from meiosis I to meiosis II in pig oocytes. Zygote 8: 119-125

Lee KH (2001) Proteomics: a technology-driven and technology-limited discovery science. Trends Biotechnol 19: 217-222

Lefievre L, Barratt CL, Harper CV, Conner SJ, Flesch FM, Deeks E, Moseley FL, Pixton KL, Brewis IA, Publicover SJ (2003) Physiological and proteomic approaches to studying prefertilization events in the human. Reprod Biomed Online 7: 419-427

Lewis SE (2005) Gene Ontology: looking backwards and forwards. Genome Biol 6: 103

Li R, Norman RJ, Armstrong DT, Gilchrist RB (2000) Oocyte-secreted factor(s) determine functional differences between bovine mural granulosa cells and cumulus cells. Biol Reprod 63: 839-845

Liu H, Sadygov RG, Yates JR, 3rd (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76: 4193-4201

Luo KL, Fan LQ, Lu GX (2008) [Proteomic identification of necrozoospermia-related proteins]. Zhonghua Nan Ke Xue 14: 431-435

Mandal A, Naaby-Hansen S, Wolkowicz MJ, Klotz K, Shetty J, Retief JD, Coonrod SA, Kinter M, Sherman N, Cesar F, Flickinger CJ, Herr JC (1999) FSP95, a testis-specific 95-kilodalton fibrous sheath antigen that undergoes tyrosine phosphorylation in capacitated human spermatozoa. Biol Reprod 61: 1184-1197

Manjunath P, Therien I (2002) Role of seminal plasma phospholipid-binding proteins in sperm membrane lipid modification that occurs during capacitation. J Reprod Immunol 53: 109-119

Martinez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R (2008) Identification of proteomic differences in asthenozoospermic sperm samples. Hum Reprod 23: 783-791

Martinez-Heredia J, Estanyol JM, Ballesca JL, Oliva R (2006) Proteomic identification of human sperm proteins. Proteomics 6: 4356-4369

Massicotte L, Coenen K, Mourot M, Sirard MA (2006) Maternal housekeeping proteins translated during bovine oocyte maturation and early embryo development. Proteomics 6: 3811-3820

Matt P, Fu Z, Fu Q, Van Eyk JE (2008) Biomarker discovery: proteome fractionation and separation in biological samples. Physiol Genomics 33: 12-17

Mattioli M, Galeati G, Bacci ML, Barboni B (1991) Changes in maturation-promoting activity in the cytoplasm of pig oocytes throughout maturation. Mol Reprod Dev 30: 119-125

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, Luthe DS, Bridges SM, Burgess SC (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics 7: 229

McCauley TC, Zhang H, Bellin ME, Ax RL (1999) Purification and characterization of fertility-associated antigen (FAA) in bovine seminal fluid. Mol Reprod Dev 54: 145-153

McCauley TC, Zhang HM, Bellin ME, Ax RL (2001) Identification of a heparin-binding protein in bovine seminal fluid as tissue inhibitor of metalloproteinases-2. Mol Reprod Dev 58: 336-341

McNatty KP, Moore LG, Hudson NL, Quirke LD, Lawrence SB, Reader K, Hanrahan JP, Smith P, Groome NP, Laitinen M, Ritvos O, Juengel JL (2004) The oocyte and its role in regulating ovulation rate: a new paradigm in reproductive biology. Reproduction 128: 379-386

Mehlmann LM (2005) Stops and starts in mammalian oocytes: recent advances in understanding the regulation of meiotic arrest and oocyte maturation. Reproduction 130: 791-799

Michaud GA, Salcius M, Zhou F, Bangham R, Bonin J, Guo H, Snyder M, Predki PF, Schweitzer BI (2003) Analyzing antibody specificity with whole proteome microarrays. Nat Biotechnol 21: 1509-1512

Miyado K, Yamada G, Yamada S, Hasuwa H, Nakamura Y, Ryu F, Suzuki K, Kosai K, Inoue K, Ogura A, Okabe M, Mekada E (2000) Requirement of CD9 on the egg plasma membrane for fertilization. Science 287: 321-324

Mlecnik B, Scheideler M, Hackl H, Hartler J, Sanchez-Cabo F, Trajanoski Z (2005) PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways.Nucleic Acids Res 2005 , 33:W633-7.

Moor RM, Mattioli M, Ding J, Nagai T (1990) Maturation of pig oocytes in vivo and in vitro. J Reprod Fertil Suppl 40: 197-210

Mor G, Visintin I, Lai Y, Zhao H, Schwartz P, Rutherford T, Yue L, Bray-Ward P, Ward DC (2005) Serum protein markers for early detection of ovarian cancer. Proc Natl Acad Sci U S A 102: 7677-7682

Mori T, Amano T, Shimizu H (2000) Roles of gap junctional communication of cumulus cells in cytoplasmic maturation of porcine oocytes cultured in vitro. Biol Reprod 62: 913-919

Moss SB, Gerton GL (2001) A-kinase anchor proteins in endocrine systems and reproduction. Trends Endocrinol Metab 12: 434-440

Moss SB, Turner RM, Burkert KL, VanScoy Butt H, Gerton GL (1999) Conservation and function of a bovine sperm A-kinase anchor protein homologous to mouse AKAP82. Biol Reprod 61: 335-342

Motlik J, Kubelka M (1990) Cell-cycle aspects of growth and maturation of mammalian oocytes. Mol Reprod Dev 27: 366-375

Motlik J, Pavlok A, Kubelka M, Kalous J, Kalab P (1998) Interplay between CDC2 kinase and MAP kinase pathway during maturation of mammalian oocytes. Theriogenology 49: 461-469

Moura AA, Chapman DA, Koc H, Killian GJ (2006) Proteins of the cauda epididymal fluid associated with fertility of mature dairy bulls. J Androl 27: 534-541

Moura AA, Chapman DA, Koc H, Killian GJ (2007) A comprehensive proteomic analysis of the accessory sex gland fluid from mature Holstein bulls. Anim Reprod Sci 98: 169-188

Moura AA, Koc H, Chapman DA, Killian GJ (2006) Identification of proteins in the accessory sex gland fluid associated with fertility indexes of dairy bulls: a proteomic approach. J Androl 27: 201-211

Nanduri B, Lawrence ML, Peddinti DS, Burgess SC (2008) Effects of Subminimum Inhibitory Concentrations of Antibiotics on the Pasteurella multocida Proteome: A Systems Approach. Comp Funct Genomics: 254836

Nikolopoulou M, Soucek DA, Vary JC (1985) Changes in the lipid content of boar sperm plasma membranes during epididymal maturation. Biochim Biophys Acta 815: 486-498

Nixon B, Aitken RJ, McLaughlin EA (2007) New insights into the molecular mechanisms of sperm-egg interaction. Cell Mol Life Sci 64: 1805-1823

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4: 1487-1502

Oliva R, de Mateo S, Estanyol JM (2009) Sperm cell proteomics. Proteomics 9: 1004-1017

Oliva R, Martinez-Heredia J, de Mateo S, Gazquez C, Oriola J, Estanyol JM, Guimera M, Balasch J, Ballesca JL (2007) Proteomics of human spermatozoa, protamine content and assisted reproduction outcome. Soc Reprod Fertil Suppl 65: 527-530

Oliva R, Martinez-Heredia J, Estanyol JM (2008) Proteomics in the study of the sperm cell composition, differentiation and function. Syst Biol Reprod Med 54: 23-36

Olson GE, Hinton BT (1985) Regional differences in luminal fluid polypeptides of the rat testis and epididymis revealed by two-dimensional gel electrophoresis. J Androl 6: 20-34

Ortega-Ferrusola C, Sotillo-Galan Y, Varela-Fernandez E, Gallardo-Bolanos JM, Muriel A, Gonzalez-Fernandez L, Tapia JA, Pena FJ (2008) Detection of "apoptosis-like" changes during the cryopreservation process in equine sperm. J Androl 29: 213-221

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

Pena FJ, Saravia F, Johannisson A, Wallgren M, Rodriguez-Martinez H (2007) Detection of early changes in sperm membrane integrity pre-freezing can estimate post-thaw quality of boar spermatozoa. Anim Reprod Sci 97: 74-83

Pennetier S, Perreau C, Uzbekova S, Thelie A, Delaleu B, Mermillod P, Dalbies-Tran R (2006) MATER protein expression and intracellular localization throughout folliculogenesis and preimplantation embryo development in the bovine. BMC Dev Biol 6: 26

Pilch B, Mann M (2006) Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol 7: R40

Pixton KL, Deeks ED, Flesch FM, Moseley FL, Bjorndahl L, Ashton PR, Barratt CL, Brewis IA (2004) Sperm proteome mapping of a patient who experienced failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: case report. Hum Reprod 19: 1438-1447

Puri P, Myers K, Kline D, Vijayaraghavan S (2008) Proteomic analysis of bovine sperm YWHA binding partners identify proteins involved in signaling and metabolism. Biol Reprod 79: 1183-1191

Reinders J, Lewandrowski U, Moebius J, Wagner Y, Sickmann A (2004) Challenges in mass spectrometry-based proteomics. Proteomics 4: 3686-3703

Richter JD (1991) Translational control during early development. Bioessays 13: 179-183

Richter JD (1999) Cytoplasmic polyadenylation in development and beyond. Microbiol Mol Biol Rev 63: 446-456

Robertson DM, McNeilage J (2004) Inhibins as biomarkers for reproductive cancers. Semin Reprod Med 22: 219-225

Ross P, Kan FW, Antaki P, Vigneault N, Chapdelaine A, Roberts KD (1990) Protein synthesis and secretion in the human epididymis and immunoreactivity with sperm antibodies. Mol Reprod Dev 26: 12-23

Rubin CS (1994) A kinase anchor proteins and the intracellular targeting of signals carried by cyclic AMP. Biochim Biophys Acta 1224: 467-479

Rubinstein S, Breitbart H (1991) Role of spermine in mammalian sperm capacitation and acrosome reaction. Biochem J 278 (Pt 1): 25-28

Saacke RG, Dalton JC, Nadir S, Nebel RL, Bame JH (2000) Relationship of seminal traits and insemination time to fertilization rate and embryo quality. Anim Reprod Sci 60-61: 663-677

Salicioni AM, Platt MD, Wertheimer EV, Arcelay E, Allaire A, Sosnik J, Visconti PE (2007) Signalling pathways involved in sperm capacitation. Soc Reprod Fertil Suppl 65: 245-259

Sasseville M, Albuz FK, Cote N, Guillemette C, Gilchrist RB, Richard FJ (2009) Characterization of Novel Phosphodiesterases in the Bovine Ovarian Follicle. Biol Reprod

Sato K, Iwasaki T, Sakakibara K, Itakura S, Fukami Y (2002) Towards the molecular dissection of fertilization signaling: Our functional genomic/proteomic strategies. Proteomics 2: 1079-1089

Sawada H, Sakai N, Abe Y, Tanaka E, Takahashi Y, Fujino J, Kodama E, Takizawa S, Yokosawa H (2002) Extracellular ubiquitination and proteasome-mediated degradation of the ascidian sperm receptor. Proc Natl Acad Sci U S A 99: 1223-1228

Schmidt C, Urlaub H (2009) iTRAQ-Labeling of In-Gel Digested Proteins for Relative Quantification. Methods Mol Biol 564: 207-226

Secciani F, Bianchi L, Ermini L, Cianti R, Armini A, La Sala GB, Focarelli R, Bini L, Rosati F (2009) Protein profile of capacitated versus ejaculated human sperm. J Proteome Res 8: 3377-3389

Sevinsky JR, Cargile BJ, Bunger MK, Meng F, Yates NA, Hendrickson RC, Stephenson JL, Jr. (2008) Whole genome searching with shotgun proteomic data: applications for genome annotation. J Proteome Res 7: 80-88

Simon AM, Goodenough DA, Li E, Paul DL (1997) Female infertility in mice lacking connexin 37. Nature 385: 525-529

Secciani F, Bianchi L, Ermini L, Cianti R, Armini A, La Sala GB, Focarelli R, Bini L, Rosati F (2009) Protein profile of capacitated versus ejaculated human sperm. J Proteome Res 8: 3377-3389

Sevinsky JR, Cargile BJ, Bunger MK, Meng F, Yates NA, Hendrickson RC, Stephenson JL, Jr. (2008) Whole genome searching with shotgun proteomic data: applications for genome annotation. J Proteome Res 7: 80-88

Simon AM, Goodenough DA, Li E, Paul DL (1997) Female infertility in mice lacking connexin 37. Nature 385: 525-529

Sirard MA, Dufort I, Coenen K, Tremblay K, Massicotte L, Robert C (2003) The use of genomics and proteomics to understand oocyte and early embryo functions in farm animals. Reprod Suppl 61: 117-129

Sirard MA, Richard F, Mayes M (1998) Controlling meiotic resumption in bovine oocytes: a review. Theriogenology 49: 483-497

Smith SL, Everts RE, Sung LY, Du F, Page RL, Henderson B, Rodriguez-Zas SL, Nedambale TL, Renard JP, Lewin HA, Yang X, Tian XC (2009) Gene expression profiling of single bovine embryos uncovers significant effects of in vitro maturation, fertilization and culture. Mol Reprod Dev 76: 38-47

Smitz JE, Cortvrindt RG (2004) In vitro growth and maturation of oocytes in human and non-human primates. Gynecol Obstet Invest 57: 18-21

Sobajima T, Aoki F, Kohmoto K (1993) Activation of mitogen-activated protein kinase during meiotic maturation in mouse oocytes. J Reprod Fertil 97: 389-394

Srinivas PR, Verma M, Zhao Y, Srivastava S (2002) Proteomics for cancer biomarker discovery. Clin Chem 48: 1160-1169

Stein KK, Go JC, Lane WS, Primakoff P, Myles DG (2006) Proteomic analysis of sperm regions that mediate sperm-egg interactions. Proteomics 6: 3533-3543

Stoevesandt O, Taussig MJ, He M (2009) Protein microarrays: high-throughput tools for proteomics. Expert Rev Proteomics 6: 145-157

Strzezek J, Wysocki P, Kordan W, Kuklinska M (2005) Proteomics of boar seminal plasma - current studies and possibility of their application in biotechnology of animal reproduction. Reprod Biol 5: 279-290

Sugiura K, Eppig JJ (2005) Society for Reproductive Biology Founders' Lecture 2005. Control of metabolic cooperativity between oocytes and their companion granulosa cells by mouse oocytes. Reprod Fertil Dev 17: 667-674

Sun QY, Breitbart H, Schatten H (1999) Role of the MAPK cascade in mammalian germ cells. Reprod Fertil Dev 11: 443-450

Sun QY, Lai L, Bonk A, Prather RS, Schatten H (2001) Cytoplasmic changes in relation to nuclear maturation and early embryo developmental potential of porcine oocytes: effects of gonadotropins, cumulus cells, follicular size, and protein synthesis inhibition. Mol Reprod Dev 59: 192-198

Sun QY, Lai L, Park KW, Kuhholzer B, Prather RS, Schatten H (2001) Dynamic events are differently mediated by microfilaments, microtubules, and mitogen-activated protein kinase during porcine oocyte maturation and fertilization in vitro. Biol Reprod 64: 879-889

Sun QY, Nagai T (2003) Molecular mechanisms underlying pig oocyte maturation and fertilization. J Reprod Dev 49: 347-359

Sutovsky P (2009) Sperm-egg adhesion and fusion in mammals. Expert Rev Mol Med 11: e11

Sutovsky P, Schatten G (2000) Paternal contributions to the mammalian zygote: fertilization after sperm-egg fusion. Int Rev Cytol 195: 1-65

Sutton ML, Cetica PD, Beconi MT, Kind KL, Gilchrist RB, Thompson JG (2003) Influence of oocyte-secreted factors and culture duration on the metabolic activity of bovine cumulus cell complexes. Reproduction 126: 27-34

Syntin P, Dacheux F, Druart X, Gatti JL, Okamura N, Dacheux JL (1996) Characterization and identification of proteins secreted in the various regions of the adult boar epididymis. Biol Reprod 55: 956-974

Talbot P, Shur BD, Myles DG (2003) Cell adhesion and fertilization: steps in oocyte transport, sperm-zona pellucida interactions, and sperm-egg fusion. Biol Reprod 68: 1-9

Tan Y, Fan L, Luo K, Zhu W, Lu G (2004) [Establishment of the two-dimensional gel electrophoretic protein map of the human sperm head]. Zhonghua Nan Ke Xue 10: 886-889

Tanghe S, Van Soom A, Nauwynck H, Coryn M, de Kruif A (2002) Minireview: Functions of the cumulus oophorus during oocyte maturation, ovulation, and fertilization. Mol Reprod Dev 61: 414-424

Tomek W, Torner H, Kanitz W (2002) Comparative analysis of protein synthesis, transcription and cytoplasmic polyadenylation of mRNA during maturation of bovine oocytes in vitro. Reprod Domest Anim 37: 86-91

Tong ZB, Bondy CA, Zhou J, Nelson LM (2002) A human homologue of mouse Mater, a maternal effect gene essential for early embryonic development. Hum Reprod 17: 903-911

Tong ZB, Gold L, Pfeifer KE, Dorward H, Lee E, Bondy CA, Dean J, Nelson LM (2000) Mater, a maternal effect gene required for early embryonic development in mice. Nat Genet 26: 267-268

Topfer-Petersen E, Ekhlasi-Hundrieser M, Tsolova M, Leeb T, Kirchhoff C, Muller P (2005) Structure and function of secretory proteins of the male genital tract. Andrologia 37: 202-204

Tulsiani DR, Zeng HT, Abou-Haila A (2007) Biology of sperm capacitation: evidence for multiple signalling pathways. Soc Reprod Fertil Suppl 63: 257-272

Turck CW, Falick AM, Kowalak JA, Lane WS, Lilley KS, Phinney BS, Weintraub ST, Witkowska HE, Yates NA (2007) The Association of Biomolecular Resource Facilities Proteomics Research Group 2006 study: relative protein quantitation. Mol Cell Proteomics 6: 1291-1298

Urner F, Leppens-Luisier G, Sakkas D (2001) Protein tyrosine phosphorylation in sperm during gamete interaction in the mouse: the influence of glucose. Biol Reprod 64: 1350-1357

Urner F, Sakkas D (2003) Protein phosphorylation in mammalian spermatozoa. Reproduction 125: 17-26

Uzbekova S, Arlot-Bonnemains Y, Dupont J, Dalbies-Tran R, Papillier P, Pennetier S, Thelie A, Perreau C, Mermillod P, Prigent C, Uzbekov R (2008) Spatio-temporal expression patterns of aurora kinases a, B, and C and cytoplasmic polyadenylation-element-binding protein in bovine oocytes during meiotic maturation. Biol Reprod 78: 218-233

Vadnais ML, Galantino-Homer HL, Althouse GC (2007) Current concepts of molecular events during bovine and porcine spermatozoa capacitation. Arch Androl 53: 109-123

Vadnais ML, Roberts KP (2007) Effects of seminal plasma on cooling-induced capacitative changes in boar sperm. J Androl 28: 416-422

Vallee M, Aiba K, Piao Y, Palin MF, Ko MS, Sirard MA (2008) Comparative analysis of oocyte transcript profiles reveals a high degree of conservation among species. Reproduction 135: 439-448

Van Montfoort AP, Geraedts JP, Dumoulin JC, Stassen AP, Evers JL, Ayoubi TA (2008) Differential gene expression in cumulus cells as a prognostic indicator of embryo viability: a microarray analysis. Mol Hum Reprod 14: 157-168

Van Regenmortel MH (2004) Biological complexity emerges from the ashes of genetic reductionism. J Mol Recognit 17: 145-148

Van Regenmortel MH (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. EMBO Rep 5: 1016-1020

Venugopal A, Chaerkady R, Pandey A (2009) Application of mass spectrometry-based proteomics for biomarker discovery in neurological disorders. Annals of indian academy of neurology 12: 3-11

Vijayaraghavan S, Liberty GA, Mohan J, Winfrey VP, Olson GE, Carr DW (1999) Isolation and molecular characterization of AKAP110, a novel, sperm-specific protein kinase A-anchoring protein. Mol Endocrinol 13: 705-717

Visconti PE, Bailey JL, Moore GD, Pan D, Olds-Clarke P, Kopf GS (1995) Capacitation of mouse spermatozoa. I. Correlation between the capacitation state and protein tyrosine phosphorylation. Development 121: 1129-1137

Visconti PE, Stewart-Savage J, Blasco A, Battaglia L, Miranda P, Kopf GS, Tezon JG (1999) Roles of bicarbonate, cAMP, and protein tyrosine phosphorylation on capacitation and the spontaneous acrosome reaction of hamster sperm. Biol Reprod 61: 76-84

von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krüger B, Snel B, Bork P (2007) STRING 7–recent developments in the integration and prediction of protein interactions.Nucleic Acids Res 2007 , 35:D358-62.

Vozzi C, Formenton A, Chanson A, Senn A, Sahli R, Shaw P, Nicod P, Germond M, Haefliger JA (2001) Involvement of connexin 43 in meiotic maturation of bovine oocytes. Reproduction 122: 619-628

Wassarman PM, Litscher ES (2001) Towards the molecular basis of sperm and egg interaction during mammalian fertilization. Cells Tissues Organs 168: 36-45

Wehrend A, Meinecke B (2001) Kinetics of meiotic progression, M-phase promoting factor (MPF) and mitogen-activated protein kinase (MAP kinase) activities during in vitro maturation of porcine and bovine oocytes: species specific differences in the length of the meiotic stages. Anim Reprod Sci 66: 175-184

Wickens M, Anderson P, Jackson RJ (1997) Life and death in the cytoplasm: messages from the 3' end. Curr Opin Genet Dev 7: 220-232

Wongsrikeao P, Kaneshige Y, Ooki R, Taniguchi M, Agung B, Nii M, Otoi T (2005) Effect of the removal of cumulus cells on the nuclear maturation, fertilization and development of porcine oocytes. Reprod Domest Anim 40: 166-170

Wu B, Ignotz G, Currie WB, Yang X (1997) Expression of Mos proto-oncoprotein in bovine oocytes during maturation in vitro. Biol Reprod 56: 260-265

Wu X, Viveiros MM, Eppig JJ, Bai Y, Fitzpatrick SL, Matzuk MM (2003) Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. Nat Genet 33: 187-191

Yanagimachi R (1994) Fertility of mammalian spermatozoa: its development and relativity. Zygote 2: 371-372

Yang X, Kubota C, Suzuki H, Taneja M, Bols PE, Presicce GA (1998) Control of oocyte maturation in cows--biological factors. Theriogenology 49: 471-482

Ying Y, Chow PH, O WS (1998) Effects of male accessory sex glands on deoxyribonucleic acid synthesis in the first cell cycle of golden hamster embryos. Biol Reprod 58: 659-663

Yuryev A, Mulyukov Z, Kotelnikova E, Maslov S, Egorov S, Nikitin A. (2006). Automatic pathway building in biological association networks.BMC Bioinformatics, 7, 171.

Zeng F, Baldwin DA, Schultz RM (2004) Transcript profiling during preimplantation mouse development. Dev Biol 272: 483-496

Zou M, Wang HF, Hu JM (2008) [The application of proteomics in epididymis researches]. Sheng Li Ke Xue Jin Zhan 39: 139-144

CHAPTER II

BOVINE GERMINAL VESICLE OOCYTE AND CUMULUS CELL PROTEOMICS[1]

[1]Reprint from Memili E, Peddinti D, Shack LA, Nanduri B, McCarthy F, Sagirkaya H, Burgess SC (2007) Bovine germinal vesicle oocyte and cumulus cell proteomics. Reproduction 133: 1107-1120.

REPRODUCTION
RESEARCH

Bovine germinal vesicle oocyte and cumulus cell proteomics

E Memili[1,2], D Peddinti[3,4], L A Shack[3,4], B Nanduri[3,4], F McCarthy[3,4], H Sagirkaya[1,5] and S C Burgess[2,3,4]

[1]*Department of Animal and Dairy Sciences, Mississippi State University, Starkville, Mississippi 39762-6100, USA,* [2]*Mississippi Agricultural and Forestry Experiment Station, Starkville, Mississippi 39762, USA,* [3]*College of Veterinary Medicine and* [4]*Institute for Digital Biology, Mississippi State University, Starkville, Mississippi 39762, USA and* [5]*Department of Reproduction and Artificial Insemination, Uludag University Veterinary Faculty, Gorukle-Bursa 16059, Turkey*

Correspondence should be addressed to E Memili; Email: em149@ads.msstate.edu

Abstract

Germinal vesicle (GV) breakdown is fundamental for maturation of fully grown, developmentally competent, mammalian oocytes. Bidirectional communication between oocytes and surrounding cumulus cells (CC) is essential for maturation of a competent oocyte. However, neither the factors involved in this communication nor the mechanisms of their actions are well defined. Here, we define the proteomes of GV oocytes and their surrounding CC, including membrane proteins, using proteomics in a bovine model. We found that 4395 proteins were expressed in the CC and 1092 proteins were expressed in oocytes. Further, 858 proteins were common to both the CC and the oocytes. This first comprehensive proteome analysis of bovine oocytes and CC not only provides a foundation for signaling and cell physiology at the GV stage of oocyte development, but are also valuable for comparative studies of other stages of oocyte development at the molecular level. Furthermore, some of these proteins may represent molecular biomarkers for developmental potential of oocytes.

Reproduction (2007) **133** 1107–1120

Introduction

Mammalian oocytes are the female gametes, their molecular biology uniquely establishes the program of life after fertilization and they are crucial in reproductive biology. Through a series of developmentally regulated events oocytes develop from primordial, primary, secondary, and tertiary follicles in the ovary. The oocyte is ovulated at the metaphase II (MII) stage. In *in vitro* conditions, however, the germinal vesicle (GV) oocyte completes MI before arresting at the MII. At fertilization, the MII oocyte and male gamete spermatozoa fuse (Matzuk *et al.* 2002, Senbon *et al.* 2003, Gilchrist *et al.* 2004). In meiotic development, nuclear maturation is manifest by GV breakdown (GVBD), condensation of chromosomes, realization of first meiosis (MI), and another arrest of development at the metaphase of the second meiosis (MII). In addition, these events related to nuclear maturation, significant changes occur in the cytoplasm including structural changes of organelles, major translational activity in which while many new proteins are synthesized, synthesis of others is terminated (Moor *et al.* 1990, Coenen *et al.* 2004). Developmentally competent MII oocytes require four periods of protein synthesis; namely, synthesis required for GVBD, MI, MII, and maintenance of MII (Khatir *et al.* 1998).

Oocytes do not develop in isolation; they are intimately involved with cumulus cells (CC). CC bind to the zona pellucida of the oocyte and connect to the oocyte cytoplasmic membrane to form a cumulus–oocyte complex (COC) through transzonal cytoplasmic process. Gap junctions allow transfer of small molecules between the oocyte and the CC (Albertini *et al.* 2001). Although this bidirectional communication and paracrine signaling between cumulus cell and oocyte are critical for oocyte growth and regulation of meiotic maturation of the oocyte (Eppig *et al.* 1993, De La Fuenta & Eppig 2001, Gilchrist *et al.* 2003, Sugiura & Eppig 2005), their nature and effects on the transcriptomes and proteomes of both are poorly defined.

Functional genomics methods now enable the analysis of transcriptomes and proteomes. From these, we can derive the molecular networks that define oocyte maturation, fertilization, and embryonic development (Pan *et al.* 2005, Sagirkaya *et al.* 2006). Here, we identify proteomes from GV stage oocytes and their surrounding CC using differential detergent fractionation (DDF) two-dimensional

ISSN 1470–1626 (paper) 1741–7899 (online)
DOI: 10.1530/REP-06-0149
Online version via www.reproduction-online.org

liquid chromatography followed by electrospray ionization tandem mass spectrometry (DDF 2-LC MS^2; McCarthy *et al.* 2005). We obtained proteomes of GV oocytes and their surrounding CC, including membrane proteins, using proteomics in a bovine model. We identified 4395 and 1092 cumulus cell- and oocyte-specific proteins. Further, 858 proteins were common to both the CC and the oocytes. Our work has provided the first experimental confirmation of 5360 of these 'predicted/hypothetical' proteins and is the first proteogenomic mapping of the recently sequenced bovine genome. Next, we used gene ontology (GO) to functionally annotate our data and this provided the largest single entry of GO annotations for the cow. We then interrogated our GO annotations to model oocyte and cumulus cell function. Specifically, because they underlie oocyte–cumulus interactions, we focus here on membrane, nuclear, and signaling proteins; receptor and ligand pairs; and transcription factors.

Materials and Methods

GV oocytes and CC

Ovaries were obtained from a local abattoir. Immature oocytes were aspirated from follicles (2–8 mm diameter) using an 18-gauge needle attached to a vacuum system (Sagirkaya *et al.* 2006). COCs (Fig. 1) were selected, washed three times in TL-HEPES supplemented with polyvinylpyrrolidone (3 mg/ml polyvinylpyrroline-40; Sigma), Na-pyruvate (0.2 mM), and gentamycin (25 μg/ml). To obtain oocytes free of CC, cumulus cell and oocyte complexes were vortexed in TL-HEPES (3 min), oocytes were collected under a stereomicroscope, further vortexed with hyaluronidase to remove adhering CC completely (3 min), washed three times in saline and stored in a cell lyses buffer at 4 °C until use. The lysis buffer consisted of digitonin (0.15 mM), EDTA (100 mM), Phenylmethylsulphonyl fluoride (100 mM), sucrose (103 mg/ml), NaCl (5.8 mg/ml), and PIPES (3 mg/ml) at pH 6.8. Oocytes were examined under a sterio microscope to ensure the complete removal of the CC. The CC removed from the oocytes after the first vortex were centrifuged, washed twice with saline, and the pellets resuspended in the lyses buffer and stored (4 °C) until use. Our method provided pure populations of CC and oocytes.

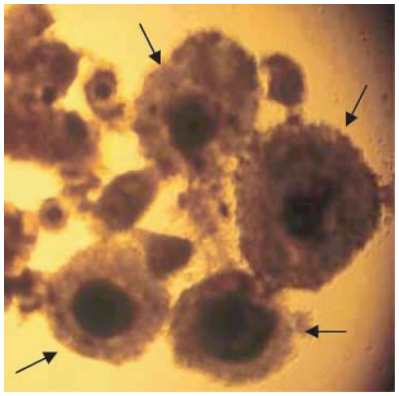

Figure 1 Morphological characteristics of bovine oocyte and their cumulus cells. Oocytes surrounded with several layers of cumulus cells (arrows) were used for this study. The is one of the most activity rich stages during oogenesis. Relatively compact cumulus cells undergo significant expansion during MI and MII stages of oocyte maturation.

Proteomics

Five hundred GV oocytes and their surrounding CC were each subjected to DDF exactly as described (McCarthy *et al.* 2005). The DDF fractions predominantly contain: DDF1, cytosolic; DDF2, membrane proteins; DDF3, cytoskeletal and nuclear proteins; and DDF4, remaining most insoluble proteins. The proteins in these DDF fractions were identified by two-dimensional liquid chromatography tandem mass spectrometry (2-DLCMS^2) exactly as described (McCarthy *et al.* 2006*a*,*b*). The resulting mass spectra were used to search subsets of the downloaded from the National Center for Biotechnology Institute (NCBI; 7/20/05) using TurboSEQUEST (Bioworks Browser 3.2; ThermoElectron, Waltham, MA, USA). We used a bovine subset of the nonredundant protein database (NRPD; 39 963 entries). Peptide matches were included only if they were ≥ 6 amino acids long and had $\Delta Cn > 0.1$ and Sequest cross-correlation (Xcorr) scores for charge states of 1.9, 2.2, and 3.75 for $+1, +2$, and $+3$ respectively (Washburn *et al.* 2001). All protein identifications and their associated MS data have been submitted to the PRoteomics IDEntifications database (PRIDE; Martens *et al.* 2005).

Modeling the proteomics data

We used GO and AgBase (McCarthy *et al.* 2006*a*,*b*) to identify the molecular functions, biological processes, and cellular components of the proteins in our dataset. Proteins without existing GO annotation, but between 70 and 90% sequence identities to presumptive orthologs with GO annotation, were GO-annotated using GOanna tool (McCarthy *et al.* 2006*a*). We next identified membrane, nuclear, and signaling proteins from our GO annotations and DDF profiles as described (McCarthy *et al.* 2006*a*). To identify receptor–ligand pairs, we used GO annotations and 'Bioinformatic Harvester' (Liebel *et al.* 2004) for proteins with human, mouse, or rat orthologs.

Since we did not find the ligands for all receptors in our data, we examined the amino acid sequences of these unidentified proteins to confirm whether they would be able to be identified by the DDF 2-DLCMS^2 method at all. To be reliably identified using our

proteomics method, a molecule must be a protein with tryptic peptides whose sequences are unique in the genome and these peptides must be within the detectable mass limits of the mass spectrometer. Also, post-translational modifications (such as glycosylation) can sterically hinder trypsin cleavage (Bark *et al.* 2001). We identified whether 'missing' proteins had peptide sequences that could be digested with trypsin (Gasteiger *et al.* 2005) whether the resulting peptides could be unique identifiers for the protein (using BLAST) and then whether or not these unique tryptic peptides would be detectable by mass spectrometry. Since 95% of our entire identified peptides were between 6 and 29 aa long (defined using our in-house 'peptide distribution analysis' program), we then removed all peptides that were <6 or >29 aa. The remaining 6–29mers were then analyzed for possible *N*- or *O*-linked glycosylation (Gupta & Brunak 2002, Julenius *et al.* 2005) that may cause steric hindrance during trypsin digestion.

To identify transcription factors we used GO annotations. We also manually inspected the entire dataset for terms that could identify transcription factors in the protein name: transcription factor, leucine zipper, DNA-binding protein, steroid hormone receptor, and corticoid receptor (http://www.gene-regulation.com/pub/databases/transfac/cl.html). Finally, we cataloged whether or not the transcription factors that we identified had previously been identified in oocytes or CC, by doing literature searches using PubMed.

Results

Proteomes

We identified 5253 and 1950 proteins in CC and GV stage oocytes respectively. Among these 858 (11.9%) were common to both cell types. Thus, this technique allowed us to identify 4395 and 1092 unique proteins in CC and oocytes respectively (Fig. 2). The lower number of proteins detected in the GV oocytes might be due to low concentration of proteins in the oocytes since fewer oocytes were used when compared with the CC. Among the 4395 proteins unique to CC, only 615 (14%) have been previously described; 3751 (85%) were annotated as 'predicted' (i.e. proteins are predicted based on sequence similarity to known proteins in other species and are frequently found in NRPD for species that have had their genomes sequenced (McCarthy *et al.* 2006*a*)); and 29 (0.65%) were annotated as 'hypothetical' (i.e. proteins predicted from nucleic acid sequences and that have not been shown to exist by experimental protein chemical evidence (Lubec *et al.* 2005)). Out of the 1092 proteins unique to oocytes, 141 (12.9%) were known, 947 (86.7%) were predicted, and only 4 (0.4%) were hypothetical. Among the 858 proteins common to both cell types, 191 (22.3%) were known, 662 (77.1%) were predicted, and only 5 (0.6%) were hypothetical (Fig. 2). This work, on only two cell types from a single organ, has contributed to the annotation of the newly sequenced bovine genome by experimentally confirming the *in vivo* expression of 5360 electronically predicted proteins (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). The proteins in DDF fractions were identified by (2-DLCMS^2). The applied method of peptide detection does not exclude the presence of a protein absolutely. Thus, the protein might be present although there was no peptide discovered.

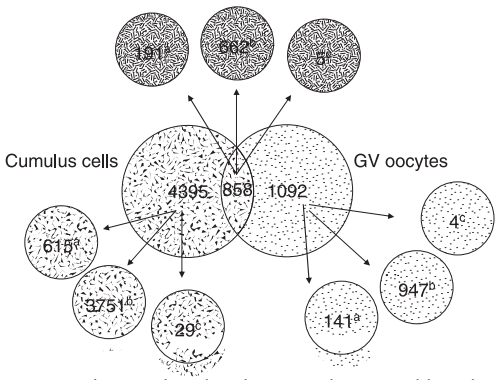

Figure 2 Distribution of predicted proteins, known and hypothetical proteins in oocytes, cumulus cells, and both cell types. [a]Known proteins; [b]predicted proteins; [c]hypothetical proteins.

A schematic of the experimental design and results indicating specific findings exhibited in specific tables and figures is shown in Fig. 3.

Membrane, intercellular signaling, and nuclear proteins

From the GO, we identified 378 membrane proteins (39% of the total known proteins): 266 unique to CC, 52 unique to oocytes, and 60 in both cell types. Our results agree with estimates that approximately one-third of all currently described genes code for membrane proteins (Wallin & von Heijne 1998, Stevens & Arkin 2000). Using GO associations, we identified 186 nuclear proteins: 73 unique to CC, 11 unique to oocytes, and 112 in both cell types. We also identified 36 proteins GO-annotated as involved in signaling: 25 unique to CC, 7 unique to CC oocyte, and 4 in both cell types. Only 154 (16.2%) proteins previously annotated as membrane proteins were present in DDF2. This difference between GO annotation and DFF fraction may be due to the presence of membrane proteins in fractions other than DDF2 (because proteins with greater numbers of trans-membrane domains tend to be present in the later DDF fractions); because some proteins may have membrane-bound isoforms that are not currently annotated as such (McCarthy *et al.* 2005) or due to errors in GO annotation.

Membrane and nuclear proteins are fundamental for inter- and intracellular signaling and are thus fundamental for modeling cell–cell interactions. We identified

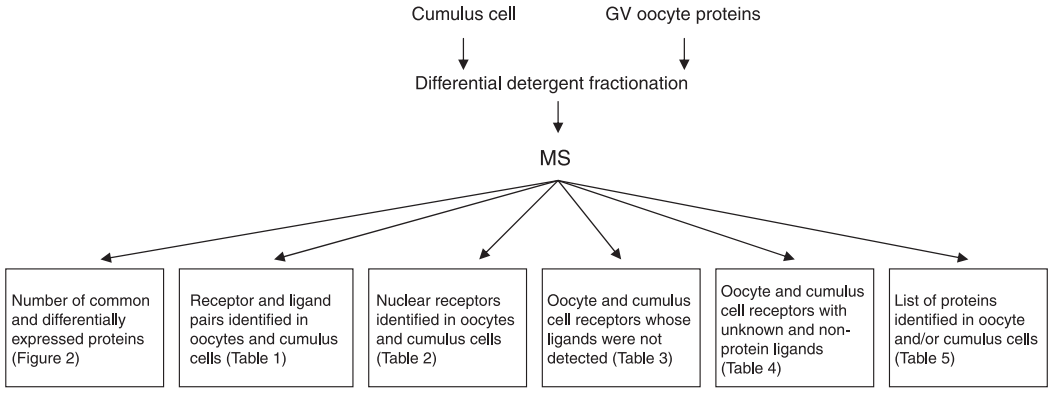

Figure 3 Schematic diagram of the experimental design and results indicating specific findings exhibited in specific tables or figures.

241 receptor–ligand pathways expressed in the CC and oocytes (Table 1). Among these were 18 growth factors (along with their binding proteins), which are likely involved in cell proliferation and cell differentiation. This is important in gametogenesis because oocyte-secreted growth factors play crucial roles in oocyte development and ovulation (Coskun *et al.* 1995). The cumulus cell dataset had numerically more growth factors (McCarthy *et al.* 2006*a*) when compared with oocytes (Matzuk *et al.* 2002) but, as a proportion of the total proteins identified from each cell type, the difference was much less striking: 0.29% (CC) versus 0.15% (oocytes). Endothelial growth factor-D, fibroblast (FGF), and epidermal growth factor (EGF) were present in both CC and oocytes, insulin-like growth factor (Igf) and transforming growth factor (TGF) were expressed only in CC (Table 1).

We also identified laminin receptors (cell adhesion molecules) in both oocytes and CC. These receptors interact with laminin, which is a major component of the basement membrane. Laminin receptors are thought to mediate the attachment, migration, and organization of cells into tissues by interacting with other extracellular matrix components (ECMs). Laminin-rich ECMs have contrasting regulatory effects on gap junction expression and thereby can alter specific cell–matrix interactions and gap junction-mediated cell-to-cell communication (Guo *et al.* 2001). This is directly relevant to the physiology of the COC, because the gap junctions between the CC and the oocyte allow transfer of molecules between CC and oocytes, as well as among the CC (Simon *et al.* 1997). We also observed 15 protein tyrosine phosphatase receptors (PTP); among these, 10 were in CC and 5 were in oocytes. PTPs are known signaling molecules regulating many cellular processes, including cell growth, differentiation, and mitotic cycle.

Nuclear hormone receptors were also present in oocytes and CC. Notably, estrogen receptor was expressed by oocytes and the estrogen receptor-binding protein was expressed by CC. Likewise, thyroid hormone receptor was expressed by the oocytes and its interacting proteins were expressed by CC. Differential expression of estrogen and thyroid hormone receptors may be a key signaling in oocyte development. Other nuclear receptors, such as peroxisome proliferators-activated receptors (PPARs), retinoic acid receptors (RXRs), and aryl hydrocarbon receptor nuclear translocators were also identified (Table 1). PPARs were identified only in CC, whereas RXRs and aryl hydrocarbon receptor were identified in both cell types. PPARs form heterodimers with retinoid X receptors (RXRs) and these heterodimers regulate transcription of various target genes, such as retinoic acid (RA)-responsive genes (BTBD11, calmin, cyclin M2, ephrin B2, HOXD10, NEDD9, RAINB6, and tenascin R; James *et al.* 2003). RAs are absolutely essential for ovarian steroid production, oocyte maturation, and early embryogenesis (Mohan *et al.* 2003).

We have identified 338 transcription factors in oocytes and CC. More transcription factors were identified in the CC (249 factors) when compared with oocytes (89 factors). However, when the total numbers of proteins are taken into account, the proportion of transcription factors was higher in oocytes (8.1%) than that of cumulus cell (5.6%). Thus, our results agree with previous data that GV oocytes are transcriptionally highly active (Memili & First 1999, Dalbies-Tran & Mermillod 2003). Furthermore, most of the transcription factors we found in both CC and oocytes belonged to the zinc finger class of transcription factors. This is reassuring as this class of transcription factors is the most common in vertebrate genomes, accounting for an estimated 3% of all gene transcription (Klug 1999). PubMed searches showed that 9 out of 19 known transcription factors were previously identified in oocytes and CC: 3 retinoid receptors and PPARs (Mohan *et al.* 2003), 4 signal transducer and activator of transcription (STAT) proteins (Boelhauve *et al.* 2005), 1 C-fos (Davis & Chen 2003), and 1 transcription activator sox 9 (Lonergan *et al.* 2003). We have identified ten transcription factors that were not identified

Table 1 Receptors and ligand pairs identified in cumulus and oocyte. This shows membrane receptors and their ligands and associated signaling molecules in cumulus and oocyte.

Oocytes			Cumulus cells	
Intracellular	Membrane	Extracellular	Membrane	Intracellular
			Predicted: similar to epidermal growth factor receptor	Predicted: similar to growth factor receptor-bound protein 2
	Predicted: similar to epidermal growth factor receptor pathway substrate		Predicted: similar to epidermal growth factor receptor pathway substrate	
			Insulin-like growth factor 2 receptor	Predicted: similar to mannose 6 phosphate receptor-binding protein 1
		Predicted: similar to IGF-II mRNA-binding protein 1		
	Predicted: similar to muscle, skeletal, receptor tyrosine kinase, part[a]	FGF2 BOVIN heparin-binding growth factor 2 precursor (HBGF-2)	Predicted: similar to platelet-derived growth factor, A chain precursor	
	Vascular endothelial growth factor-D		Vascular endothelial growth factor-D	
Predicted: similar to C1q and tumor necrosis factor-related protein 2	Predicted: similar to tumor necrosis factor receptor superfamily member		Predicted: similar to tumor necrosis factor receptor superfamily member[a]	Predicted: similar to C1q and tumor necrosis factor-related protein 2
		Predicted: similar to insulin-like growth factor IB precursor (IGF-IB)	IGF-I receptor	
		Predicted: similar to insulin-like growth factor-binding protein-like		
		Predicted: similar to interferon-α	Interferon, α; receptor	
	Predicted: similar to interleukin-1 receptor-associated kinase 1	ICBO1B interleukin-1 β precursor	Predicted: similar to interleukin 1 receptor-like 2, partial[a]	
	Predicted: similar to interleukin-1 receptor-like 1 precursor[a]	ICBO1B interleukin-1 β precursor	Predicted: similar to interleukin-1 receptor-associated kinase 1	
	Predicted: similar to interleukin-1 receptor-associated kinase 1		Predicted: similar to interleukin-1 receptor-associated kinase 1	
			MPRD_BOVIN cation-dependent mannose-6-phosphate receptor precursor[a]	Predicted: similar to mannose 6 phosphate receptor-binding protein 1
	Predicted: similar to glutamate receptor, ionotropic, *N*-methyl-D-aspar[a]		Predicted: similar to glutamate receptor, ionotropic, *N*-methyl-D-aspar[a]	
	EAA1 _BOVIN excitatory amino acid transporter 1 (sodium-dependent glutamate/asp)		Excitatory amino acid transporter 1 (sodium-dependent glutamate/asp)	
			Predicted: similar to glutamate receptor δ-2 subunit precursor[a]	Predicted: similar to glutamate receptor-interacting protein 2
			Predicted: similar to mannose receptor, C type 2, partial[a]	Predicted: similar to mannose-6-phosphate isomerase
Predicted: similar to GDP-mannose pyrophosphorylase B isoform 2			Predicted: similar to GDP-mannose pyrophosphorylase B isoform 2	
		Latent TGF-β-binding protein-2	Predicted: similar to transforming growth factor β type II receptor[a]	
		Predicted: similar to bone morphogenetic protein 10 precursor (BMP-10)		

Table 1 *(Continued).*

Oocytes			Cumulus cells	
Intracellular	Membrane	Extracellular	Membrane	Intracellular
			Predicted: similar to transforming growth factor β 3 precursor	
			Predicted: similar to glutamate receptor 1 precursor (GluR-1)[a]	Predicted: similar to glutamate receptor-interacting protein 1
	Predicted: similar to activin A type IB receptor isoform b precursor	Predicted: similar to bone morphogenetic protein 10 precursor		
			Predicted: similar to glutamate receptor KA2 precursor[a]	Predicted: similar to glutamate receptor-interacting protein 1
	Predicted: similar to ALK tyrosine kinase receptor precursor		Predicted: similar to ALK tyrosine kinase receptor precursor[a]	Predicted: similar to protein tyrosine kinase
	Predicted: similar to laminin receptor 1 (ribosomal protein SA)[a]	Laminin Bl	Predicted: similar to laminin receptor 1 (ribosomal protein SA)[a]	
		Predicted: similar to laminin α-2 chain precursor		
		Predicted: similar to laminin α 5		
		Predicted: similar to laminin β 2-like chain		
		Predicted: similar to laminin α 3 subunit isoform 1		
		Predicted: similar to laminin α 3 subunit isoform 1		
			Predicted: similar to phospholipase A2 receptor[a]	
Predicted: similar to phospholipase A2, activating protein				Predicted: similar to phospholipaseA2-like
				Predicted: similar to cytosolic phospholipase A2
	Predicted: similar to glutamate receptor KA1 precursor, partial[a]		Predicted: similar to glutamate receptor KA1 precursor[a]	
	Excitatory amino acid transporter 1 (sodium-dependent glutamate/asp)		Excitatory amino acid transporter 1 (sodium-dependent glutamate/asp)	
			Predicted: similar to glutamate receptor, ionotropic, δ 1, D14[a]	Predicted: similar to glutamate receptor interacting protein 2
	Predicted: similar to glutamate receptor, ionotropic, kainate 1 isoform		Predicted: similar to glutamate receptor, ionotropic, kainate 1 isoform[a]	Predicted: similar to glutamate receptor-interacting protein 1 (GRIP1)
			Excitatory amino acid transporter 1 (sodium-dependent glutamate/asp)	
			Predicted: similar to glutamate receptor, metabotropic 8, partial[a]	Predicted: similar to glutamate receptor-interacting protein 2
			Predicted: similar to metabotropic glutamate receptor 5 precursor	Predicted: similar to glutamate receptor-interacting protein 2
			Predicted: similar to metabotropic glutamate receptor 4 precursor	
	Predicted: similar to kinase insert domain receptor (a type III receptor)[a]			

Table 1 *(Continued).*

Oocytes			Cumulus cells	
Intracellular	Membrane	Extracellular	Membrane	Intracellular
		Vascular endothelial growth factor-D		
		Predicted: similar to apolipoprotein B-100 precursor	Predicted: similar to apolipoprotein E receptor 2, partial[a]	
		Predicted: similar to apolipoprotein B		
		Predicted: similar to neuregulin 2 isoform 3	Predicted: similar to receptor tyrosine-protein kinase erbB-3 precursor[a]	
Predicted: similar to pro-neuregulin-3 precursor (Pro-NRG3)			Predicted: similar to pro-neuregulin-3 precursor (Pro-NRG3)	
	Predicted: similar to protein tyrosine phosphatase, receptor type, N p		Predicted: similar to protein tyrosine phosphatase, receptor type, f p	
Predicted: similar to protein tyrosine phosphatase 4a1	Predicted: similar to protein tyrosine phosphatase, receptor type, U i		Predicted: similar to protein tyrosine phosphatase, receptor type, U i	Predicted: similar to protein tyrosine phosphatase 4a1
Predicted: similar to protein tyrosine phosphatase, nonreceptor type			Predicted: similar to protein tyrosine phosphatase, receptor type, Q i	Predicted: similar to protein tyrosine phosphatase, nonreceptor type
Protein tyrosine phosphatase, nonreceptor type 13 (APO-1/CD95 (Fas)-a)			Predicted: similar to protein tyrosine phosphatase, receptor type, K p	
			Predicted: similar to glycoprotein receptor gp330/ megalin precursor[a]	Predicted: similar to myomegalin
	Predicted: similar to EPH receptor A8 isoform 1 precursor		Predicted: similar to Eph receptor A2, partial[a]	
	Predicted: similar to Ephrin type-A receptor 1 precursor		Predicted: similar to EPH receptor A8 isoform 1 precursor, partial[a]	
			Predicted: similar to Ephrin type-A receptor 1 precursor[a]	
	Predicted: similar to interleukin-1 receptor-like 1 precursor[a]		Predicted: similar to Epha5 protein, partial	
	Predicted: similar to interleukin-1 receptor-associated kinase 1		Predicted: similar to interleukin-1 receptor-associated kinase 1	
			Predicted: similar to transforming growth factor β 3 precursor	
	T-cell receptor β J8		Bovine T-cell receptor γ chain	
			T-cell receptor β chain variable segment	
	T-cell receptor δ chain			
	Predicted: similar to T-cell receptor α chain			
	MHC class I heavy chain		MHC class I heavy chain	
	MHC class H β-chain		MHC class n β-chain	
		Predicted: similar to putative CCL21 chemokine	Chemokine receptor 7[a]	

Table 1 *(Continued).*

Oocytes			Cumulus cells	
Intracellular	Membrane	Extracellular	Membrane	Intracellular
			Predicted: similar to chemokine receptor	
	Predicted: similar to fibroblast growth factor receptor 1 isoform 1 precursor[a]		Predicted: similar to fibroblast growth factor receptor 1 isoform 1 precursor[a]	
		Heparin-binding growth factor 2 precursor (HBGF-2)		
		Predicted: similar to netrin 4	Predicted: similar to netrin receptor Unc5h4, partial[a]	
	Predicted: similar to netrin receptor UNC5C precursor	Predicted: similar to netrin 4	Predicted: similar to netrin receptor UNC5C precursor[a]	
	Predicted: similar to polycystic kidney disease and receptor for egg jelly[a]		Predicted: similar to polycystic kidney disease and receptor for egg jelly	
	Predicted: similar to polycystin 2, partial			
		Predicted: similar to transferrin	Predicted: similar to transferrin receptor[a]	
		TRFE_BOVIN serotransferrin precursor (transferrin)		

[a]Indicates receptor and ligand present in the same cell type.

previously in bovine oocytes and CC, and these include a forkhead transcription factor, nuclear transcription factor-Yα, Pax6, basic transcription factor 3a, zinc finger DHHC, DNA polymerase δ subunit zinc finger protein 313, zinc finger protein 470, and zinc finger protein ZFY. We have also identified 83 predicted proteins as transcription factors in oocytes and 236 predicted proteins as transcription factors in cumulus cells (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/).

'Missing' ligands

Ligands for 121 receptors were not identified, of which only 27 are proteins (Table 3). For the remaining 94, either the ligand is unknown (30 ligands) or known, but it is not a protein; axiomatically in either event the ligand cannot be identified by DDF 2-LCMS[2] (64 ligands; Table 4). Out of the 27 known protein ligands, 7 have no entries in the NCBI, which rendered them undetectable by the Sequest search. Eight of the remaining 20 have no unique peptides; 38 (of 60 peptides in total) are probably *O*-glycosylated and 2 are probably *N*-glycosylated. Therefore, only 20 unique peptides, representing 7 proteins, could theoretically be detected (Table 3).

Discussion

Although most basic reproductive biology work is done in the mouse (Eppig *et al.* 1993), significant species differences in oocyte biology exist (Sutton *et al.* 2003). Here, we used the bovine system because it is important for both agricultural and biomedical studies. Coenen *et al.* (2004) pioneered proteomics of bovine female gametogenesis. Using radio labeling and two-dimensional gel electrophoresis, they demonstrated three major patterns of translational activity during bovine oogenesis (one at the initiation of maturation, 0–4 h; one in the middle, 4–16 h; and one after completion of MI, 6–28 h) suggesting a developmentally regulated series stage-specific protein synthesis. However, the identities, functions, and expression patterns of these proteins are largely unknown. Here, we studied GV stage oocytes because these are highly active both transcriptionally and translationally (Memili & First 1999). Furthermore, interactions between the oocyte and its surrounding CC at this stage are crucial for development of a matured oocyte (MII) – the only cell type that can be fertilized to initiate a new organism. The GV stage is also one of the most active stages in the regulation of cumulus cell functions (Gilchrist *et al.* 2004). Although our methods used tenfold fewer cells to identify a ten time larger proteome, our work complements that of Coenen *et al.* (2004). Our comprehensive approach using DDF to model bovine oocytes also has significant impact on annotation of the bovine genome by demonstrating the

Table 2 Nuclear receptors identified in cumulus and oocyte. This shows nuclear receptors other than receptors related to membrane and their associated signaling molecules of cumulus and oocyte.

Oocyte			**Cumulus cells**	
Nucleus	Intracellular	**Extracellular**	Intracellular	Nucleus
Predicted: similar to estrogen-related receptor γ			Predicted: similar to estrogen receptor-binding protein Sulfotransferase, estrogen-preferring	
Predicted: similar to thyroid hormone receptor β			Predicted: similar to thyroid hormone receptor interactor 12 Predicted: similar to thyroid hormone receptor interactor 3 Predicted: similar to thyroid hormone receptor interactor 6	
	Predicted: similar to thyroid hormone receptor interactor 11		Predicted: similar to thyroid hormone receptor interactor 11 Predicted: similar to thyroid hormone receptor-associated protein	
				Predicted: similar to glucocorticoid receptor (GR) Predicted: similar to glucocorticoid receptor DNA-binding factor 1
	Heat shock 90 kDa protein 1, α		Heat shock 90 kDa protein 1, α	Predicted: similar to aryl hydrocarbon receptor nuclear translocator-1
			Cellular retinoic acid-binding protein 1	
			Cellular retinoic acid-binding protein 2	Predicted: similar to nuclear receptor ROR-β
Predicted: similar to retinoic acid receptor, α			Cellular retinoic acid-binding protein 1 Cellular retinoic acid-binding protein 2	
				Peroxisome proliferators-activated receptor α[a] Predicted: similar to peroxisome proliferator-activated receptor-binding protein
Predicted: similar to aryl hydrocarbon receptor nuclear translocator 2[a]	90 kDa heat shock protein β		90 kDa heat shock protein β	
			Cellular retinoic acid-binding protein 1	
			Cellular retinoic acid-binding protein 2	Predicted: similar to retinoic acid receptor β (RAR-β)
			Cellular retinoic acid-binding protein 1	Predicted: similar to retinoic acid receptor γ
Predicted: similar to retinoic acid receptor, α Predicted: similar to nuclear receptor coactivator 7 Predicted: similar to nuclear receptor-binding SET domain protein 1 isoform			Cellular retinoic acid-binding protein 2	Predicted: similar to retinoic acid receptor RXR-β

[a]Indicates receptor and ligand present in the same cell type.

existence of 5360 'predicted' and 38 'hypothetical' proteins for the first time (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/).

Not only are oocyte proteomes virtually undescribed, but there is also a general lack of knowledge of how interactions between the oocytes and surrounding CC lead to oocyte maturation. Interactions between oocytes and CC are considered essential for proper maturation or 'programming' of oocytes, which is crucial for normal fertilization and embryonic development (Buccione *et al.* 1990). CC are unique in that they are

Table 3 Cumulus and oocyte receptors whose ligands were not detected. This lists known cumulus and oocyte ligands that were not detected by mass spectrometry. Although the ligands were not detected the expression of its receptor indicates possible signaling mechanisms.

Cell type[a]	Protein name	Ligand
C	5-Hydroxytryptamine receptor 2A	5-Hydroxytryptamine
	Bradykinin receptor B2	Bradykinin
	Chemokine receptor 7	Mip-3-β chemokine|chemokine ligand 21
	Chemokine (C-X-C motif) ligand 13-like	CCR2b and CCR3 receptors
	Prolactin receptor long form	Prolactin
	Thyrotropin (TSH) receptor variant	Thyrotropin
	Predicted: similar to AXL receptor tyrosine kinase	Growth arrest-specific gene 6
	Predicted: similar to chemokine (C–C motif) receptor 9 isoform A, part	Chemokine scya25
	Predicted: similar to ephrin receptor EphB3 precursor	EphB1
	Predicted: similar to glucagon-like peptide 1 receptor precursor	Glucagon-like peptide 1
	Predicted: similar to glucagon-like peptide 2 receptor precursor	Glucagon-like peptide 2
	Predicted: similar to interleukin 20 receptor, α	Interleukin 20
	Predicted: similar to interleukin 4 receptor α	Interleukin 4 and IL13
	Predicted: similar to melatonin-related receptor	TOMM20 and TOMM40
	Predicted: similar to mineralocorticoid receptor, partial	Mineralocorticoids (mc) such as aldosterone
	Predicted: similar to opioid-binding protein/cell adhesion molecule–ligand	Opioid
	Predicted: similar to opioid growth factor receptor-like 1	Opioid growth factor
	Predicted: similar to prostaglandin D2 receptor	Prostaglandin D2
	Predicted: similar to ryanodine receptor 2	Ryanodine
	Predicted: similar to vasoactive intestinal polypeptide receptor 2	Vasoactive intestinal polypeptide
O	Predicted: similar to interleukin 3 receptor, α precursor	Interleukin-3
	Predicted: similar to NT-3 growth factor receptor precursor	Neutotropin 3
	Predicted: similar to proteinase-activated receptor 2 precursor	Trypsin and trypsin-like enzymes
	Predicted: similar to ryanodine receptor 2	Ryanodine
CO	ANPRB_BOVIN atrial natriuretic peptide receptor B precursor (ANP-B)	Atrial natriuretic peptide (anp)|brain natriuretic peptide (bnp)
	Predicted: similar to ephrin receptor EphB2 isoform 2 precursor	EphB2
	Predicted: similar to interleukin-17B receptor precursor	IL17B and IL17E

[a]C, cumulus cell; O, oocyte; CO, both cell types.

differentiated somatic cells essential for development of a competent oocyte. A comparative functional analysis of oocyte–cumulus cell biology between mouse and livestock oocytes is important to fully understand early mammalian development. For example, differences have been demonstrated in oocyte regulation of cumulus cell metabolism, and in cumulus cell expansion between mouse and bovine (Zuelke & Brackett 1992, Eppig *et al.* 1993, Sutton *et al.* 2003). Our work provides the first detailed definition of both CC and oocytes at the same time in development.

We used both physical and enzymatic separations to isolate pure cell populations (Memili & First 1999). We expected many proteins to be common to both CC and oocytes, particularly heat shock proteins, histones, ribosomal proteins, mitochondrial proteins, and proteins related to basic ubiquitous cellular and molecular functions (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). We detected peroxiredoxin 4 in the oocytes (Table 5, supporting data). Also detected in pig oocytes, peroxiredoxin proteins have important roles in the maintenance of intracellular redox balance and protection of cells against oxidative stress due to reactive oxygen radicals (Ellederova *et al.* 2004). This suggests a conserved mammalian mechanism for cellular protection against oxidative stress. Our previous work and studies by others demonstrated that bovine oocytes have high transcriptional activity early on during GV leading to the MII stage in which mRNAs and proteins constitute a reservoir of molecular support for early embryogenesis following fertilization (Memili & First 1999, Dalbies-Tran & Mermillod 2003, Vallee *et al.* 2005). However, proteins are the primary functional units of the genome. Thus, we initiated the foundation for comprehensive proteome modeling of the dynamics of oocyte development through cell–cell interactions with the oocyte and the CC at the GV stage.

Mainly driven by the paracrine growth factors secreted by the oocyte, bidirectional interactions between the oocytes and the CC are essential for the development of competent MII oocytes, to support early embryogenesis, and for developmental potential of embryos for fetal development (Gilchrist *et al.* 2003). We detected expected proteins, including growth factors along with their binding proteins, such as Igfs and TGF in CC and oocytes respectively (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). We detected other expected proteins in the oocyte included zona pellucida proteins, many zinc finger proteins consistent with a high level of transcriptional activity, and heat shock proteins (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). The expected cumulus cell proteins included prohormone convertase, Igf2r, and binding proteins. Although oocytes have gamete and

Table 4 Cumulus and oocyte receptors with unknown and nonprotein ligands. Receptors whose ligands were not detected are listed.

Cell type[a]	Protein name	Ligand
C	MPRD_BOVIN cation-dependent mannose-6-phosphate receptor precursor	Mannose 6 phosphate
	Leptin receptor long form	Unknown
	Transient receptor potential cation channel TRPC4 middle region 1	Unknown
	Toll-like receptor 2	Lipopolysaccharide
	Predicted: similar to candidate taste receptor T1R2, partial	Unknown
	Predicted: similar to c-kit receptor	Unknown
	Predicted: similar to G protein-coupled receptor	Unknown
	Predicted: similar to G protein-coupled receptor 103	Unknown
	Predicted: similar to G protein-coupled receptor 149	Unknown
	Predicted: similar to G protein-coupled receptor 45	Unknown
	Predicted: similar to G protein-coupled receptor 82	Unknown
	Predicted: similar to G protein-coupled receptor 88	Unknown
	Predicted: similar to G protein-coupled receptor family C, group 5	Unknown
	Predicted: similar to γ-aminobutyric acid (GABA) B receptor 1	GABA
	Predicted: similar to γ-aminobutyric acid type B receptor, subunit	GABA
	Predicted: similar to γ-aminobutyric acid type B receptor, subunit	GABA
	Predicted: similar to γ-aminobutyric-acid receptor β-2 subunit	GABA
	Predicted: similar to G protein-coupled receptor SALPR	Unknown
	Predicted: similar to hypocretin receptor 2	Unknown
	Predicted: similar to killer cell immunoglobulin-like receptor KIR3DL1	Unknown
	Predicted: similar to leukemia inhibitory factor receptor precursor	Unknown
	Predicted: similar to muscarinic acetylcholine receptor M5	Acetyl choline
	Predicted: similar to neuronal acetylcholine receptor protein, α-6	Acetyl choline
	Predicted: similar to neuronal acetylcholine receptor protein, β-2	Acetyl choline
	Predicted: similar to neuronal acetylcholine receptor protein, β-3	Acetyl choline
	Predicted: similar to nuclear receptor subfamily 2, group E, member 1	Unknown
	Predicted: similar to nuclear receptor subfamily 4, group A, member 2	Unknown
	Predicted: similar to olfactory receptor	Oderants
	Predicted: similar to olfactory receptor 10A3 (HTPCRX12)	Oderants
	Predicted: similar to olfactory receptor 1257	Oderants
	Predicted: similar to olfactory receptor 12D2 (Hs6M1-20)	Oderants
	Predicted: similar to olfactory receptor 2C3	Oderants
	Predicted: similar to olfactory receptor 5H2	Oderants
	Predicted: similar to olfactory receptor 5U1 (Hs6M1-28)	Oderants
	Predicted: similar to olfactory receptor 6M1, partial	Oderants
	Predicted: similar to olfactory receptor MOR107-1, partial	Oderants
	Predicted: similar to olfactory receptor MOR14-2, partial	Oderants
	Predicted: similar to olfactory receptor MOR156-5, partial	Oderants
	Predicted: similar to olfactory receptor MOR157-1	Oderants
	Predicted: similar to olfactory receptor MOR235-2	Oderants
	Predicted: similar to olfactory receptor MOR241-1	Oderants
	Predicted: similar to olfactory receptor MOR256-13	Oderants
	Predicted: similar to olfactory receptor MOR258-6	Oderants
	Predicted: similar to olfactory receptor MOR264-5	Oderants
	Predicted: similar to olfactory receptor MOR267-8	Oderants
	Predicted: similar to olfactory receptor MOR34-1	Oderants
	Predicted: similar to olfactory receptor Olfr366	Oderants
	Predicted: similar to olfactory receptor Olr105	Oderants
	Predicted: similar to olfactory receptor Olr1466	Oderants
	Predicted: similar to olfactory receptor Olr1537	Oderants
	Predicted: similar to olfactory receptor Olr245	Oderants
	Predicted: similar to olfactory receptor Olr315	Oderants
	Predicted: similar to olfactory receptor Olr374	Oderants
	Predicted: similar to olfactory receptor Olr39	Oderants
	Predicted: similar to olfactory receptor Olr4	Oderants
	Predicted: similar to olfactory receptor Olr641	Oderants
	Predicted: similar to olfactory receptor Olr659	Oderants
	Predicted: similar to olfactory receptor Olr879	Oderants
	Predicted: similar to olfactory receptor, family 10, subfamily X	Oderants
	Predicted: similar to olfactory receptor, family 2, subfamily M	Oderants
	Predicted: similar to olfactory receptor, family 2, subfamily T	Oderants
	Predicted: similar to olfactory receptor, family 9, subfamily Q	Oderants
	Predicted: similar to orphan nuclear receptor NR4A1	Unknown
	Predicted: similar to short transient receptor potential channel 7	Unknown
	Predicted: similar to toll-like receptor 7 precursor (UNQ248/PRO285)	Lipopolysaccharide

Table 4 *(Continued).*

Cell type[a]	Protein name	Ligand
O	Toll-like receptor 9	Lipopolysaccharide
	Predicted: similar to feline leukemia virus subgroup C receptor-related	Unknown
	Predicted: similar to G protein-coupled receptor 128	Unknown
	Predicted: similar to G protein-coupled receptor 154	Unknown
	Predicted: similar to killer cell immunoglobulin-like receptor KIR3DL1	Unknown
	Predicted: similar to nuclear receptor subfamily 5, group A	Unknown
	Predicted: similar to olfactory receptor 1J4 (HTPCRX01)	Oderants
	Predicted: similar to olfactory receptor 5T2	Oderants
	Predicted: similar to olfactory receptor 5T2	Oderants
	Predicted: similar to olfactory receptor MOR173-1	Oderants
	Predicted: similar to olfactory receptor MOR195-1	Oderants
	Predicted: similar to olfactory receptor MOR234-3	Oderants
	Predicted: similar to olfactory receptor MOR256-15	Oderants
	Predicted: similar to olfactory receptor Olfr1357	Oderants
	Predicted: similar to olfactory receptor Olr245	Oderants
	Predicted: similar to olfactory receptor Olr701	Oderants
	Predicted: similar to olfactory receptor, family 51, subfamily E	Oderants
	Predicted: similar to scavenger receptor class A, member 3 isoform 2	Unknown
	Predicted: similar to sphingosine 1-phosphate receptor Edg-5	Lysosphingolipid sphingosine 1-phosphate
	Predicted: similar to transient receptor potential cation channel	Unknown
CO	Hyaluronic acid-mediated motility receptor	Hyaluronic acid
	Predicted: similar to G protein-coupled receptor 133, partial	Unknown
	Predicted: similar to G protein-coupled receptor 171	Unknown
	Predicted: similar to G protein-coupled receptor 64, partial	Unknown
	Predicted: similar to inositol 1,4,5-trisphosphate receptor type 1	Inositol 1,4,5-trisphosphate
	Predicted: similar to inositol 1,4,5-trisphosphate receptor type 2	Inositol 1,4,5-trisphosphate
	Predicted: similar to olfactory receptor 5B2 (OST073)	Oderants
	Predicted: similar to olfactory receptor 7A5	Oderants
	Predicted: similar to transient receptor potential cation channel	Unknown

[a]C, cumulus cells; O, oocytes; CO, both cell types.

totipotency-related proteins but CC are differentiated, we detected many more unique proteins in CC than oocytes (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). Another reason for this discrepancy may be the relative lack of previous research on CC. A PubMed search shows that there are 36 times more papers describing research on oocytes than CC, which is probably because the oocyte is the unique progenitor for life. However, CC are essential to oocyte development, and for reproductive biology and are as important as oocytes (Sugiura & Eppig 2005). Our model is that oocytes orchestrate their environmental conditions by signaling cumulus cell development and physiology and that the soluble and membrane-bound signals from CC support oocyte development. This is because oocytes are dependent on CC in metabolic processes, such as glycolysis and amino acid uptake (Buccione *et al.* 1990). Here, we have been able to reconstruct signaling pathways from the intracellular space and cell membranes to the nucleus.

Paracrine growth factors secreted by oocytes are involved in a number of developmentally important events, including expansion of cumulus cell numbers and functions, regulation of follicular cell functions, and regulation of ovulatory and post-ovulatory events (Gilchrist *et al.* 2001). Among the expected growth factors, receptors, and ligands found in CC and oocytes (Table 1), there were remarkable numbers of nuclear receptors and binding proteins, for example, the RXRs in oocytes, and cellular RA-binding proteins in the CC (Table 1). Our evidence of retinoid signaling is consistent with the existing literature (30). RA, which is a metabolite of vitamin A, plays important roles in growth and differentiation by changing expression of certain genes (Mangelsdorf *et al.* 1994). RA improves development of bovine preimplantation embryos *in vitro* (Livingston *et al.* 2004) and supplementation of 9-*cis* RA in oocyte maturation medium influences trophectoderm differentiation and total cell number of the inner cell mass (Hidalgo *et al.* 2003).

Surrounding the oocyte and is made of three glycoproteins, zona pellucida has a role in fertilization and cleavage. We did not apply special treatment to the zona pellucida but we know that we could solubilize it because we identified proteins ZP2, ZP3, and ZP4 in DDF3 fraction (Supplementary Table 5, which can be viewed online at www.reproduction-online.org/supplemental/). However, the ZP has few known proteins (ZP1, 2, 3, and 4) and we may have identified previously unidentified ZP proteins but, because we did not specifically focus on the ZP, we cannot definitively identify these proteins' locations to the ZP. Notably we did not detect ZP1. This could be because ZP1 protein has no entry in the database we have used for sequest searchers which render them undetectable.

In conclusion, we have established a method that provides a basis for the proteomics of bovine oocyte

and surrounding cumulus cell biology, which will allow modeling the complex cell–cell interactions in oocyte development. This complements transcription analyses, and together the two methods may be used in the future for systems biology modeling of early mammalian development. We have also established the foundations necessary for further structural and functional annotation of the bovine genome aimed at identifying markers for developmental competency that are essential for selecting oocytes for mammalian reproduction.

Acknowledgements

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work. This study was funded by the Life Sciences and Biotechnology Institute, Mississippi Agriculture, Forestry and Experiment Station, Mississippi State University (Manuscript number: J_10983).

References

Albertini DF, Combelles CM, Benecchi E & Carabatsos MJ 2001 Cellular basis for paracrine regulation of ovarian follicle development. *Reproduction* **121** 647–653.

Bark SJ, Muster N, Yates JR III & Siuzdak G 2001 High temperature protein mass mapping using a thermophilic protease. *Journal of the American Chemical Society* **123** 1774–1775.

Boelhauve M, Sinowatz F, Wolf E & Paula-Lopes FF 2005 Maturation of bovine oocytes in the presence of leptin improves development and reduces apoptosis of *in vitro*-produced blastocysts. *Biology of Reproduction* **73** 737–744.

Buccione R, Schroeder AC & Eppig JJ 1990 Interactions between somatic cells and germ cells thought mammalian oogenesis. *Biology of Reproduction* **43** 543–547.

Coenen K, Massicotte L & Sirard MA 2004 Study of newly synthesized proteins during bovine oocyte maturation *in vitro* using image analyses of two-dimensional gel electrophoresis. *Molecular Reproduction and Development* **67** 313–322.

Coskun S, Uzumcu M, Lin YC, Friedman CI & Alak BM 1995 Regulation of cumulus cell stereidogenesis by the porcine oocyte and preliminary characterization of oocyte-produced factor(s). *Biology of Reproduction* **53** 670–675.

Dalbies-Tran R & Mermillod P 2003 Use of heterologous complementary DNA array screening to analyze bovine oocyte transcriptome and its evolution during maturation. *Biology of Reproduction* **68** 252–261.

Davis JS & Chen DB 2003 Epidermal growth factor induces c-fos and c-jun mRNA via Raf-1/MEK1/ERK-dependent and -independent pathways in bovine luteal cells. *Molecular and Cellular Endocrinology* **200** 141–154.

Ellederova Z, Halada P, Man P, Kubelka M, Motlik J & Kovarova H 2004 Protein patterns of pig oocytes during *in vitro* maturation. *Biology of Reproduction* **71** 1533–1539.

Eppig JJ, Wigglesworth K & Chesnel F 1993 Secretion of cumulus expansion enabling factor by mouse oocytes; relationship to oocyte growth and competence to resume meiosis. *Developmental Biology* **158** 400–409.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A 2005 Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook, ch 2*, pp 571–608. Ed. John M Walker. Totowa, NJ: Humana Press.

Gilchrist RB, Ritter LJ & Armstrong DT 2001 Mouse oocyte mitogenic activity is developmentally coordinated throughout folliculogenesis and meiotic maturation. *Developmental Biology* **240** 289–298.

Gilchrist RB, Morissey MP, Ritter LJ & Armstrong DT 2003 Comparison of oocyte factors and transforming growth factor-beta in the regulation of DNA synthesis in bovine granulosa cells. *Molecular and Cellular Endocrinology* **201** 87–95.

Gilchrist RB, Ritter LJ & Armstrong DT 2004 Oocyte-somatic cell interactions during follicle development in mammals. *Animal Reproduction Science* **82–83** 431–446.

Guo Y, Martinez-Williams C, Yellowley CE, Donahue HJ & Rannels DE 2001 Connexin expression by alveolar epithelial cells is regulated by extracellular matrix. *American Journal of Physiology. Lung Cellular and Molecular Physiology* **280** 191–202.

Gupta R & Brunak S 2002 Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing* 310–322.

Hidalgo CO, Diéz C, Duque P, Facal N & Gómez E 2003 Pregnancies and improved early embryonic development with bovine oocytes matured *in vitro* with 9-*cis*-retinoic acid. *Reproduction* **125** 409–416.

James SY, Lin F, Kolluri SK, Dawson MI & Zhang XK 2003 Regulation of retinoic acid receptor beta expression by peroxisome proliferator-activated receptor gamma ligands in cancer cells. *Cancer Research* **63** 3531–3538.

Julenius K, Molgaard A, Gupta R & Brunak S 2005 Prediction, conservation analysis, and structural characterization of mammalian mucin-type *O*-glycosylation sites. *Glycobiology* **15** 153–164.

Khatir H, Lonergan P & Mermillod P 1998 Kinetics of nuclear maturation and protein profiles of oocytes from prepubertal and adult cattle during *in vitro* maturation. *Theriogenology* **50** 917–929.

Klug A 1999 Zinc finger peptides for the regulation of gene expression. *Journal of Molecular Biology* **293** 215–218.

De La Fuenta R & Eppig JJ 2001 Transcriptional activity of the mouse oocyte genome: companion granulosa cells modulate transcription and chromatin remodeling. *Developmental Biology* **229** 224–236.

Liebel U, Kindler B & Pepperkok R 2004 'Harvester': a fast meta search engine of human protein resources. *Bioinformatics* **20** 1962–1963.

Livingston T, Eberhardt D, Edwards JL & Godkin J 2004 Retinol improves bovine embryonic development *in vitro*. *Reproductive Biology and Endocrinology* **2** 83.

Lonergan P, Rizos D, Gutierrez-Adan A, Moreira PM, Pintado B, de la Fuente J & Boland MP 2003 Temporal divergence in the pattern of messenger RNA expression in bovine embryos cultured from the zygote to blastocyst stage *in vitro* or *in vivo*. *Biology of Reproduction* **69** 1424–1431.

Lubec G, Afjehi-Sadat L, Yang JW & John JP 2005 Searching for hypothetical proteins: theory and practice based upon original data and literature. *Progress in Neurobiology* **77** 90–127.

Mangelsdorf DJ, Umesono K & Evans RM 1994 The retinoid receptors. In *The Retinoids: Biology, Chemistry, and Medicine*, 2 edn, pp 319–350. Eds MB Sporn, AB Roberts & DJ Goodman. New York: Raven Press, Ltd.

Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, Gevaert K, Vandekerckhove J & Apweiler R 2005 PRIDE: the proteomics identifications database. *Proteomics* **5** 3537–3545. Erratum in: Proteomics 2005; **5** 4046.

Matzuk MM, Burns KH, Viveiros MM & Eppig JJ 2002 Intercellular communication in the mammalian ovary: oocytes carry the conversation. *Science* **296** 2178–2180.

McCarthy FM, Burgess SC, van den Berg BH, Koter MD & Pharr GT 2005 Differential detergent fractionation for non-electrophoretic eukaryote cell proteomics. *Journal of Proteome Research* **4** 316–324.

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Burrell DG, Hill DP, Dolan ME, Williams WP *et al.* 2006*a* AgBase: a functional genomics resource for agriculture. *BMC Genomics* (In press).

McCarthy FM, Cooksey AM, Wang N, Bridges SM, Pharr GT & Burgess SC 2006*b* Modeling a whole organ using proteomics: the avian bursa of fabricius. *Proteomics* **6** 2759–2771.

Memili E & First NL 1999 Control of gene expression at the onset of bovine embryonic development. *Biology of Reproduction* **61** 1198–1207.

Mohan M, Thirumalapura NR & Malayer J 2003 Bovine cumulus-granulosa cells contain biologically active retinoid receptors that can respond to retinoic acid. *Reproductive Biology and Endocrinology* **1** 104.

Moor RM, Mattioli M, Ding J & Nagai T 1990 Maturation in pig oocytes *in vivo* and *in vitro*. *Journal of Reproduction and Fertility* **40** 197–210.

Pan H, O'brien MJ, Wigglesworth K, Eppig JJ & Schultz RM 2005 Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development *in vitro*. *Developmental Biology* **286** 493–506.

Sagirkaya H, Misirlioglu M, Kaya A, First NL, Parrish JJ & Memili E 2006 Developmental and molecular correlates of bovine preimplantation embryos. *Reproduction* **131** 895–904.

Senbon S, Hirao Y & Miyano T 2003 Interactions between the oocytes and the surrounding cumulus cells in the follicular development: lessons from *in vitro* culture. *Journal of Reproduction and Development* **49** 259–269.

Simon AM, Goodenough DA, Li E & Paul DL 1997 Female infertility in mice lacking connexin 37. *Nature* **385** 525–529.

Stevens TJ & Arkin IT 2000 Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **39** 417–420.

Sugiura K & Eppig JJ 2005 Society for reproductive biology founders' lecture; control of metabolic cooperativity between oocytes and their companion granulose cells by mouse oocytes. *Reproduction, Fertility, and Development* **17** 667–674.

Sutton ML, Cetica PD, Beconi MT, Kind KL, Gilchrist RB & Thompson JG 2003 Influence of oocyte-secreted factors and culture duration on metabolic activity of bovine cumulus cell complexes. *Reproduction* **126** 127–134.

Vallee M, Gravel C, Palin MF, Reghenas H, Stothard P, Wishart DS & Sirard MA 2005 Identification of novel and known oocyte-specific genes using complementary DNA subtraction and microarrays analyses in three different species. *Biology of Reproduction* **73** 61–71.

Wallin E & von Heijne G 1998 Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7** 1029–1038.

Washburn MP, Wolters D & Yates JR III 2001 Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19** 242–247.

Zuelke KA & Brackett BG 1992 Effects of luteinizing hormone on glucose metabolism in cumulus-enclosed bovine oocytes matured *in vitro*. *Endocrinology* **131** 2690–2696.

Received 14 August 2006
First decision 10 October 2006
Revised manuscript received 20 February 2007
Accepted 16 March 2007

CHAPTER III

PROTEOMICS-BASED SYSTEMS BIOLOGY MODELING OF BOVINE GERMINAL VESICLE STAGE OOCYTE AND CUMULUS CELL INTERACTION[1]

[1] Reprint from Peddinti D, Memili E, Burgess SC (2010) Proteomics-Based Systems Biology Modeling of Bovine Germinal Vesicle Stage Oocyte and Cumulus Cell Interaction. PLoS ONE 5(6): e11240. doi:10.1371/journal.pone.0011240

OPEN ACCESS Freely available online

PLoS one

Proteomics-Based Systems Biology Modeling of Bovine Germinal Vesicle Stage Oocyte and Cumulus Cell Interaction

Divyaswetha Peddinti[1,3]*, Erdogan Memili[2,3], Shane C. Burgess[1,3,4,5]

1 Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, Mississippi, United States of America, **2** Department of Animal and Dairy Sciences, Mississippi State University, Mississippi State, Mississippi, United States of America, **3** Institute for Digital Biology, Mississippi State University, Mississippi State, Mississippi, United States of America, **4** Life Sciences and Biotechnology Institute, Mississippi State University, Mississippi State, Mississippi, United States of America, **5** Mississippi Agricultural and Forestry Experimental Station, Mississippi State University, Mississippi State, Mississippi, United States of America

Abstract

Background: Oocytes are the female gametes which establish the program of life after fertilization. Interactions between oocyte and the surrounding cumulus cells at germinal vesicle (GV) stage are considered essential for proper maturation or 'programming' of oocytes, which is crucial for normal fertilization and embryonic development. However, despite its importance, little is known about the molecular events and pathways involved in this bidirectional communication.

Methodology/Principal Findings: We used differential detergent fractionation multidimensional protein identification technology (DDF-Mud PIT) on bovine GV oocyte and cumulus cells and identified 811 and 1247 proteins in GV oocyte and cumulus cells, respectively; 371 proteins were significantly differentially expressed between each cell type. Systems biology modeling, which included Gene Ontology (GO) and canonical genetic pathway analysis, showed that cumulus cells have higher expression of proteins involved in cell communication, generation of precursor metabolites and energy, as well as transport than GV oocytes. Our data also suggests a hypothesis that oocytes may depend on the presence of cumulus cells to generate specific cellular signals to coordinate their growth and maturation.

Conclusions/Significance: Systems biology modeling of bovine oocytes and cumulus cells in the context of GO and protein interaction networks identified the signaling pathways associated with the proteins involved in cell-to-cell signaling biological process that may have implications in oocyte competence and maturation. This first comprehensive systems biology modeling of bovine oocytes and cumulus cell proteomes not only provides a foundation for signaling and cell physiology at the GV stage of oocyte development, but are also valuable for comparative studies of other stages of oocyte development at the molecular level.

Citation: Peddinti D, Memili E, Burgess SC (2010) Proteomics-Based Systems Biology Modeling of Bovine Germinal Vesicle Stage Oocyte and Cumulus Cell Interaction. PLoS ONE 5(6): e11240. doi:10.1371/journal.pone.0011240

Editor: Jay M. Baltz, Ottawa Hospital Research Institute and University of Ottawa, Canada

Received January 27, 2010; **Accepted** May 28, 2010; **Published** June 21, 2010

Funding: This study was funded by the Life Sciences and Biology Institute and Mississippi Agricultural and Forestry Experiment Station, Mississippi State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dpeddinti@cvm.msstate.edu

Introduction

Germinal vesicle (GV) breakdown is fundamental for maturation of fully grown, developmentally competent mammalian oocytes. Intercellular communication between oocytes and cumulus cells at GV stage is essential for proper maturation or 'programming' of oocytes, which is crucial for fertilization and embryonic development [1,2]. 'Gap junctions' in the regions of oocyte and cumulus cells association allow nutrient and paracrine factor transport between oocytes and cumulus cells [2,3,4]. Cumulus cell removal before maturation, or the obstruction of gap junctions, suppresses oocyte maturation [5,6,7,8,9]. Furthermore, cumulus cells are proposed to protect oocytes by preventing oxidative stress-induced cell death and DNA damage by increasing oocyte glutathione content [10] and thus functionally influence oocyte competence. In turn, via secreted factors, oocytes regulate folliculogenesis by promoting: granulosa cell proliferation, differentiation, and gene expression as well as cumulus cell expansion [2,11]. Folliculogenesis fails in the absence of oocyte paracrine signaling, (whether due to genetic deficiency or experimental oocyte ablation) [2,12,13]. Although this oocyte and cumulus cell bidirectional communication is essential for competent oocyte development, the molecular details underlying this communication remain poorly defined. There is thus still a lack of reliable molecular markers and valid definition of a high quality oocyte have impeded the selection of optimal oocytes necessary for assisted reproductive techniques (ARTs) at a high efficiency in humans as well as farm animal species.

Published studies with mouse model show that cumulus cells play an important role in nutritional support of the developing oocyte in the form of pyruvate [14,15,16] and stimulation of this nutritional support of cumulus cells is in turn dependent upon the presence of paracrine factors secreted by the oocytes [17]. Although most basic reproductive biology work is done in the

mouse [18], significant species differences in oocyte biology exist between humans and mice [19,20]. The bovine is a relevant animal model for studies of oocyte and cumulus cell communication in human because oocyte biology, and many aspects of ovarian follicular dynamics, is similar between these two single ovulating species [14,15]. Bovine fertility is also important on its own merit; it has implications in agro-economics involving cattle industry worldwide. Evidences using both the bovine and porcine models show that attachment of cumulus cells to the oocyte during meiotic maturation and fertilization is critical for promoting subsequent embryo development [7,8,21,22].

Proteins primarily determine cell phenotypes and here we used a shotgun proteomics approach that allows us to relatively quantify which proteins are actually expressed in the cell compartments (as opposed to what might be or have the potential to be). This is especially important in oocytes, where there is no linear correlation between amounts of mRNA and the proteins they encode [23]. We previously analyzed the proteomes of bovine germinal vesicle (GV) stage oocytes and their surrounding cumulus cells using differential detergent fractionation two-dimensional liquid chromatography electrospray ionization tandem mass spectrometry (DDF 2D-LC ESI MS/MS) [19]. In the previous study, we reported the first descriptive map of bovine GV oocytes and their potentially-interacting cumulus cells with specific emphasis on membrane, nuclear proteins, receptor-ligand pairs, and transcription factors. Here, and in contrast, using separately-harvested bovine GV oocyte and cumulus cells, we did a separate proteomics experiment but this time using an updated bovine proteome, more stringent search criteria, and much better structural and functional annotation. This has allowed us to do more comprehensive quantitative computational systems biology modeling of oocyte and cumulus cell communication.

Biological systems utilize highly complex, interrelated networks and pathways to function and in contrast to our previous work [19], here we used two complementary computational systems biology modeling approaches: Gene Ontology (GO; [24]) –based and canonical genetic-network-based, to derive understanding from our large dataset. As part of this process, we functionally-annotated all oocyte and cumulus proteins we identified. This improved these protein's GO annotation quality score (GAQ) [25], and enhanced our ability to do GO quantitative modeling to identify the biological processes that are either agonistic or antagonistic to GV oocyte development and which may affect oocyte competence and maturation. Our data also provide a foundation for further structural and functional annotation of the bovine genome specifically-focused on identifying genes associated with developmental competency that could be used for selecting oocytes in manipulating mammalian reproduction. Complementing our biological process modeling, our canonical genetic network modeling identified signaling pathways likely to be involved in the bidirectional communication between oocytes and cumulus cells. Together our analyses can serve as a basis for further "omic" or reductionist research on oocyte and cumulus cell communication and related reproductive abnormalities.

Results

GV oocyte and cumulus cell proteomes

We identified 811 and 1,247 proteins in GV oocyte and cumulus cells, respectively (Table S1). All protein identifications and MS/MS data have been submitted to the PRoteomics IDEntifications database (PRIDE [26]; accession #: 8691, 8692, 8693, 8694, 8695 and 8696, representing the 6 new datasets generated in this work). Of the total 2,058 proteins, 352 (20.4%) were common to both cell types and 459, and 895 were unique to GV oocyte, and cumulus cells, respectively (Figure 1); 371 proteins were differentially-expressed between GV oocyte and cumulus cells (301 proteins had higher and 70 had lower expression in the cumulus cells, compared to the GV oocytes [Table S2]). Only 702 proteins (41.1%) were annotated as 'known' and their expression at protein level have been experimentally validated (23% cumulus cell specific, 6.3% GV oocyte specific, and 11.3% common to both [Figure 1]). Approximately 43% of the identified proteins were annotated as 'predicted' based on sequence homology and their expression at protein level has not previously been confirmed [27]. Our proteomics data has contributed to the bovine genome annotation by experimentally-confirming the in vivo expression of 742 electronically predicted proteins (Table S1). We also identified 6.5% and 5.7% of 'hypothetical' proteins (i.e. proteins predicted from nucleic acid sequences and that have not been shown to exist by experimental protein chemical evidence [28]) specific to GV oocyte and cumulus cells, respectively and 3.2% common to both cell types.

It is complicated to directly compare between the data set in present study with our previous study [19]. First this is because here we used more sophisticated proteomics methods. Secondly we used a bovine RefSeq database [08/28/2007; 25,078 entries] with very many different protein identification numbers (GI numbers) and protein names; previously *Bos taurus*, had no RefSeq database and so we had to derive a non-redundant protein database [NRPD; National Center for Biotechnology Institute. (NCBI); 07/20/2005; 39,963 entries] using search terms 'Bos' and 'taurus'. Together this means that any estimate of overlap based on GI or name will be an underestimate. Here we BLAST searched the protein sequences of the smaller current dataset against the larger previous dataset and found an overlap of 439 (26%) (Figure S1).

GO Modeling

Predicted and hypothetical proteins do not have any functional annotation associated with them and they represent ~60% of total proteins in our oocyte and cumulus datasets. To compensate for this lack of functional annotation, and to provide the best foundation for biological modeling, we GO-annotated all proteins in our GV oocyte and cumulus data sets. We annotated all the

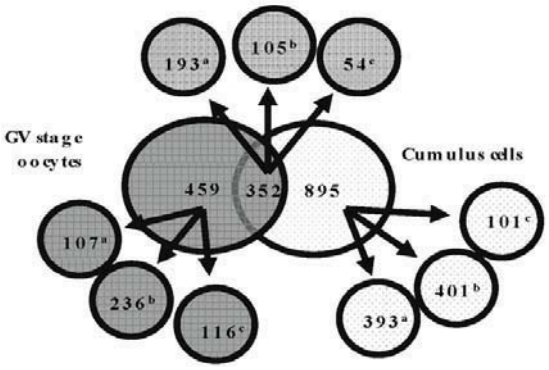

Figure 1. Comparison of proteins identified in germinal vesicle (GV) oocytes and cumulus cells. Distribution of predicted, known, and hypothetical proteins in GV oocytes and cumulus cells is shown. Superscript a, b and c = known, predicted and hypothetical proteins, respectively.
doi:10.1371/journal.pone.0011240.g001

proteins identified in GV oocyte and cumulus cells using GO and obtained annotation for 765 and 1159 proteins of GV oocyte and cumulus cell, respectively. Compared with our previous work, the GAQ score, which is measure of GO annotation quality [25] was almost doubled from 49.2 to 75 (Figure 2). This allowed us to do comprehensive modeling of bovine GV oocyte and cumulus cell communication. Grouping of biological process annotations into more generalized GO categories using a generic GO slim revealed 19 functional categories were represented in GV oocyte and cumulus cell proteomic datasets. Functional categories and percentages of proteins in each category from oocyte and cumulus are shown in Figure 3. The largest GO category represented in GV oocyte was related to cell communication with translation, transport, RNA metabolism, reproduction, and cytoskeletal organization and biogenesis also well represented (42.7% of the oocyte proteome). The largest represented GO category in cumulus cells was transport with cell communication, generation of precursor metabolites and energy, transcription, translation, and protein modification also substantially represented (40.6% of the cumulus cell proteome). Membrane and nuclear proteins are fundamental for inter- and intracellular signaling and are thus fundamental for modeling cell–cell interactions. From the cellular component GO, we identified 404 membrane proteins (24% of the total proteins): 248 unique to CC, 84 unique to oocytes, and 72 in both cell types. Using GO associations, we also identified 273 nuclear proteins: 172 unique to CC, 46 unique to oocytes, and 55 in both cell types (Table S3).

We next focused on the 371 proteins differentially expressed between GV oocyte and cumulus cells. Application of the AgBase generic GO Slim [29] revealed that 7 functional categories were represented in these differentially expressed proteins. In comparison to oocytes, cumulus cells had significantly higher expression of proteins involved in three biological processes: generation of precursor metabolites and energy, transport, and cell communication (Figure 4).

Our GO based quantitative modeling also showed that GV stage oocytes are biased towards biological processes such as cell cycle regulation, signal transduction, DNA transcription, protein metabolism and modification, generation of precursor metabolites and energy, response to oxidative stress, protein amino acid phosphorylation, and cytoskeletol organization and biogenesis but biased against apoptosis (Figure 5).

Canonical genetic network analysis

Analysis of protein-protein interactions as part of complexes, pathways and biological networks is complimentary to analysis of functional annotations and here we used canonical pathway analysis. Among the 811 proteins identified in GV oocytes, 727 proteins had information about their contribution in canonical networks and functions/pathways. We identified 30 canonical networks and 49 functions. Functions of proteins involved in the top five networks were related to protein synthesis, DNA replication, recombination and repair, cell-to-cell signaling and interaction, molecular transport, amino acid metabolism, reproductive system development and function, small molecule biochemistry, and cellular function and maintenance. Of the 1247 proteins identified in cumulus cells, 1114 had information about their contribution in canonical genetic networks and functions/pathways, respectively. We identified 46 networks and 50 functions. Functions of proteins involved in top five networks were related to cell-to-cell signaling and interaction, molecular transport, protein synthesis, nucleic acid metabolism, cellular function and maintenance, small molecule biochemistry, molecular transport, RNA trafficking and post translational modification. The top five networks (ranked based on statistical significance), and their associated proteins are shown in Table 1 and 2 for GV oocyte and cumulus cell, respectively.

Twenty-three and 39 canonical pathways were significantly represented in GV oocytes and cumulus cells, respectively. Oxidative phosphorylation is the top represented canonical pathway significant only in cumulus cells. In this pathway, expression of 21 proteins was significantly altered and all these proteins have higher expression in cumulus cells compared to GV oocyte. These proteins include 9 ATP (adenosine triphosphate) synthases (ATP5A1, ATP5B, ATP5C1, ATP5F1, ATP5H, ATP5J, ATP5L, ATP5O, and ATP6V1E1), 3 cytochrome-c-

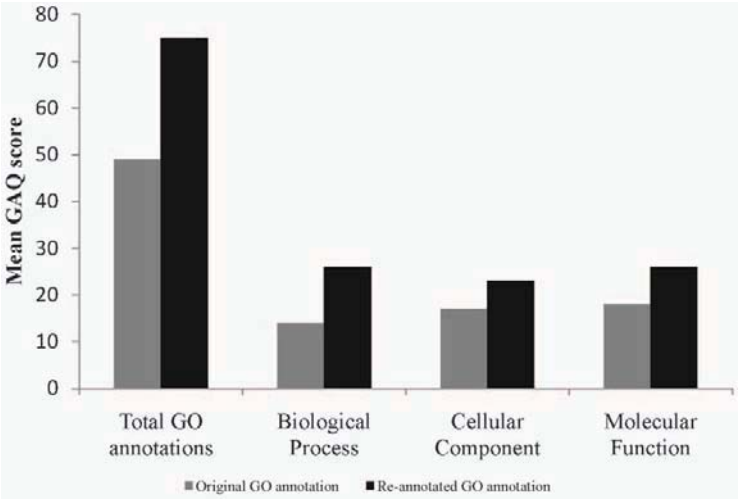

Figure 2. Mean Gene ontology Annotation Quality (GAQ) scores. Mean Gene ontology Annotation Quality (GAQ) score of original and improved Gene onltology annotations of germinal vesicle (GV) oocyte and cumulus cell proteome data sets before and after reannotation.
doi:10.1371/journal.pone.0011240.g002

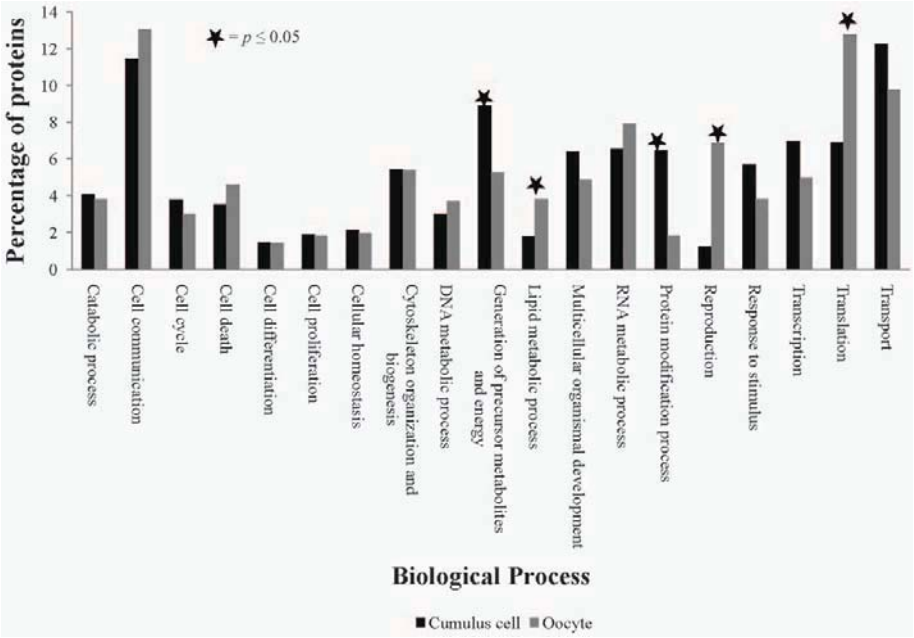

Figure 3. Gene Ontology (GO) modeling of germinal vesicle (GV) oocyte and cumulus cell proteomes. Distribution of percentages of GV stage oocyte and cumulus cell proteins involved in various biological processes. Significant differences in percentage of proteins involved in various identified GO categories in between GV oocyte and cumulus cells were evaluated by student's t-test.
doi:10.1371/journal.pone.0011240.g003

oxidases (COX2, COX17, and COX5A), cytochrome C-1(CYC1), 6 NADH (reduced nicotinamide adenine dinucleotide) dehydrogenase (ubiquinone) complexes (NDUFA5, NDUFA9, NDUFB, NDUFC2, NDUFS2, and NDUFV2) and 2 ubiquinol cytochrome c reductase core proteins (UQCRC1 and UQCRC2) (Table S2). Glycolysis and pyruvate metabolism pathways were

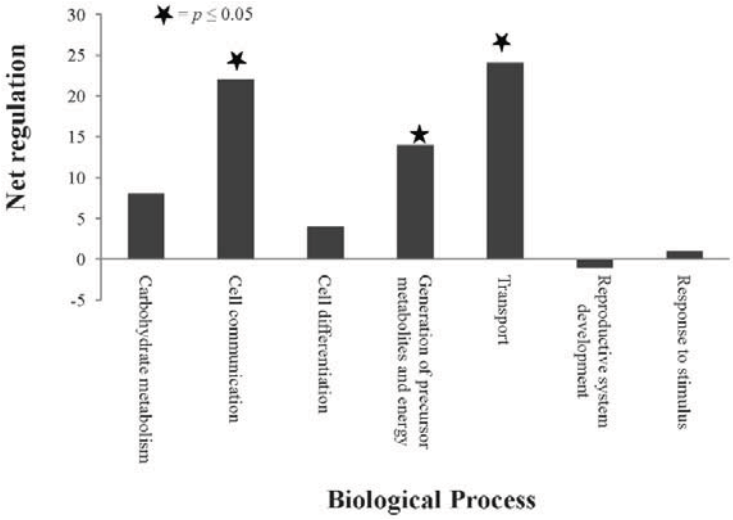

Figure 4. Overall effects in GO Slims of differentially expressed proteins of germinal vesicle (GV) oocyte and cumulus cells. Biological process GO annotations of all differentially-expressed proteins between GV oocyte and cumulus cells were used to generate GO Slims. For each GO Slim the difference in the numbers of proteins with higher expression and the number of proteins with lower expression in cumulus cells (relative to GV oocyte) was calculated to estimate the net regulatory effect.
doi:10.1371/journal.pone.0011240.g004

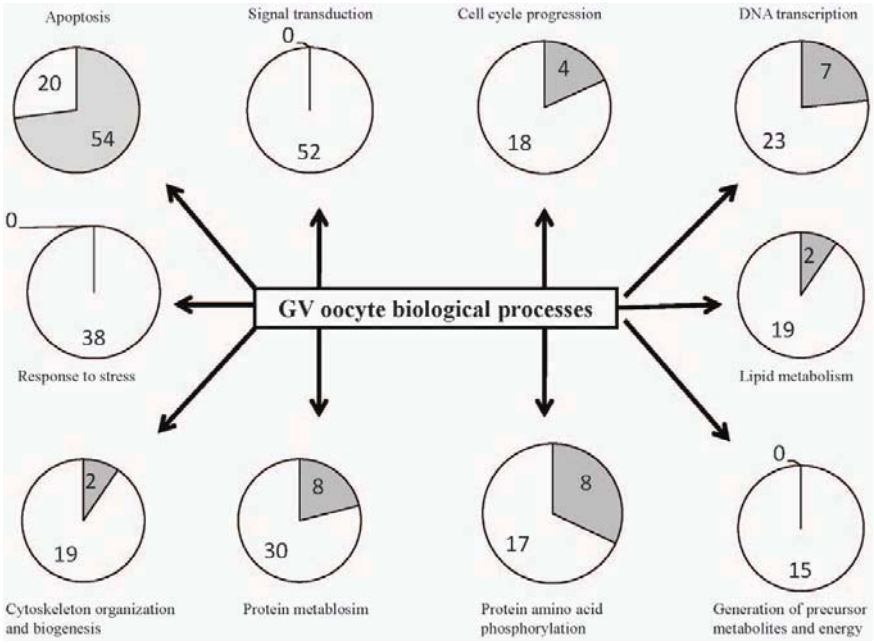

Figure 5. Germinal vesicle (GV) oocyte biological processes. Numbers of proteins agonistic (white) or antagonistic (grey) for each biological process including signal transduction, cell cycle regulation, DNA transcription, apoptosis regulation, protein metabolism and modification, generation of precursor metabolites and energy, cytoskeleton organization and biogenesis, and response to stress and calculated the net effect.
doi:10.1371/journal.pone.0011240.g005

significantly represented in both GV oocyte and cumulus cell proteomes. Expression of 12 key enzymes involved in glycolysis/pyruvate metabolism, including triosephosphate isomerase (TPIS), enolase 1 and 2 (ENO1 and ENO2), pyruvate kinase muscle (PKM), fructose-bisphosphate aldolase A (ALDOA), fructose-bisphosphate aldolase C (ALDOC), lactate dehydrogenase A (LDHA), phosphoglycerate mutase 1 (PGAM1), phosphoglycerate kinase 1 and 2(PGK1 and PGK2), glucose phosphate isomerise (GPI), and aldehyde dehydrogenase 3 family member A2 (ALDH3A2) was higher in cumulus cells compared to oocyte (Table S2).

Cell-to-cell signaling network

Bidirectional communication between oocytes and cumulus cells is essential for the development and function of both cell types [1,11,15]. We focused on the 91 identified proteins GO-annotated as involved in cell-to-cell signaling; 29 of these proteins were differentially-expressed between GV oocyte and cumulus cells (23 were greater and 6 lower in the cumulus cells compared to GV oocytes [Table S4]). The canonical network generated by IPA using proteins involved in cell-to-cell signaling is shown in Figure 6. We next overlaid canonical pathway information on to this network to identify important signaling pathways involved in the intricate cross talk between the GV oocyte and its surrounding cumulus cells. Major pathways associated with proteins involved in this cell-to-cell signaling network are related to integrin signaling, actin cytoskeleton signaling, mitogen-activated protein kinase (MAPK) signaling, phosphoinositide 3-kinase (PI3K) signaling, and ephrin receptor signaling (Table 3). Expression of six proteins– actin, beta (ACTB), actinin, alpha 4 (ACTN4), integrin alpha 2(ITGA2), integrin alpha V (ITGAV), integrin beta 1 (ITGB1), and talin 1(TLN1)—involved in integrin signaling and actin cytoskeleton signaling was higher in cumulus cells compared to GV oocyte (Figure 7) (Table S4). Expression of six proteins—zona pellucida glycoprotein 2[ZP2], zona pellucida 3[ZP3], periredoxin 2[PRDX2], complement component 3[C3], milk fat globule-EGF factor 8 protein [MFGE8] and vitronectin [VTN]—involved in cell-to-cell signaling was significantly higher in GV oocyte compared to cumulus cells (Figure 6; Table S4).

Discussion

Here we provide a foundational, proteomics-based descriptive computational systems-biology modeling of oocyte and cumulus cell interaction at the GV stage. Bidirectional communication between oocytes and cumulus cells is considered essential for proper maturation or 'programming' of oocytes, which is crucial for normal fertilization and embryonic development [15]. Cumulus cells are unique in that they are differentiated somatic cells essential for development of a competent oocyte. In turn, oocytes through the secretion of secreted factors (OSF's), regulate a multitude of key cumulus cell functions, which may in turn produce positive regulatory factors that pass back to the oocyte, improving subsequent development [17]. However, despite their importance, little is known about the molecular events and pathways involved in the bidirectional communication between oocyte and cumulus cells. The proteins identified in this study in combination with the functional modeling using GO and IPA can serve as a basis for future hypothesis-driven research on follicle development, oocyte and cumulus cell communication, oocyte maturation, and related reproductive abnormalities.

In this study, we identified 811, and 1247 proteins of oocyte, and cumulus cells respectively, which is a significant increase in

Table 1. The top five biological networks in bovine GV oocyte proteome.

Network IDs	Proteins in Network[a)]	Score[b)]	Focus Proteins	Top Functions
1	*Akt*, **ANGPTL3, C17ORF61, C7ORF20, DDB1, EEF2, EEF1A1, EEF1A2, EEF1B2, EEF1D, EEF1G, ENO1, ENO3**, *Enolase*, **FAM62A, GPI, HARS, HRNR, HYOU1, ILF2 (includes EG:3608), KRT10, PDCD5, PHIP, PLA2G1B, PPT1**, *Protein-synthesizing GTPase*, **RAD23B, RAI14, RPS7, SRGAP1, TSFM (includes EG:10102), TUBB4, TUFM, WARS, YWHAZ**	52	32	Protein synthesis, Lipid metabolism, Small- molecule biochemistry
2	**AHCY, AK2, AURKA, CACYBP, DDX6, EIF4ENIF1, EIF5A, ELAVL1, G3BP1, H2AFX, HIST3H2A, HMGB2, HNRNPK**, *Importin alpha/beta*, *Importin beta*, NME2, **NUP50, PCBP2, PCBP1 (includes EG:5093), PDIA6**, *Pkc(s)*, **RAG1, RALY, RAN, RBMX, RPL23, RPS2, RPS20, TAGLN, TERF1, TGM1, TLN1, UBE2N, UGP2, UQCRC2**	52	32	Molecular transport, Protein trafficking, Amino acid metabolism
3	**ACR**, *Adenosine-tetraphosphatase*, **ALDH2**, *ATP synthase*, **ATP5A1, ATP5B, ATP5C1, ATP5G2, ATP5O**, *ATPase*, **BAT1**, *ERK*, **ETFA (includes EG:2108), ETFB (includes EG:2109), FAU, FHIT**, *GOT*, **GOT2**, *H+-transporting two-sector ATPase*, **HSPD1, HSPE1, M6PRBP1, MAPKAPK3, MFGE8, MVP, NSF, PJA1, RAB1A, RAB2A, RAB7A, RABEP1**, *Rsk*, **STX12, ZP2, ZP3**	42	28	Cell-to-cell signaling and interaction, Reproductive system development and function, DNA replication, recombination, and repair
4	*26s Proteasome*, **AGA, COX5B, DNAJA1, DNMT1, EPHA2, FKBP4**, *HSP*, *Hsp70*, *Hsp90*, **HSP90AA1, HSP90AB1, HSPA2, HSPA5, HSPA6, HSPA8, HSPA9, HSPB1, HSPH1, IARS2, MDH1, MTHFD1**, *Nos*, **PARK7, PARP1, PFN1, PGK1**, *PI3K*, **PPA1, PSAP, PTGES3 (includes EG:10728), RPL7, RPS3A**, *Shc*, **STIP1**	*42*	28	Cellular function and maintenance, Cellular compromise
5	**AICDA, Ant, CYCS (includes EG:54205)**, *Cytochrome c oxidase*, *DDOST*, *Dolichyl-diphosphooligosaccharide-protein glycotransferase*, *Glutathione peroxidase*, **GPX1, IL1F5, INPP1**, *Ldh*, **LDHA, LDHB**, *NF-kappaB (family)*, *NFkB (complex)*, **PECAM1**, *peroxidase (miscellaneous)*, **PRDX1, PRDX2, PRDX3, PRDX4, PRDX5, PRDX6, RAB3C, RPN1, RPN2, SLC25A3, SLC25A4, SLC25A10, SLC25A13**, *Sod*, **SOD2, TRPM8, TXN, VDAC1**	38	26	Small molecule biochemistry, Molecular transport, Cellular function and maintenance

a) The focus proteins are indicated with gene names and shown in bold letters.
b) A score of >2 is considered statistically significant.
doi:10.1371/journal.pone.0011240.t001

Table 2. The top five biological networks in bovine cumulus cell proteome.

Network IDs	Proteins in Network[a)]	Score[b)]	Focus Proteins	Top Functions
1	**ACADM, ATP1A2, BCKDHA, C7ORF20, CAPNS1, CSPG4, DPYSL2**, *Enolase*, *ERK*, **ETFA (includes EG:2108), ETFB (includes EG:2109), FAU, FHIT, GAK, HSPE1, IPO7, LGI1, MAP1B, MFGE8, MPZL1, MVP, NUMA1, PELP1, POSTN, PRKCSH**, *Rab11*, **RAB11B, RAB1A, TUBA3C, TUBA4A, TUBB2C**, *Tubulin*, ULBP3, ZP2, ZP3	43	31	Cell-To-Cell Signaling and Interaction, Reproductive system development and function
2	**ACTR6, AK2, ANXA4, AP3D1, BRP44, CALU, EEF2, EEF1A1, EEF1A2, EEF1B2, EEF1D, FOS**, *Immunoproteasome Pa28/20s*, **NIPSNAP1, OPTN**, *Proteasome PA700/20s*, *Protein-synthesizing GTPase*, ***PSMA*, PSMA1, PSMA2, PSMA4, PSMA6, PSMA7, PSMB2, PSMB4, PSME1, PSME2, PTPRN, RPLP1, RPLP2, RPLP0 (includes EG:6175), RPS18, SEC23B, TKT, TUFM**	41	31	Protein synthesis, Molecular transport, Nucleic acid metabolism
3	*Adaptor protein 2*, *Ap1*, **AP1G1, BLVRA, BTF3, CD58**, *Ck2*, *Clathrin*, **CLTC, CSNK2A1, EIF5A, FKBP3, FKBP10, GTF2F1, HNRNPA2B1, HNRNPL, IGF2R, NUCKS1**, *Peptidylprolyl isomerase*, *Phosphatidylinositol4,5 kinase*, **PIN1, PIP4K2A, POLR3C, PPIB, PSMA3, RGS19, RPL5 (includes EG:6125), SSB (includes EG:6741), TF, TFR2, TFRC, TOP1, TUBB4, VPS35, WARS**	38	29	Cellular function and maintenance, Small molecule biochemistry, Molecular transport
4	*3-hydroxyacyl-CoA dehydrogenase*, **ACAT1**, *Acetyl-CoA C-acetyltransferase*, **ATIC, B4GALT1, CYLD, DAD1, DDOST**, *Dolichyl-diphosphooligosaccharide-protein glycotransferase*, **ECH1, ECHS1**, *Enoyl-CoA hydratase*, **ENPP1**, *GOT*, **GOT2, HADHA, HADHB, HSD17B4, HSD17B10**, *NFkB (complex)*, **NME2, OSTC, PDIA6, PGRMC1**, *PPARα-RXRα*, **PRDX4, RPN1, RPN2, SFXN1, SLC27A1, STK10, STT3A, TOMM20, TOMM40, TOMM70A**	36	28	Lipid metabolism, Small molecule biochemistry, Post-translational modification
5	**ALPL, ANP32B, C21ORF33, CBX2, CCND1, CCT8, CDC37, DDX3X, DHRS12, EIF3D, EIF4A2, EIF4ENIF1, ELAVL1, ENO3, H1F0, H3F3B**, *Hdac*, *Histone h3*, *Histone h4*, **HNRNPM**, *Importin beta*, **INHA, KPNB1**, *Mi2*, **NCAPD2, PNO1, RAN, SF3B1, SF3B3**, *SWI-SNF*, *Tcf/lef*, **THRAP3, TNPO1, VRK1, XPO1**	36	28	Molecular Transport, RNA Trafficking

a) The focus proteins are indicated with gene names and shown in bold letters.
b) A score of >2 is considered statistically significant.
doi:10.1371/journal.pone.0011240.t002

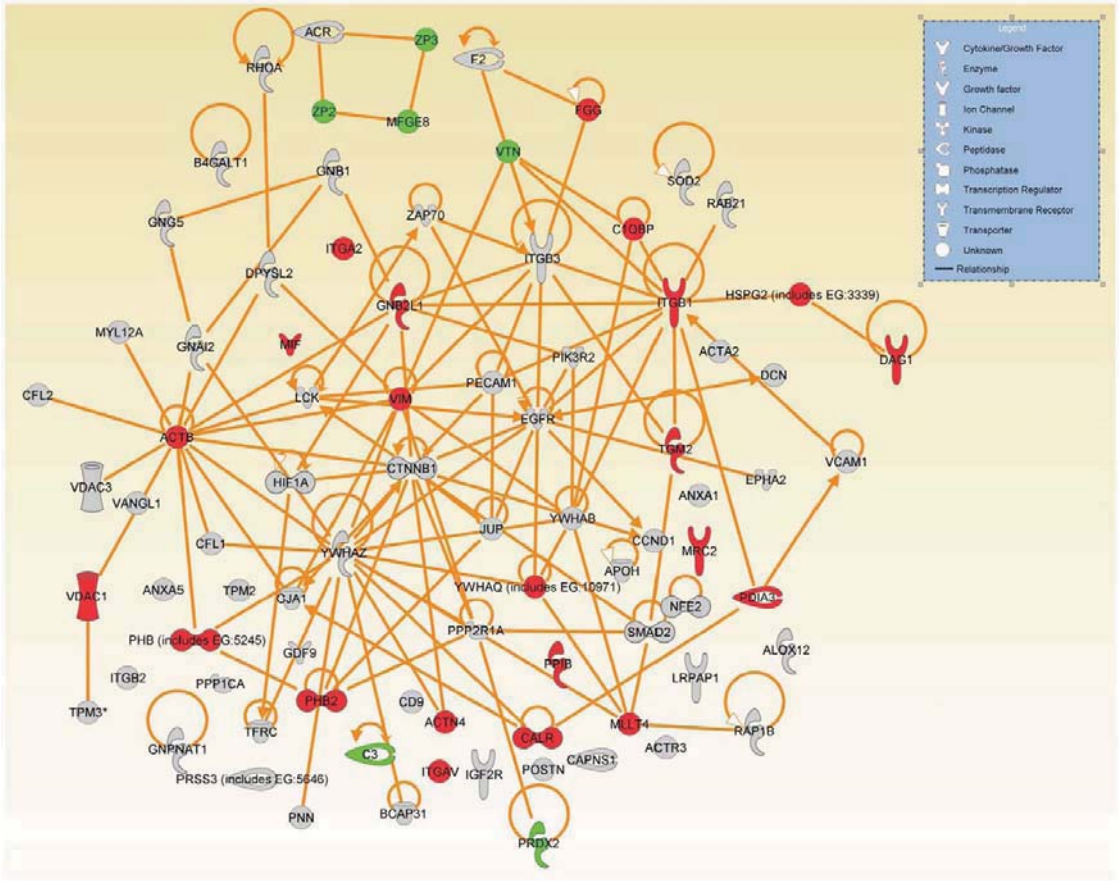

Figure 6. Cell-to-cell signaling and interaction network. Network generated with proteins involved in cell-to-cell signaling biological process using Ingenuity Pathway Analysis (IPA) as described in materials and methods. Each node represents a protein; proteins in shaded nodes were found in either GV oocyte or cumulus cell or both (see Table 3). Proteins in red and green nodes were higher and lower, respectively, in cumulus cells compared with GV oocytes.
doi:10.1371/journal.pone.0011240.g006

information compared to other female gametal proteomic studies [30,31,32,33,34,35]. Like all proteomics studies, we identified relatively small number of proteins compared to the numbers of mRNAs that can be identified in transcriptome studies. However, mRNAs are at least one step removed from phenotype; proteins are the functional molecules of cells and there is, in general, a very low correlation between amounts of mRNAs and amounts of the encoded protein [36,37,38,39]. Together this means that mRNA is a much less accurate predictor of phenotype than is protein. Even though the coverage of genetic pathways may be lower from proteomics experiments than transcriptomics experiments, because of the statistics that underlie the pathway analysis we are still able to identify relevant pathways that are important to the phenotype.

This current work is based on a separate previous proteomics study [19] which was smaller and generated less than one third of the data than this study did and did not use tissue replicates. It also uses a newer version of the bovine proteome. Though both studies follow accepted standards in the field, there are some significant technical differences between this and the previous study which exemplify the maturation of the field of proteomics. In this present study, protein identification is more stringent and we used the probabilistic approach of decoy database searching [40]. Decoy database searching is based on the real data and uses a decoy database derived from this real database; as this maintains the same amino acid composition and the search space characteristics. Not only does this method allow a probability estimation for every peptide identification is a false positive (and thus protein) but also, because it uses mass spectra generated in the experiment, it does not rely on arbitrary cut off values derived from unrelated experiments [40]. Here we have more confident protein identifications than in the previous work, though we are likely to have increased our type II error. Even though we analyzed almost three times as many mass spectra in this new study, we ended up "confidently" identifying fewer proteins compared to previous study. In the previous study, we identified 5253 proteins were expressed by the cumulus cells and 1950 proteins were expressed by GV oocytes; in this study, we identified 1247 proteins in cumulus cells, and 811 proteins in GV oocytes.

Table 3. Top five signaling pathways associated with cell to cell signaling network.

Pathway name[A]	# of focus proteins	Focus proteins[B]	P-value
Integrin signaling	15	ACTA2, ↑ACTB, ↑ACTN4, ACTR3, CAPNS1, ↑ITGA2, ITGB2, ↑ITGAV,	1.01E-12
		↑ITGB1, ITGB3, MYL12A, PIK3R2,	
		PPP1CA, RAP1B, RHOA, ↑TLN	
Actin cytoskeleton signaling	13	ACTA2, ↑ACTB, ↑ACTN4, ACTR3, CFL1	5.48E-10
		CFL2, F2, ↑ITGA2, ↑ITGB1, MYL12A,	
		PIK3R2, PPP1CA, RHOA	
Ephrin receptor signaling	12	ACTR3, CFL1, CFL2, EPHA2, GNAI2,	1.11E-09
		GNB1, ↑GNB2L1, GNG5, ↑ITGA2,	
		↑ITGB1, RAP1B, RHOA,	
PI3K signaling	9	CCND1, CTNNB1, ↑ITGA2, ↑ITGB1, PIK3R2 PPP2RIA, YWHAB,	5.44E-08
		↑YWHAQ, YWHAZ	
MAPK signaling	9	↑ITGA2, ↑ITGB1, PIK3R2, PPP2RIA, PPP1CA, RAP1B, YWHAB,	2.09E-06
		↑YWHAQ, YWHAZ	

[A]MAPK : Mitogen activated protein kinase; PI3K ; Phosphoinositide 3-kinase.
[B]↑ higher expression in cumulus cells compared to GV oocyte.
doi:10.1371/journal.pone.0011240.t003

Maternal effect proteins are the proteins that are present in the oocyte, remain expressed throughout oocyte maturation and early embryonic development until maternal-to-embryonic transition (MET) and they may also be required for successful activation of the embryonic genome. We identified 11 maternal effect proteins in bovine oocytes: Maternal Antigen that Embryos Require (Mater; a.k.a NACHT, leucine rich repeat and PYD containing 5, *Nalp5*,), zona pellucid proteins-2, 3 and 4 (ZP-2, -3, and -4), growth differentiation factor 9 (GDF 9), beta-actin (ACTB), heat shock protein-70 (HSP-70), peroxiredoxins 1 and 2 (PRDX-1, -2), DNA (cytosine-5) methyl transferase one (DNMT1), and fibroblast growth factor-8 (FGF-8) all of which are important to oogensis and early embryonic development [41,42,43,44,45,46].

Our previous publication described an initial GO-based functional analysis of bovine GV oocyte and cumulus cells using 'known proteins' and focusing on the cellular component Gene

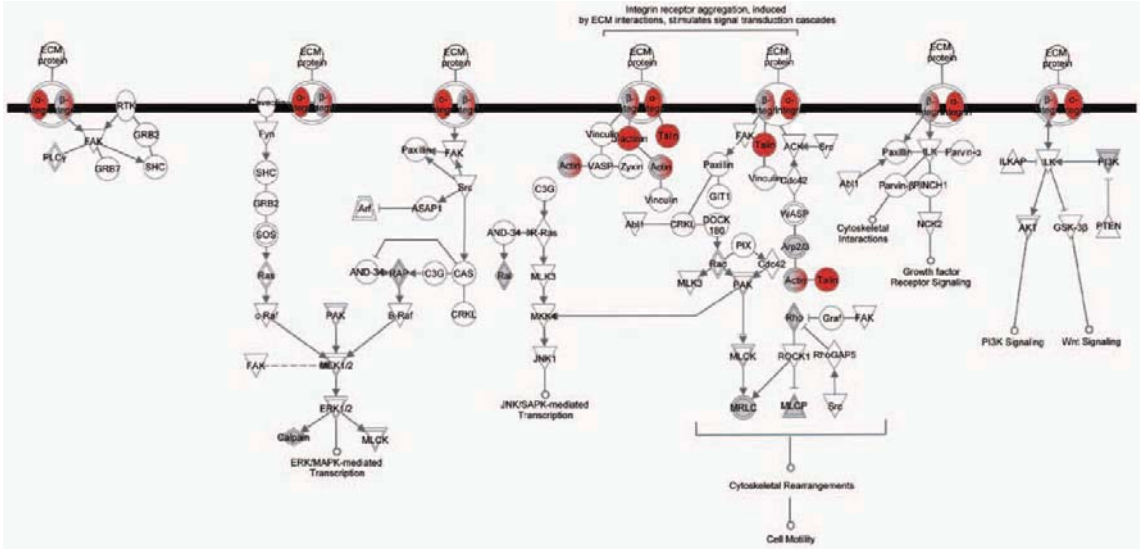

Figure 7. Integrin signaling pathway Integrin signaling pathway generated by the Ingenuity Pathway Analysis (IPA) software. Integrin and actin cytoskeleton signaling pathways were the top two pathways associated with cell-to-cell signaling. Each node represents a protein; the proteins in shaded nodes in the pathway are identified/relate to an identified protein in the proteomic analysis. While proteins in clear nodes are part of the pathway but have not been identified in the GV oocyte or cumulus datasets. Proteins in red nodes were shown higher expression in cumulus cells compared to GV oocyte.
doi:10.1371/journal.pone.0011240.g007

Ontology to identify membrane and nuclear proteins. Here, we take a much more comprehensive approach to include not only 'known' but also 'predicted' and 'hypothetical' proteins. By definition, predicted and hypothetical proteins do not have any functional annotation associated with them and they represent ~60% of total proteins in our oocyte and cumulus datasets. Hence we manually annotated all these predicted and hypothetical proteins based on sequence similarity and orthology from human, mouse and rat proteins. Overall our approach, while much more laborious and requiring recognized GO biocuration skills, provides more comprehensive information on the biological processes associated with bovine GV oocyte and cumulus cell function. Not only does this improved data quality allow a much better modeling, but we almost doubled the Gene ontology Annotation Quality (GAQ) score [25] (a quantitative measure of the GO annotation) of the proteins in our dataset and this will be valuable for others (the annotations are available at the AgBase database ([29]; www.agbase.msstate.edu). By manual literature biocuration, we also captured the quantitative positive and negative regulatory roles that these GV oocyte proteins play in the various biological processes. Overall we could identify and quantify proteins agonistic or antagonistic to biological processes including signal transduction, cell cycle regulation, DNA transcription, apoptosis regulation, protein metabolism and modification, generation of precursor metabolites and energy, cytoskeleton organization and biogenesis, and response to stress in GV oocyte; all of which may have putative roles in oocyte competence and maturation. This type of quantitative modeling is not possible with the available pathway analysis tools. Furthermore, we have analyzed the protein datasets identified in this experiment in the context of protein interaction networks and pathways using Ingenuity Pathway Analysis. This pathway and network modeling also reveals signaling pathways associated with the proteins involved in cell-to-cell signaling biological process which may have implications in various reproductive processes such as oocyte development, and maturation.

Our GO-based modeling identified that the percentage of proteins involved in translation in GV oocytes is twice as high as in cumulus cells and this is consistent with other's data showing that GV oocytes are transcriptionally and translationally active and also that proteins synthesized in this stage might be crucial to achieve maturation and successful subsequent development [19,47]. We identified a comparative up regulation of three biological processes—generation of precursor metabolites and energy, transport, and cell communication—in cumulus cells compared to oocytes. Up regulation of generation of precursor metabolites and energy in cumulus cells is consistent with a model that compensates for oocytes inability to metabolize glucose in which surrounding cumulus cells may absorb and metabolize glucose to provide products that can be utilized by oocytes for energy metabolism [48,49]. In support, cumulus cells nutritionally-support the developing oocytes by providing pyruvate [14,15,16]. Furthermore, stimulation of this nutritional support is in turn dependent on the presence of paracrine factors secreted by the oocyte (oocyte secreted factors (OSFs)); OSFs stimulate the glycolytic activity in cumulus cells by promoting the expression of genes involved in glycolysis [17] in mice, and our bovine model also suggests the glycolytic support of cumulus cells by showing higher expression of twelve proteins involved in glycolysis in cumulus cells associated with GV oocyte. Glycolysis is the metabolic pathway that converts glucose into pyruvate and the free energy released in this process is used to form the high energy compounds, ATP and NADH. We also identified oxidative phosphorylation as the most prominent metabolic pathway significantly represented only in cumulus cells, but not in oocyte suggesting a hypothesis that cumulus cells may also provide energy to oocytes in the form of ATP. Expression of core proteins (ATP synthases, cytochrome oxidases, NADH dehydrogenases, and ubiquinol cytochrome c reductase) involved in oxidative phosphorylaton was higher in cumulus cells compared to oocytes. NADH dehydrogenases and cytochrome C oxidases are, respectively, the first and last enzymes of the mitochondrial electron transport chain helping to establish a transmembrane difference of proton electrochemical potential that the ATP synthase then uses to synthesize ATP [50].

Cumulus cells provide a network of gap junction transmembrane channels facilitating two way communications for nutrient or paracrine factor exchange between oocytes and cumulus cells [1,3]. Several small metabolites (such as some energy substrates, nucleotides, and amino acids) are absorbed by oocytes mostly, or entirely, from surrounding cumulus cells via gap junctions [15,51]. Our bovine model also demonstrated an increase in transport in cumulus cells and this is in agreement with previous studies that cumulus cells have key role in amino acid and energy substrate uptake and transport of these substances to the oocyte.

Cumulus cells are essential for protecting oocytes from oxidative stress-induced apoptosis and DNA damage and their communication with oocytes is essential for development of an oocyte competent to undergo fertilization and embryogenesis [5,52]. We identified 91 proteins involved in cell-to-cell signaling from oocytes and cumulus cells. To effect intercellular communication several signaling pathways are necessary. We identified the integrin and actin cytoskeleton signaling pathways as the top two associated pathways with cell-to-cell signaling. Integrin signaling converges on cell cycle regulation, directing cells to live or die, to proliferate or to exit the cell cycle and differentiate [53]. Expression of ACTB, ACTN4, ITGAV, ITGA2, ITGB1, and TLN1, involved in integrin signaling and actin cytoskeleton signaling was higher in cumulus cells compared to oocytes. These proteins are the cell surface receptors present in the follicular basement membrane and around follicular cells and participate in cell attachment to matrix and mediate mechanical and chemical signals from it [54]. These signals in turn regulate the activities of cytoplasmic kinases, growth factor receptors, and ion channels that control the organization of the intracellular actin cytoskeleton [53,54]. The actin cytoskeleton of the human oocyte plays a key role in cell surface morphology as well processes associated with oocyte maturation [55]. Together this data suggests a hypothesis that oocytes depend on the presence of cumulus cells to generate specific cellular signals to coordinate their growth and maturation.

In addition to integrin and actin cytoskeleton signaling pathways, we also identified MAPK signaling, PI3K signaling, and ephrin receptor signaling pathways may involved in bidirectional oocyte-cumulus cell communication. MAPK and PI3K signaling are involved in activation of several membrane signaling molecules followed by sequential activation of protein kinases which involves highly regulated and modulated cascades of phosphorylation events [56]. Protein phosphorylation during this signaling plays a major role in the oocyte meiotic maturation and many protein kinases activated during oocyte maturation are also involved in cumulus cell proliferation and differentiation [57]. Expression of three proteins ITGA2, ITGB, and YWHAQ involved in MAPK and PI3K signaling was higher in cumulus cells compared to oocytes. YWHAQ, a protein of tyrosine 3-monooxygenase/tryptophan 5- monooxygenase activation protein family, modulate/complement intracellular events involving

phosphorylation dependent switching or protein modification in mammalian oocytes. The YWHAQ protein has been shown to protect phosphorylated proteins from inopportune dephosphorylation [58]. This is suggesting a hypothesis that cumulus cells have important role in oocyte maturation by regulating the protein phosphorylation, an important event during mammalian oocyte maturation.

We also observed expression of peroxiredoxins (PRDX's) involved in cell-to-cell signaling was significantly higher in oocytes compared to cumulus cells. PRDX's are peroxidases involved in antioxidant defence and intracellular signaling. Cumulus cells induce PRDX up regulation in oocytes [59] and our work also suggests that cumulus cells may have a key role in protecting oocytes from oxidative stress induced apoptosis and DNA damage via inducing oocytes to produce peroxiredoxins.

In conclusion, our systems biology modeling of bovine oocyte and cumulus proteomes, in the context of gene ontology and canonical protein interaction networks identified ninty-one proteins involved in the cell-to-cell signaling biological process that may have role in bidirectional communication between oocytes and cumulus cells. Most of the proteins involved in the cell-to-cell signaling biological process are the components of integrin, actin cytoskeleton, MAPK and PI3K signaling pathways that may have implications in various reproductive processes such as oocyte development, and maturation. In addition, GO-based analysis of differentially expressed proteins facilitated the identification that compared to GV oocytes, cumulus cells have higher expression of proteins involved in the generation of precursor metabolites and energy, transport, and cell communication. This systems biology model of bovine oocytes and cumulus cell potential interacting proteomes not only provides a foundation for understanding signaling and cell physiology at the GV stage of oocyte development, but is also valuable for comparative studies of other stages of oocyte development at the molecular level. The proteomes and biological process identified in this study can also serve as reference points for further comparative studies on immature and abnormal oocyte and cumulus cells elucidating the underlying molecular mechanisms involved in normal and pathological oocyte-cumulus cell communication. Furthermore, some of the proteins involved in cell-to-cell signaling may have value as molecular biomarkers which could be useful for assessing oocyte quality.

Materials and Methods

GV oocytes and cumulus cells

Separate from our previous work [19], ovaries from Holstein cows were collected by Biomed Inc. (Madison, WI, USA) from a local slaughter house. Cumulus–oocyte complexes (COC) were aspirated only from the follicles with a diameter of 2–8 mm using an 18-gauge needle attached to a vacuum system [60]. Only oocytes with intact cumulus cell layers and homogenously granulated cytoplasm were selected, washed three times in TL-HEPES supplemented with polyvinylpyrrolidone (3 mg/ml polyvinyl-pyrroline-40; Sigma), Na-pyruvate (0.2 mM) and gentamycin (25 µg/ml). To obtain oocytes free of cumulus cells, COCs were vortexed in TL-HEPES (3 min), oocytes were collected under a stereomicroscope, further vortexed with hyaluronidase to remove adhering cumulus cells completely (3 min), washed thrice in saline and stored in a cell lysis buffer (4°C) until use. The cumulus cells separated from the oocytes after the first vortex were centrifuged, washed twice with saline, and the resulted pellets were resuspended in the lysis buffer and stored (4°C) until use.

Protein extraction using differential detergent fractionation (DDF)

DDF sequentially extracts proteins from different cellular compartments using a series of detergents and provides information on cell location as well as increasing proteome coverage. Three replicates of five hundred GV oocytes and their surrounding CC were each subjected to DDF exactly as previously described [61]. Briefly, cytosolic proteins were extracted by six sequential incubations in a buffer containing digitonin (10 min each); next a fraction containing predominantly membrane proteins was isolated by incubating the cells in 10% Triton X-100 buffer for 30 min and then removing the soluble protein. Nuclear DDF buffer containing deoxycholate (DOC) was then added to the remaining pellet and subjected to freeze-thawing to disrupt the intact nuclear membrane. Nuclear proteins were collected from the resulting soluble fraction and the sample was then aspirated through an 18-guage needle and treated with a mixture of DNase I (50U, Invitrogen, Carlsbad, CA) and RNase A (50 mg, Sigma-Aldrich, St Louis, MO) at 37°C for 1 h to digest nucleic acids. The remaining undissolved pellet was then treated with a buffer containing 5% SDS.

Proteomics

Protein quantification and trypsin digestion of DDF fractions were done exactly as described [19]. Briefly, proteins were precipitated with 25% tricholoroacetic acid to remove salts and detergents, then resuspended in 0.1 M ammonium bicarbonate with 5% HPLC grade acetonitrile (ACN), reduced (Dithiothreitol, 5mM, 65°C, 5 min), alkylated (iodoacetamide, 10 mM, 30°C, 30 min) and then trypsin digested until there was no visible pellet (sequencing grade modified trypsin, Promega; 1:50 w/w 37°C, 16 h). Peptides were desalted using a peptide macrotrap (Michrom BioResources, Inc., Auburn, CA) and eluted using a 0.1% trifluoroacetic acid, 95% ACN solution. Desalted peptides were dried in a vacuum centrifuge and resuspended in 20µL of 0.1% formic acid and 5% ACN.

Two-dimensional liquid chromatography (LC) analysis was accomplished by strong cation exchange (SCX) followed by reverse phase (RP) coupled directly in line with an electrospray ionization (ESI) ion trap tandem mass spectrometer (LCQ; ThermoElectron Corp., San Jose, Calif, USA) essentially as described in [19]. The salt gradient applied in this study was different from our previous published method and was applied in steps of 0, 10, 15, 20, 25, 30, 35, 40, 45, 50, 57, 64, 90, and 700 mM ammonium acetate in 5% acetonitrile (ACN) and 0.1% formic acid. The reverse phase gradient used 0.1% formic acid in ACN and increased the ACN concentration in a linear gradient from 5% to 30% in 20 minutes and then 30% to 95% in 7 minutes, followed by 5% for 10 minutes for 0, 10, 15, 25, 30, 45, 64, 90, and 700 mM salt gradient steps. For 20, 35, 40, 50, and 57 mM salt gradient steps, ACN concentration was increased in a linear gradient from 5% to 40% in 65 minutes, 95% for 15 minutes, and 5% for 20 minutes.

Mass spectra and tandem mass spectra were searched against an *in silico* trypsin-digested bovine database using TurboSEQUEST™ (Bioworks Browser 3.2; ThermoElectron). The bovine database used in this study was different from the previous publication and here we used database of bovine Reference Sequence (RefSeq) proteins downloaded from the National Center for Biotechnology Institute [NCBI; 08/28/2007; 25,078 entries]. The Reference Sequence (RefSeq) collection aims to provide a comprehensive, integrated, non-redundant, and well-annotated set of sequences of proteins [62]. Trypsin digestion including mass changes due to cysteine carbamidomethylation (C, 57.02 Da) and

methionine mono- and di-oxidation (15.99 Da and 32 Da), were included in the search criteria. The peptide (MS precursor ion) mass tolerance was set to 1.5 Da and the fragment ion (MS^2) mass tolerance was set to 1.0 Da. Rsp Value less than 5.

As a primary filter we first limited our Sequest search output to include only peptides ≥6 amino acids long, with ΔCn ≥0.1 and Sequest cross correlation (Xcorr) scores of 1.9, 2.2, and 3.7 for +1, +2, and +3 charge states, respectively. We next used a decoy database search strategy [40] (using the same search criteria as the real database search) to calculate p-values for peptide identifications as this allows us to assign the probability of a false identification based on the real data from the experiment itself [40,63,64]. Since the accuracy of peptide identification depends on the charge state we calculated P-values for +1, +2, and +3 charge states separately. The probability that peptide identification from the original database is a random match (P-value) is estimated based on the probability that a match against the decoy database will achieve the same Xcorr [65,66]. Protein probabilities were calculated exactly as described [67,68] and we used only proteins identified by peptides with a $p<0.05$ for further analysis. All protein identifications and MS/MS data have been submitted to the PRoteomics IDEntifications database (PRIDE [26]; accession numbers 8691, 8692, 8693, 8694, 8695 and 8696). PRIDE submission requirements are based on the proposed guidelines by proteomics standards initiative [69] and include all the peptides identified for each protein with their sequence, charge state, Xcorr, and delta cn.

We used an isotope-free quantification method [70] and a custom program *ProtQuant* [71] to identify differences in protein expression between oocyte and cumulus cell datasets. *ProtQuant* is a java based tool for label-free quantification that uses a spectral counting method with increased specificity (and thus decreased false positive i.e. type I errors). This increased specificity is achieved by incorporating the quantitative aspects of the Sequest cross correlation (XCorr) into the spectral counting method. *ProtQuant* also computes the statistical significance of differential expression by one-way ANOVA ($\alpha \leq 0.05$).

Gene Ontology (GO) Annotation

To identify the biological processes of all proteins in our datasets we used Gene Ontology annotations and the GO resources and tools available at AgBase [29]. We had to overcome the limitation that most literature for any species is not yet curated and so functional annotations from this literature are not yet available at GO databases. First, we used *GORetriver* to obtain all existing GO annotations available for known proteins in our datasets [72]. To obtain GO annotations for proteins without existing annotation, but between 70–90% amino acid sequence identity to presumptive orthologs with GO annotation were annotated using GO*anna* by manually checking the similarity between our proteins and orthologs [72]. GO biological process annotations for proteins were grouped into more generalized categories using *GOSlim viewer* [72]. Significant differences in percentage of proteins involved in various identified GO categories in between GV oocyte and cumulus cells were evaluated by *student's t-test*. Differences at $p<0.05$ were considered statistically significant.

GO annotation quality (*GAQ*) scores were calculated exactly as described [25] to quantify the improvement in GO annotation quality of GV oocyte and cumulus proteins due to re-annotation in this study compared to existing annotations. The *GAQ* score is a quantitative measure of the GO annotation of a set of gene products based on the number of GO annotations available, the level of detail of the annotation and the types of evidence used to make these GO annotations.

GO based quantitative modeling

Our GO based quantitative modeling of GV stage oocyte was based on specific hypothesis framed in GO biological process (GOBP) terms defining the phenotype of oocyte competence and maturation. Although GOBP terms exists for gene products that effect and affect oocyte competence and maturation, and there is functional literature on these genes, the literature and GO are unconnected [42]. To connect the data with the GO, we manually annotated the literature of GV oocyte proteins to compare the number of proteins that were either agonistic or antagonistic for each biological process including signal transduction, cell cycle regulation, DNA transcription, apoptosis regulation, protein metabolism and modification, generation of precursor metabolites and energy, cytoskeleton organization and biogenesis, and response to stress and calculated the net effect of each process in GV oocyte; all of which may have a putative role in oocyte competence and maturation.

Modeling using Ingenuity pathway analysis

To gain insights into the biological pathways and networks that are statistically significantly represented in our proteomic datasets we used Ingenuity Pathways Analysis (IPA; Ingenuity Systems, California). We imported protein accessions from our GV oocyte and cumulus cell datasets into IPA. IPA selects "focus proteins" to be used for generating biological networks. Focus proteins are the proteins from our datasets that are mapped to corresponding gene objects in the Ingenuity Pathways Knowledgebase (IPKB) and are known to interact with other proteins based on published, peer reviewed content in the IPKB. Based on these interactions IPA builds networks with a size of no more than 35 genes or proteins. A p-value for each network and canonical pathway is calculated according to the fit of the user's set of significant genes/proteins. IPA computes a score for each network from the p-value that indicates the likelihood of the focus proteins in a network being found together due to chance. We selected only networks scoring ≥2, which have >99% confidence of not being generated by chance [73] [74]. Biological functions are assigned to each network by using annotations from scientific literature and stored in the IPKB. A Fisher exact test is used to calculate the p-value determining the probability of each biological function/disease or pathway being assigned by chance. We used $P \leq 0.05$ to select highly significant biological functions and pathways represented in our proteomic datasets [73].

Supporting Information

Figure S1 Comparison of proteins identified in present study with previous published study by Memili et al., 2007.
Found at: doi:10.1371/journal.pone.0011240.s001 (0.01 MB PDF)

Table S1 List of proteins identified in germinal vesicle (GV) stage oocyte and cumulus cells. Proteins identified by DDF-MudPIT and their distribution in GV oocyte and cumulus cells. GI numbers of the identified proteins (column A) and corresponding protein names (assigned by NCBI; column B). Protein distribution in GV oocyte or cumulus cells or common to both (O: Oocyte; CC: Cumulus Cells; C: common to both; column C). Numbers of peptides and Sequest cross correlation scores (ΣXcorr) in columns D, F and E, G for oocyte and cumulus cells respectively.
Found at: doi:10.1371/journal.pone.0011240.s002 (0.29 MB XLS)

Table S2 List of differentially expressed proteins in cumulus cells compared with GV oocytes. Gene name, numbers of peptides, Sequest cross correlation (ΣXcorr) score and P value for of each differentially expressed protein in GV oocyte and cumulus cells.
Found at: doi:10.1371/journal.pone.0011240.s003 (0.10 MB XLS)

Table S3 Distribution of GV oocyte and cumulus cell proteins by subcellular compartments. The classification of the identified GV oocyte and cumulus proteins among subcellular compartments was performed using cellular component GO annotations.
Found at: doi:10.1371/journal.pone.0011240.s004 (0.02 MB XLS)

Table S4 Proteins involved in cell-to-cell signaling biological process.
Found at: doi:10.1371/journal.pone.0011240.s005 (0.04 MB XLS)

Acknowledgments

We would like to acknowledge Dr. Susan M Bridges, Dr. Bindu Nanduri and Bryce Magee for helping us with Biomarker discovery tool, Dr. Teresia Buza for GAQ score and Dr. Tibor Pechan for running the mass spectrometer.

Author Contributions

Conceived and designed the experiments: DP EM SB. Performed the experiments: DP. Analyzed the data: DP. Contributed reagents/materials/analysis tools: SB. Wrote the paper: DP. Performed the proteomics sample preparation, data generation, analyzed and interpreted proteomic data, computational systems biology modeling and analysis and wrote the draft of the manuscript: DP. Facilitated sample collection, contributed to design of the study, provided expert knowledge and interpretation in reproductive biology: EM. Helped analyze and interpret the statistical analysis of the proteomics data and helped to draft the manuscript: SCB. Reviewed manuscript draft: SBC.

References

1. Eppig JJ (1991) Intercommunication between mammalian oocytes and companion somatic cells. Bioessays 13: 569–574.
2. Gilchrist RB, Ritter LJ, Armstrong DT (2004) Oocyte-somatic cell interactions during follicle development in mammals. Anim Reprod Sci 82–83: 431–446.
3. Grazul-Bilska AT, Reynolds LP, Redmer DA (1997) Gap junctions in the ovaries. Biol Reprod 57: 947–957.
4. Tanghe S, Van Soom A, Nauwynck H, Coryn M, de Kruif A (2002) Minireview: Functions of the cumulus oophorus during oocyte maturation, ovulation, and fertilization. Mol Reprod Dev 61: 414–424.
5. Assidi M, Dufort I, Ali A, Hamel M, Algriany O, et al. (2008) Identification of Potential Markers of Oocyte Competence Expressed in Bovine Cumulus Cells Matured with Follicle-Stimulating Hormone and/or Phorbol Myristate Acetate In Vitro. Biol Reprod.
6. Vozzi C, Formenton A, Chanson A, Senn A, Sahli R, et al. (2001) Involvement of connexin 43 in meiotic maturation of bovine oocytes. Reproduction 122: 619–628.
7. Fatehi AN, Zeinstra EC, Kooij RV, Colenbrander B, Bevers MM (2002) Effect of cumulus cell removal of in vitro matured bovine oocytes prior to in vitro fertilization on subsequent cleavage rate. Theriogenology 57: 1347–1355.
8. Mori T, Amano T, Shimizu H (2000) Roles of gap junctional communication of cumulus cells in cytoplasmic maturation of porcine oocytes cultured in vitro. Biol Reprod 62: 913–919.
9. Wongsrikeao P, Kaneshige Y, Ooki R, Taniguchi M, Agung B, et al. (2005) Effect of the removal of cumulus cells on the nuclear maturation, fertilization and development of porcine oocytes. Reprod Domest Anim 40: 166–170.
10. de Matos DG, Furnus CC, Moses DF (1997) Glutathione synthesis during in vitro maturation of bovine oocytes: role of cumulus cells. Biol Reprod 57: 1420–1425.
11. Matzuk MM, Burns KH, Viveiros MM, Eppig JJ (2002) Intercellular communication in the mammalian ovary: oocytes carry the conversation. Science 296: 2178–2180.
12. Su YQ, Sugiura K, Wigglesworth K, O'Brien MJ, Affourtit JP, et al. (2008) Oocyte regulation of metabolic cooperativity between mouse cumulus cells and oocytes: BMP15 and GDF9 control cholesterol biosynthesis in cumulus cells. Development 135: 111–121.
13. Sugiura K, Su YQ, Diaz FJ, Pangas SA, Sharma S, et al. (2007) Oocyte-derived BMP15 and FGFs cooperate to promote glycolysis in cumulus cells. Development 134: 2593–2603.
14. Biggers JD, Whittingham DG, Donahue RP (1967) The pattern of energy metabolism in the mouse oocyte and zygote. Proc Natl Acad Sci U S A 58: 560–567.
15. Buccione R, Schroeder AC, Eppig JJ (1990) Interactions between somatic cells and germ cells throughout mammalian oogenesis. Biol Reprod 43: 543–547.
16. Kim H, Schuetz AW (1991) Regulation of nuclear membrane assembly and maintenance during in vitro maturation of mouse oocytes: role of pyruvate and protein synthesis. Cell Tissue Res 265: 105–112.
17. Sugiura K, Pendola FL, Eppig JJ (2005) Oocyte control of metabolic cooperativity between oocytes and companion granulosa cells: energy metabolism. Dev Biol 279: 20–30.
18. Eppig JJ, Wigglesworth K, Chesnel F (1993) Secretion of cumulus expansion enabling factor by mouse oocytes: relationship to oocyte growth and competence to resume meiosis. Dev Biol 158: 400–409.
19. Memili E, Peddinti D, Shack LA, Nanduri B, McCarthy F, et al. (2007) Bovine germinal vesicle oocyte and cumulus cell proteomics. Reproduction 133: 1107–1120.
20. Sutton ML, Cetica PD, Beconi MT, Kind KL, Gilchrist RB, et al. (2003) Influence of oocyte-secreted factors and culture duration on the metabolic activity of bovine cumulus cell complexes. Reproduction 126: 27–34.
21. Amano T, Mori T, Matsumoto K, Iritani A, Watanabe T (2005) Role of cumulus cells during maturation of porcine oocytes in the rise in intracellular Ca2+ induced by inositol 1,4,5-trisphosphate. Theriogenology 64: 261–274.
22. Zhang L, Jiang S, Wozniak PJ, Yang X, Godke RA (1995) Cumulus cell function during bovine oocyte maturation, fertilization, and embryo development in vitro. Mol Reprod Dev 40: 338–344.
23. Richter JD (1991) Translational control during early development. Bioessays 13: 179–183.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29.
25. Buza TJ, McCarthy FM, Wang N, Bridges SM, Burgess SC (2008) Gene Ontology annotation quality analysis in model eukaryotes. Nucleic Acids Res 36: e12.
26. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, et al. (2005) PRIDE: the proteomics identifications database. Proteomics 5: 3537–3545.
27. McCarthy FM, Cooksey AM, Wang N, Bridges SM, Pharr GT, et al. (2006) Modeling a whole organ using proteomics: the avian bursa of Fabricius. Proteomics 6: 2759–2771.
28. Lubec G, Afjehi-Sadat L, Yang JW, John JP (2005) Searching for hypothetical proteins: theory and practice based upon original data and literature. Prog Neurobiol 77: 90–127.
29. McCarthy FM, Bridges SM, Wang N, Magee GB, Williams WP, et al. (2007) AgBase: a unified resource for functional analysis in agriculture. Nucleic Acids Res 35: D599–603.
30. Bhojwani M, Rudolph E, Kanitz W, Zuehlke H, Schneider F, et al. (2006) Molecular analysis of maturation processes by protein and phosphoprotein profiling during in vitro maturation of bovine oocytes: a proteomic approach. Cloning Stem Cells 8: 259–274.
31. Coonrod SA, Wright PW, Herr JC (2002) Oolemmal proteomics. J Reprod Immunol 53: 55–65.
32. Ellederova Z, Halada P, Man P, Kubelka M, Motlik J, et al. (2004) Protein patterns of pig oocytes during in vitro maturation. Biol Reprod 71: 1533–1539.
33. Hamamah S, Matha V, Berthenet C, Anahory T, Loup V, et al. (2006) Comparative protein expression profiling in human cumulus cells in relation to oocyte fertilization and ovarian stimulation protocol. Reprod Biomed Online 13: 807–814.
34. Ma M, Guo X, Wang F, Zhao C, Liu Z, et al. (2008) Protein expression profile of the mouse metaphase-II oocyte. J Proteome Res 7: 4821–4830.
35. Meng Y, Liu XH, Ma X, Shen Y, Fan L, et al. (2007) The protein profile of mouse mature cumulus-oocyte complex. Biochim Biophys Acta 1774: 1477–1490.
36. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, et al. (2002) Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics 1: 304–313.
37. Cox B, Kislinger T, Emili A (2005) Integrating gene and protein expression data: pattern analysis and profile mining. Methods 35: 303–314.
38. Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19: 1720–1730.
39. Nie L, Wu G, Culley DE, Scholten JC, Zhang W (2007) Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. Crit Rev Biotechnol 27: 63–75.
40. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4: 207–214.
41. Dalbies-Tran R, Papillier P, Pennetier S, Uzbekova S, Monget P (2005) Bovine mater-like NALP9 is an oocyte marker gene. Mol Reprod Dev 71: 414–421.

42. Fair T, Carter F, Park S, Evans AC, Lonergan P (2007) Global gene expression analysis during bovine oocyte in vitro maturation. Theriogenology 68 Suppl 1: S91–97.
43. Pennetier S, Perreau C, Uzbekova S, Thelie A, Delaleu B, et al. (2006) MATER protein expression and intracellular localization throughout folliculogenesis and preimplantation embryo development in the bovine. BMC Dev Biol 6: 26.
44. Pfeffer PL, Sisco B, Donnison M, Somers J, Smith C (2007) Isolation of genes associated with developmental competency of bovine oocytes. Theriogenology 68 Suppl 1: S84–90.
45. Racedo SE, Wrenzycki C, Lepikhov K, Salamone D, Walter J, et al. (2009) Epigenetic modifications and related mRNA expression during bovine oocyte in vitro maturation. Reprod Fertil Dev 21: 738–748.
46. Thelie A, Papillier P, Pennetier S, Perreau C, Traverso JM, et al. (2007) Differential regulation of abundance and deadenylation of maternal transcripts during bovine oocyte maturation in vitro and in vivo. BMC Dev Biol 7: 125.
47. Memili E, First NL (1999) Control of gene expression at the onset of bovine embryonic development. Biol Reprod 61: 1198–1207.
48. Donahue RP, Stern S (1968) Follicular cell support of oocyte maturation: production of pyruvate in vitro. J Reprod Fertil 17: 395–398.
49. Leese HJ, Barton AM (1985) Production of pyruvate by isolated mouse cumulus cells. J Exp Zool 234: 231–236.
50. Weber J, Senior AE (2003) ATP synthesis driven by proton transport in F1F0-ATP synthase. FEBS Lett 545: 61–70.
51. Sugiura K, Eppig JJ (2005) Society for Reproductive Biology Founders' Lecture 2005. Control of metabolic cooperativity between oocytes and their companion granulosa cells by mouse oocytes. Reprod Fertil Dev 17: 667–674.
52. Bettegowda A, Patel OV, Lee KB, Park KE, Salem M, et al. (2008) Identification of Novel Bovine Cumulus Cell Molecular Markers Predictive of Oocyte Competence: Functional and Diagnostic Implications. Biol Reprod.
53. Giancotti FG, Ruoslahti E (1999) Integrin signaling. Science 285: 1028–1032.
54. Monniaux D, Huet-Calderwood C, Le Bellego F, Fabre S, Monget P, et al. (2006) Integrins in the ovary. Semin Reprod Med 24: 251–261.
55. Veselska RRJ (2001) Cortical actin cytoskeleton in human oocytes: a comparison with mouse oocytes. Scripta medica 74: 265–274.
56. Seger R, Krebs EG (1995) The MAPK signaling cascade. Faseb J 9: 726–735.
57. Su YQ, Wigglesworth K, Pendola FL, O'Brien MJ, Eppig JJ (2002) Mitogen-activated protein kinase activity in cumulus cells is essential for gonadotropin-induced oocyte meiotic resumption and cumulus expansion in the mouse. Endocrinology 143: 2221–2232.
58. Snow AJ, Puri P, Acker-Palmer A, Bouwmeester T, Vijayaraghavan S, et al. (2008) Phosphorylation-dependent interaction of tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein (YWHA) with PADI6 following oocyte maturation in mice. Biol Reprod 79: 337–347.
59. Leyens G, Verhaeghe B, Landtmeters M, Marchandise J, Knoops B, et al. (2004) Peroxiredoxin 6 is upregulated in bovine oocytes and cumulus cells during in vitro maturation: role of intercellular communication. Biol Reprod 71: 1646–1651.
60. Sagirkaya H, Misirlioglu M, Kaya A, First NL, Parrish JJ, et al. (2006) Developmental and molecular correlates of bovine preimplantation embryos. Reproduction 131: 895–904.
61. Peddinti D, Nanduri B, Kaya A, Feugang JM, Burgess SC, et al. (2008) Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. BMC Syst Biol 2: 19.
62. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35: D61–65.
63. Elias JE, Gibbons FD, King OD, Roth FP, Gygi SP (2004) Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nat Biotechnol 22: 214–219.
64. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2: 43–50.
65. Park GW, Kwon KH, Kim JY, Lee JH, Yun SH, et al. (2006) Human plasma proteome analysis by reversed sequence database search and molecular weight correlation based on a bacterial proteome analysis. Proteomics 6: 1121–1132.
66. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, et al. (2005) Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res 4: 53–62.
67. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75: 4646–4658.
68. MacCoss MJ, Wu CC, Yates JR, 3rd (2002) Probability-based validation of protein identifications using a modified SEQUEST algorithm. Anal Chem 74: 5593–5599.
69. Orchard S, Jones AR, Stephan C, Binz PA (2007) The HUPO pre-congress Proteomics Standards Initiative workshop. HUPO 5th annual World Congress. Long Beach, CA, USA 28 October–1 November 2006. Proteomics 7: 1006–1008.
70. Nanduri B, Lawrence ML, Vanguri S, Pechan T, Burgess SC (2005) Proteomic analysis using an unfinished bacterial genome: the effects of subminimum inhibitory concentrations of antibiotics on Mannheimia haemolytica virulence factor expression. Proteomics 5: 4852–4863.
71. Bridges SM, Bryce GB, Wang N, Williams WP, Burgess SC, et al. (2007) ProtQuant: a tool for the label -free quantification of mudPIT proteomics data. BMC Bioinformatics 8.
72. McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, et al. (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics 7: 229.
73. Gerling IC, Singh S, Lenchik NI, Marshall DR, Wu J (2006) New data analysis and mining approaches identify unique proteome and transcriptome markers of susceptibility to autoimmune diabetes. Mol Cell Proteomics 5: 293–305.
74. Huang Y, Yan J, Lubet R, Kensler TW, Sutter TR (2006) Identification of novel transcriptional networks in response to treatment with the anticarcinogen 3H-1,2-dithiole-3-thione. Physiol Genomics 24: 144–153.

CHAPTER IV

COMPREHENSIVE PROTEOMIC ANALYSIS OF BOVINE SPERMATOZOA OF VARYING FERTILITY RATES AND IDENTIFICATION OF BIOMARKERS ASSOCIATED WITH FERTILITY[1]

[1] Reprint from D. Peddinti, B. Nanduri, A. Kaya, J. Feugang, S.C. Burgess, E. Memilli. Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. BMC Syst Biol. 2008 Feb 22;2:19. PMID: 18294385.

BMC Systems Biology

BioMed Central

Research article

Open Access

Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility

Divyaswetha Peddinti[†1,4], Bindu Nanduri[†1,4], Abdullah Kaya[2], Jean M Feugang, Shane C Burgess[1,4,5] and Erdogan Memili[*†3]

Address: [1]Department of Basic Sciences, Mississippi State University, Mississippi State, MS 39762, USA, [2]Alta Genetics, Inc., Watertown, WI, USA, [3]Department of Animal and Dairy Sciences, Mississippi State University, Mississippi State, MS 39762, USA, [4]Institute for Digital Biology, Mississippi State University, Mississippi State, MS 39762, USA and [5]Mississippi Agricultural and Forestry Experimental station, Mississippi State, MS 39762, USA

Email: Divyaswetha Peddinti - dpeddinti@cvm.msstate.edu; Bindu Nanduri - bnanduri@cvm.msstate.edu; Abdullah Kaya - akaya@altagenetics.com; Jean M Feugang - jn181@ads.msstate.edu; Shane C Burgess - burgess@cvm.msstate.edu; Erdogan Memili* - em149@ads.msstate.edu

* Corresponding author †Equal contributors

Published: 22 February 2008

BMC Systems Biology 2008, **2**:19 doi:10.1186/1752-0509-2-19

Received: 10 July 2007

Accepted: 22 February 2008

This article is available from: http://www.biomedcentral.com/1752-0509/2/19

Abstract

Background: Male infertility is a major problem for mammalian reproduction. However, molecular details including the underlying mechanisms of male fertility are still not known. A thorough understanding of these mechanisms is essential for obtaining consistently high reproductive efficiency and to ensure lower cost and time-loss by breeder.

Results: Using high and low fertility bull spermatozoa, here we employed differential detergent fractionation multidimensional protein identification technology (DDF-Mud PIT) and identified 125 putative biomarkers of fertility. We next used quantitative Systems Biology modeling and canonical protein interaction pathways and networks to show that high fertility spermatozoa differ from low fertility spermatozoa in four main ways. Compared to sperm from low fertility bulls, sperm from high fertility bulls have higher expression of proteins involved in: energy metabolism, cell communication, spermatogenesis, and cell motility. Our data also suggests a hypothesis that low fertility sperm DNA integrity may be compromised because cell cycle: G_2/M DNA damage checkpoint regulation was most significant signaling pathway identified in low fertility spermatozoa.

Conclusion: This is the first comprehensive description of the bovine spermatozoa proteome. Comparative proteomic analysis of high fertility and low fertility bulls, in the context of protein interaction networks identified putative molecular markers associated with high fertility phenotype.

Background

Male infertility is a major problem for mammalian reproduction. The nature of sub-fertility due to the male is as complex as that of the female [1]. Infertility due to male factor contributes approximately 40% of the infertility cases in humans. For this reason it is very important to investigate the factors that affect male fertility. Here we used bovine spermatozoa to model human male fertility

because cattle provide several advantages as a model for male factor infertility. These include good breeding records fertility data records and progeny records. In cattle breeding, Artificial insemination (AI), a common breeding technique, utilizes semen from genetically superior sires to inseminate cows. In the United States more than ~70% of cows are bred by AI but only ~50% of these matings result in successful full term pregnancy [2]. The underlying molecular events/mechanisms that determine the fertilizing potential of a semen sample are not well defined. A thorough understanding of these mechanisms is essential for obtaining consistently high reproductive efficiency and to ensure lower cost and time-loss by breeder.

Fertility traits of semen can be categorized as compensable or uncompensable [1,3-7]. Defects in compensable traits (motility and morphology) can be overcome by increasing the number of spermatozoa per insemination [1]. Defects in uncompensable traits affect the function of spermatozoa during the later stages of fertilization and in embryonic development [1,8] and as such cannot be compensated. Uncompensable traits include nuclear vacuoles [9], morphological deficiencies that do not suppress movement [4], defective chromatin structure [10]. Low fertility in bulls has an uncompensable component that includes reduced cleavage rate and delayed pronuclear formation following in vitro fertilization [1,11]. Currently available fertility assays assess the defects that affect functional competence of spermatozoa (i.e. capacitation, acrosome reaction, sperm-oocyte interaction) [8,12], however these cannot definitively predict fertility. At present, the molecular nature of sperm fertility defects or biomarkers for accurate fertility prediction is not known [13].

Spermatozoa are transcriptionally inactive so the only comprehensive method to understand the molecular functions in spermatozoa is via proteomics [13]. Published proteomic studies with bull spermatozoa described the sub-proteome of the sperm and functions of proteins from its surrounding cells. Accessory gland (AG) proteins were shown to modulate important sperm functions after ejaculation and in the female reproductive tract such as capacitation, acrosome reaction, sperm-oocyte interaction, and sperm protection [14]. It is known that fertile associated antigen (FAA), a heparin binding protein from seminal vesicles and prostate glands, binds to spermatozoa membrane and modulates heparin-sperm interactions that are indicative of fertility [15]. Two seminal plasma proteins such as, prostaglandin-D-synthetase and osteoponin were more abundant in the semen of high fertility bulls when compared to low fertility bulls [16,17].

Here we describe a comprehensive proteomic analysis of bull sperm using differential detergent fractionation (DDF) two-dimensional liquid chromatography followed by electrospray ionization tandem mass spectrometry (DDF 2-LC ESI MS^2; [18]). We compared protein expression profiles of sperm from high and low fertility bulls to characterize the differences in fertility at the protein level. Our results show that expression of 2051 and 2281 proteins was specific to high and low fertility bull spermatozoa, respectively and 1518 proteins were common to both. Differential expression of 125 proteins was significant between high and low fertility bull spermatozoa and these proteins are potential biomarkers for bovine male fertility. Biological systems utilize highly complex, interrelated metabolic and signaling pathways to function. Therefore, to identify signaling pathways involved in fertility, we carried out systems modeling of our proteomic datasets using Gene Ontology (GO) and Ingenuity Pathway Analysis (IPA). We identified differences in the signaling pathways between high and low fertility bull spermatozoa and found that EGF and PDGF signaling pathways were specific to high fertility.

Results

Proteome profiles of spermatozoa from high and low fertility bulls

We identified 3569 and 3799 proteins in high and low fertility group spermatozoa respectively (see additional file 1). Among these 1518 (20.4%) were common to both groups and 2051 and 2281 proteins were unique to high and low fertility groups respectively (Figure 1). Only those proteins identified by at least three peptides were included in the analysis for differential expression and we identified 125 proteins as differentially-expressed between the high and low fertility spermatozoa. Compared to low fertility bull spermatozoa, expression of 74 proteins increased and there was a decrease in the expression of 51 proteins in high fertility spermatozoa (Table 1). Only a small proportion of proteins identified in this study have been previously described (15.1% of the high fertility group specific and 14.3% of the low fertility group specific proteins (Figure 1)). The majority of the identified proteins are 'predicted' (i.e. predicted based on sequence similarity to known proteins in other species and are frequently found in NRPD database for species that have had their genomes sequenced [19]). We contributed to the annotation of the newly sequenced bovine genome by experimentally confirming the in vivo expression of 4,313 electronically predicted proteins (see additional file 1). We also identified 10.6% and 9.8% 'hypothetical' (i.e. proteins predicted from nucleic acid sequences and that have not been shown to exist by experimental protein chemical evidence [20]) proteins specific to high fertility and low fertility spermatozoa respectively.

Predicted and hypothetical proteins do not have any functional annotation associated with them and they repre-

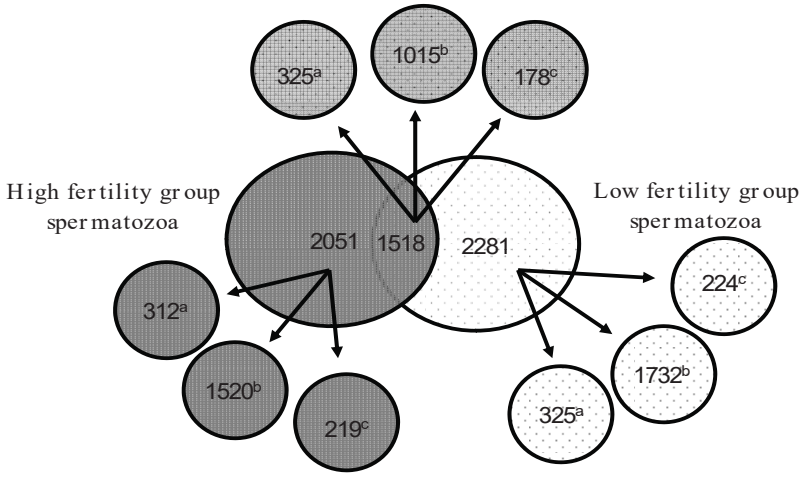

Figure 1
Comparison of proteins identified in high fertility and low fertility spermatozoa. Distribution of predicted, known and hypothetical proteins is shown. [a] known proteins, [b] predicted proteins, [c] hypothetical proteins.

sent ~80% of differentially expressed proteins between high and low fertility spermatozoa (Table 1). This poses a problem for meaningful biological modeling of our data without carrying out some functional annotation first. Therefore, we annotated all differentially expressed proteins in our data sets using AgBase GO resources.

Membrane and nuclear proteins

Membrane and nuclear proteins are fundamental for inter and intra cellular signaling and are thus fundamental for modeling cell-cell interactions. Sperm oocyte fusion is a key element for fertilization. This process is facilitated by sperm surface proteins and leads to specific binding of the sperm surface-active component with the egg zona pellucida and, ultimately, sperm-egg fusion [21]. To identify proteins from the sperm membrane and the nucleus which function in cell fusion, we focused on membrane and nuclear proteins identified in our datasets. Based on the GO associations of known proteins, 40.6% (395) are membrane proteins. We also identified 112 nuclear proteins based on GO associations. Biological process annotation of membrane proteins revealed that majority of membrane proteins involved in transport (33%), cell communication (18%) and metabolism (17%).

We GO annotated all differentially expressed proteins and applied the generic GO Slim [22] to identify 7 functional super-categories represented in differentially expressed proteins in high fertility spermatozoa. Most GO Slim categories, including processes such as metabolism, cell communication and cell motility showed overall up regulation of protein expression in the high fertility group while transport proteins showed an overall down regulation in the high fertility group (Figure 2).

High fertility and low fertility sperm proteomes: molecular network and pathway analysis

Protein identification from biological samples on a global scale is important. However, there is a need to move beyond this level of analysis; Instead of simply enumerating a list of proteins, the analysis needs to include their interactions as parts of complexes, pathways and biological networks. To achieve this level of analysis with our high fertility and low fertility spermatozoa proteomic datasets we used Ingenuity Pathway Analysis (IPA). At IPA thresholds for significance, 71, and 73 networks and 68, and 73 functions/diseases were significantly represented in the proteomes of high fertility and low fertility spermatozoa respectively. The top 10 functions/diseases (ranked based on significance), and the associated signaling pathways are shown in Table 2 and Table 3 for proteomes of high and low fertility groups respectively. Analysis of the top 10 functions revealed that functions like cellular movement, cell to cell signaling and interaction were identified only in the high fertility sperm proteome (Table 2). Whereas, functions like cell death and reproductive system disease were identified only in the low fertility sperm proteome (Table 3).

Compared to low fertility sperm proteome (9), the high fertility sperm proteome (20) had a 2-fold enrichment in signaling pathways. However, the number of significant

Table 1: Differentially expressed proteins.

Accesion	Name	Peptides		∑Xcorr		P-value	Regulation
		HF	LF	HF	LF		
115496714	Actin-like 7B	16	7	60.06	42.11	0.02243	up
77736067	Acyl-CoA thioesterase 9	14	4	39.23	16.48	2.11E-04	up
41386786	A-kinase anchor protein 4	679	581	2581.8	2424.3	0.001694	up
30794280	Albumin	7	1	27.09	5.24	1.49E-04	up
60302887	Aldose reductase	1	3	3.03	12.82	0.03293	down
27807289	Annexin A2	4	10	29.17	26.46	0.04155	down
84490369	ATP synthase, H+ transporting, mitochondrial	16	8	48.01	31.88	0.0333	up
28603752	ATP synthase, H+ transporting, mitochondrial F0	7	1	18.99	7.14	0.005907	up
28461221	ATP synthase, H+ transporting, mitochondrial F1 complex	201	174	765.55	700.71	0.006727	up
28461251	ATPase inhibitory factor 1 precursor	18	10	45.95	30.03	0.01581	up
27807145	Casein kinase 2, alpha prime polypeptide	3	0	11.25	0	5.01E-04	up
60101831	Cytochrome c oxidase subunit III	5	4	17.45	11.47	0.01008	up
84000107	Glycoprotein (transmembrane) nmb	0	3	6.9	8.98	0.0133	down
84000035	Hypothetical protein LOC504736	3	1	8.66	2.04	0.03204	up
115497288	Hypothetical protein LOC506544	4	7	13.54	18.39	0.04931	down
78369248	Hypothetical protein LOC509274	18	10	77.94	51.72	0.01217	up
115496338	Hypothetical protein LOC516024	27	22	109.76	67.93	1.54E-05	up
115495377	Hypothetical protein LOC520260	44	58	201.9	232.07	0.04396	down
114052468	Hypothetical protein LOC532785	8	19	32.19	55.25	0.003313	down
115496742	Hypothetical protein LOC534599	27	11	88.93	61.57	0.02739	up
84000301	Hypothetical protein LOC534927	9	3	32.16	14.67	0.01593	up
115495951	Hypothetical protein LOC540767	20	12	62.88	41.8	0.004421	up
94966950	Hypothetical protein LOC614199	4	0	10.84	1.66	0.007311	up
84000391	Hypothetical protein LOC615316	11	5	35.17	24.34	0.02524	up
115497750	Hypothetical protein LOC617117	19	12	58.06	44.14	0.03432	up
116004271	Hypothetical protein LOC767959	4	0	12.9	3.03	0.04429	up
27805989	Lysyl oxidase-like 4	1	3	1.59	7.16	0.03124	down
27806307	Mitochondrial ATP synthase, O subunit	13	8	51.79	34.78	0.03731	up
28461275	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex	13	3	48.9	27.96	0.05046	up
28461255	NADH dehydrogenase (ubiquinone) 1 beta subcomplex	2	4	3.67	13.56	0.01051	down
62751972	Potassium voltage-gated channel shaker-related	3	0	5.86	0	1.21E-04	up
119891540	PREDICTED: glutathione S-transferase kappa 1	3	3	10.97	8.47	0.01147	up
119887606	PREDICTED: hypothetical protein	1	3	2.03	8.02	0.04382	down
119903031	PREDICTED: hypothetical protein	5	15	25.57	46.02	2.23E-04	down
119908822	PREDICTED: hypothetical protein	3	1	7.23	1.65	0.03051	up
119888977	PREDICTED: hypothetical protein	11	3	37.09	22.02	0.03232	up
119905186	PREDICTED: hypothetical protein	0	3	0	6.3	1.90E-05	down
76661674	PREDICTED: hypothetical protein	1	4	2.22	9.87	0.01595	down
119901076	PREDICTED: hypothetical protein	1	5	9.84	11.22	0.04157	down
119918378	PREDICTED: hypothetical protein	3	15	16.31	36.68	5.24E-04	down
61843441	PREDICTED: hypothetical protein	0	4	0	9.99	4.37E-05	down
119901737	PREDICTED: hypothetical protein	4	6	20.4	18.24	0.02245	down
119904572	PREDICTED: hypothetical protein	3	0	6.15	4.95	0.01124	up
119884876	PREDICTED: hypothetical protein	3	2	9.65	2.72	0.03549	up
76631114	PREDICTED: hypothetical protein	35	15	106.7	81.25	0.02224	up
119923822	PREDICTED: hypothetical protein	63	51	205.06	170.94	0.001919	up
76644873	PREDICTED: hypothetical protein isoform 2	3	0	7.78	1.83	0.03381	up
76645752	PREDICTED: hypothetical protein isoform 4	1	4	1.62	9.68	0.003268	down
119912558	PREDICTED: hypothetical protein isoform 6	4	0	10.87	0	7.37E-12	up
119893872	PREDICTED: hypothetical protein LOC535130	5	1	14.93	12.61	0.03256	up
76687954	PREDICTED: hypothetical protein, partial	3	0	7.7	1.95	0.04264	up
119925886	PREDICTED: hypothetical protein, partial	3	0	6.76	5.69	0.01623	up
119895251	PREDICTED: profilin 3	2	4	12.89	19.46	0.01716	down
119922439	PREDICTED: similar to 1700016M24Rik protein	3	0	8.17	0	1.04E-04	up
119879571	PREDICTED: similar to AAT1-alpha	22	12	69.77	50.74	0.01412	up
119912554	PREDICTED: similar to Ace protein	15	4	55.74	20.29	8.69E-06	up
61878077	PREDICTED: similar to Actin-related protein T1	7	2	16.82	7.44	0.02775	up
119928361	PREDICTED: similar to ADAM metallopeptidase	7	2	23.85	10.3	0.01155	up
119913547	PREDICTED: similar to ADAM metallopeptidase with thrombospondin type 1 motif, 17 preproprotein	5	0	12.58	0	1.24E-07	up
119903267	PREDICTED: similar to ALMS1 protein	4	0	10.2	2.89	0.02294	up
119892487	PREDICTED: similar to Ankyrin repeat domain-containing	1	3	1.88	7.82	0.03473	down

Table 1: Differentially expressed proteins. *(Continued)*

76657564	PREDICTED: similar to calmodulin	10	5	23.53	16.83	0.04331	up
119901005	PREDICTED: similar to centrosomal protein 110kD	3	0	8.38	2.06	0.03921	up
119893858	PREDICTED: similar to chromosome 13 open reading	0	3	0	6.84	4.35E-04	down
61814552	PREDICTED: similar to Cytochrome c oxidase subunit	13	10	41.99	47.17	0.01639	up
119904416	PREDICTED: similar to diaphanous homolog 3	0	3	0	7.64	3.85E-04	down
119915202	PREDICTED: similar to DNAH8, partial	12	3	33.72	16.39	0.001961	up
119927503	PREDICTED: similar to DNAH8, partial	6	1	22.2	8.36	0.04519	up
119911633	PREDICTED: similar to EF-hand calcium binding domain 5	0	3	0	7.52	3.23E-05	down
119888835	PREDICTED: similar to EPH receptor A8	4	0	9.46	1.9	0.009866	up
119895747	PREDICTED: similar to FAT tumor suppressor 2	12	3	28.56	32.84	1.15E-04	up
119919673	PREDICTED: similar to ferritin L subunit isoform	1	4	0.57	10.42	3.74E-05	down
119909426	PREDICTED: similar to fertilin alpha	12	27	63.45	89.9	0.003791	down
119919953	PREDICTED: similar to filamin	0	3	0	6.94	4.54E-04	down
76662361	PREDICTED: similar to GFHL3075 isoform 3	3	0	8.3	7.63	0.01657	up
61820991	PREDICTED: similar to GK2 protein	16	9	54.86	31.79	0.01383	up
119901324	PREDICTED: similar to HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1 isoform 1	2	3	6.09	6.74	0.01418	down
119895512	PREDICTED: similar to HIST1H4I protein	3	0	6.36	1.06	0.007699	up
119915532	PREDICTED: similar to histone H2b-616	4	10	17.41	34.08	0.005828	down
76613952	PREDICTED: similar to histone H4	2	2	7.48	5.88	0.02561	down
76642199	PREDICTED: similar to Izumo sperm-egg fusion 1	39	24	133.67	99.29	0.005314	up
119890207	PREDICTED: similar to KIAA0191	18	26	70.99	64.29	0.01024	down
119891377	PREDICTED: similar to KIAA0225 isoform 1	1	4	1.01	9.43	5.50E-04	down
119890395	PREDICTED: similar to KIAA0467 protein	0	3	0	7.91	3.74E-06	down
119902048	PREDICTED: similar to KIAA1305 protein	0	3	0	7.91	0.001712	down
119906772	PREDICTED: similar to KIAA1429 protein isoform	0	3	0	6.98	4.55E-04	down
119912552	PREDICTED: similar to KIAA1636 protein	3	0	7.44	0	1.53E-04	up
119891313	PREDICTED: similar to KIAA1793 protein	12	8	52.91	37.11	0.0228	up
119909205	PREDICTED: similar to KIAA2017 protein isoform	6	0	17.51	6.69	0.01123	up
76664109	PREDICTED: similar to LOC505732 protein	4	6	14.08	12.97	0.01832	down
76641602	PREDICTED: similar to LOC507431 protein isoform	10	7	31.93	36.41	0.03413	up
119902010	PREDICTED: similar to LOC512571 protein	47	34	189.9	155.47	0.03649	up
119905900	PREDICTED: similar to NDRG3	0	3	0	7.3	1.80E-04	down
76612380	PREDICTED: similar to nestin	0	3	0	6.85	1.25E-04	down
119911939	PREDICTED: similar to netrin-1	2	4	2.84	8.74	0.01328	down
119894490	PREDICTED: similar to obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF, partial	1	3	1.45	7.16	0.03376	down
119905455	PREDICTED: similar to Pitrilysin metallopeptidase	1	6	2.49	22.6	3.23E-05	down
76618065	PREDICTED: similar to Pou6f1 protein	4	1	15.92	11.31	0.006931	up
119894859	PREDICTED: similar to Protein KIAA1543 isoform	1	3	1.36	8.01	0.008488	down
119893105	PREDICTED: similar to protein kinase A binding protein	135	113	480.21	392.89	2.35E-05	up
76627105	PREDICTED: similar to RAB2B, member RAS oncogene	7	16	29.46	52.58	0.004367	down
119912290	PREDICTED: similar to RIKEN cDNA 4121402D02	3	0	7.29	0	2.00E-07	up
119914167	PREDICTED: similar to RIKEN cDNA A530050D06 gene	6	8	12.57	21.82	0.007372	down
119903563	PREDICTED: similar to RNA polymerase I polypept	3	6	7.15	15.59	0.03155	down
119910233	PREDICTED: similar to sca1	3	0	7.43	1.65	0.0301	up
119916698	PREDICTED: similar to Septin 12	0	3	0	8.2	2.26E-07	down
119903556	PREDICTED: similar to sulfotransferase K1	0	3	0	6.79	1.49E-04	down
119902145	PREDICTED: similar to telomerase-associated protein	3	0	8.27	0	8.45E-06	up
119914302	PREDICTED: similar to trans-Golgi p230	4	0	8.82	2.85	0.01983	up
119917225	PREDICTED: similar to TRRAP protein	2	5	2.92	11.86	0.002132	down
119917582	PREDICTED: similar to TUBA	3	1	9.42	3.39	0.03158	up
119912117	PREDICTED: similar to Tumor necrosis factor receptor	7	0	17.77	0	6.66E-12	up
119903686	PREDICTED: similar to ubiquitin specific protease 34 isoform1	1	4	2.31	9.99	0.01666	down
77736091	Prohibitin	7	3	22.59	9.08	0.03313	up
114052901	Rhabdoid tumor deletion region gene 1	10	2	31.71	15.41	0.009593	up
84000339	Sperm associated antigen 6	4	1	13.76	2.2	0.02495	up
87196516	Sperm mitochondria-associated cysteine-rich protein	3	0	9.82	3.51	0.03455	up
115495195	Tektin 1	27	19	107.5	76.3	0.0155	up
84000201	Transmembrane protein 5	3	1	6.27	6.97	0.004588	up
61888856	Triosephosphate isomerase	40	26	158.4	122.55	0.005981	up
27807143	Ubiquinol-cytochrome c reductase core protein II	7	2	33.24	16.14	0.02472	up

List of differentially expressed proteins in high fertility (HF) group spermatozoa when compared to low fertility group. (LF) spermatozoa In this table we provided the information about number of peptides, Sequest cross correlation (∑Xcorr) score and P value for each protein in high and low fertility group spermatozoa respectively.

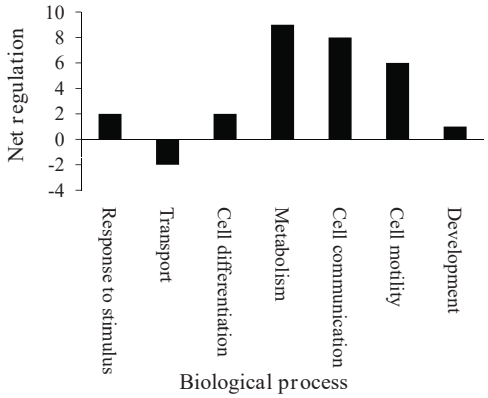

Figure 2
Overall effects in GO Slims of differentially expressed proteins of high and low fertility spermatozoa. Biological process GO annotations of all significantly altered proteins between high and low fertility spermatozoa were used to generate GO Slims. For each GO Slim, the difference in the numbers of proteins with increased expression and the number of proteins with decreased expression (relative to low fertility spermatozoa) was calculated to estimate the net regulatory effect.

metabolic pathways represented was comparable between the low (8) and high (9) fertility spermatozoa. Epidermal growth factor (EGF) signaling was the most prominent signaling pathway specific to high fertility sperm (Figure 3). EGF signaling is known to promote proliferation, survival, and differentiation of a wide variety of mammalian cells [23]. In addition to the EGF signaling pathway, platelet derived growth factor (PDGF) signaling, peroxisome proliferated activator receptor (PPAR) signaling, interleukin(IL) -4 signaling, NF-kβ signaling, chemokine signaling, and insulin growth factor (IGF)-1 signaling were identified only in high fertility spermatozoa. In low the fertility group, Cell cycle: G2/M DNA damage check point regulation was the most significant pathway followed by integrin signaling.

Proteins with significantly altered expression: molecular network and pathway analysis

Systems analysis of global proteomes revealed that some signaling pathways are differentially represented between the high and low fertility group spermatozoa. To further analyze these differentially expressed pathways, we carried out IPA analysis with just the 125 differentially expressed proteins. In high fertility spermatozoa, expression of 74 proteins was increased when compared to low fertility spermatozoa. IPA analysis identified three significant networks with scores of 22, 19, and 13 respectively. Proteins identified in the top three networks are participants in EGF signaling, PDGF signaling, oxidative phosophorylation, and pyruvate metabolism pathways. Expression of two proteins, ATP synthase, H+ transporting, mitochondrial F1 complex (ATP5B), and cytochrome c oxidase subunit III (COX3) involved in oxidative phosphorylation and casein kinase II involved in EGF signaling and PDGF signaling were higher in the high fertility spermatozoa compared to low fertility spermatozoa (Table 1). IPA also identified pyruvate metabolism as the most sig-

Table 2: Top ten functions/diseases and their respective top ten signaling pathways in high fertility group spermatozoa.

Functions & diseases	Signaling Pathways									
	EGF signaling	PDGF signaling	Integrin signaling	Amyloid processing	Complement and Coagulation cascade	PPAR signaling	Neurotrophin signaling	Huntingtons disease signaling	IGF1 Signaling	Apoptosis signaling
Cell cycle	12	14	13	10	4	11	11	13	14	10
Cellular movement	9	12	25	10	11	13	9	12	11	14
Connective tissue development and function	3	4	7	5	2	2	3	2	5	2
Cellular assembly and Organization	3	5	15	9	3	4	4	5	4	7
Cell morphology	15	17	31	10	6	10	15	16	14	12
Cardio-vascular disease	4	5	3	5	3	2	3	7	3	7
Lipid metabolism	2	2	2	2	1	2	2	3	2	2
Small molecule Biochemistry	9	10	12	12	1	6	8	9	12	12
Cell to cell signaling and interaction	6	8	22	3	13	3	5	6	7	7
Post translational modification	8	10	14	13	1	7	7	9	9	10

EGF: Epidermal Growth Factor; PDGF: Platelet Derived Growth Factor; IGF1: Insulin Growth Factor-1.

Table 3: Top ten functions/diseases and their respective top ten signaling pathways in low fertility group spermatozoa.

	Signaling Pathways									
Functions & Diseases	Cell cycle:G2/M DN Adamage check point regulation	Integrin signaling	Apoptosis signaling	MAPK signaling	Amyloid Processing	VEGF signaling	G-protein coupled receptor signaling	PTEN signaling	Actin Cyto-skeleton signaling	Axonal guidance signaling
1.Cell cycle	10	11	11	10	5	8	8	10	10	7
2.Cell morphology	3	26	13	18	7	11	18	14	31	26
3.Post translational modification	9	12	9	11	9	5	11	9	14	11
4.Cellular assembly and Organization	3	18	4	10	5	5	7	4	23	20
5.Lipid metabolism	0	2	3	5	5	3	5	2	2	4
6.Small molecule biochemistry	5	9	10	11	8	4	13	7	11	11
7.Connective tissue disorder	1	12	1	11	2	4	10	6	11	1
8.Gene Expression	8	6	8	16	6	5	16	11	6	8
9.Cell death	7	17	14	17	8	8	14	13	15	18
10.Reproductive system disease	9	5	6	9	2	7	4	8	14	12

MAPK: Mitogen Activated Protein Kinase; VEGF: Vascular Endothelial Growth Factor; PTEN: Phosphatase and tensin homolog deleted on chromosome ten.

nificant pathway in up regulated proteins of high fertility spermatozoa. In the low fertility sperm proteome, expression of 51 proteins increased when compared to high fertility spermatozoa. IPA analysis identified two significant networks in highly expressed proteins of low fertility sperm. Proteins identified in the top two significant networks are participants in integrin signaling and estrogen receptor signaling.

Discussion

Male fertility can be described as the success by spermatozoa to fertilize oocytes and of the resulting zygotes continue on through embryonic and fetal development until birth [11]. In this study we used bovine spermatozoa to study fertility as it can serve as a model for understanding human male infertility and reproductive diseases. Studying Bovine male fertility on its own merit has implications in agro-economics involving cattle industry worldwide.

A spermatozoon must reach the site of fertilization and be capacitated for successful fertilization to occur. A subsequent step is the acrosome reaction characterized by fusion of a spermatozoon outer acrosomal membrane with overlying plasma membrane [8]. The molecular mechanisms and signal transduction pathways mediating the processes of capacitation and acrosome reaction have been partially defined [8]. Bull sperm cytosolic fraction proteomic analysis showed enrichment for tyrosine kinases which are essential for phosphorylation of specific sperm proteins during capacitation [24]. The abundance of a variety of proteins from cells surrounding the sperm has been proposed to indicate male fertility [2,14,15]. Most of the studies used 2-dimensional electrophoresis (2-DE) for isolation and identification of sperm proteins [13,25-28]. To our knowledge this is the first comprehensive non-electrophoretic proteomic study of bull sperm proteome. The aim of our study was to identify proteins that were differentially expressed between high and low fertility bull spermatozoa and interrelated metabolic and signaling pathways that have a role in fertility.

We identified 125 proteins as differentially expressed in between the high and low fertility sperm even though 1518 proteins were common to both groups and about 2000 were unique to each. The reasons for this apparent discrepancy are that we took a conservative approach to the statistical analysis: only proteins identified by at least three peptides were included in the analysis for differential expression and the statistical method used in *ProtQuant* is very conservative. *ProtQuant* specifically address the issue of "missing" mass spectra that occurs in all 2-D LC MS^2 -based expression proteomics methods. No other published method (either non-isotopic or isotopic) addresses this issue. Missing mass spectra are due to the inherent limitations of the mass spectrometers, the probabilistic nature of sampling and the cutoffs used to determine "true" assignments of peptides to mass spectra [29]. *ProtQuant* is highly conservative method which is based on sum of Xcorr method itself increases the specificity of spectral counting and reduce the type I errors of differential expression. Regardless, proteins were analyzed from each of three of the areas represented in Figure 1 and differentially-expressed proteins occurred in all three (i.e.

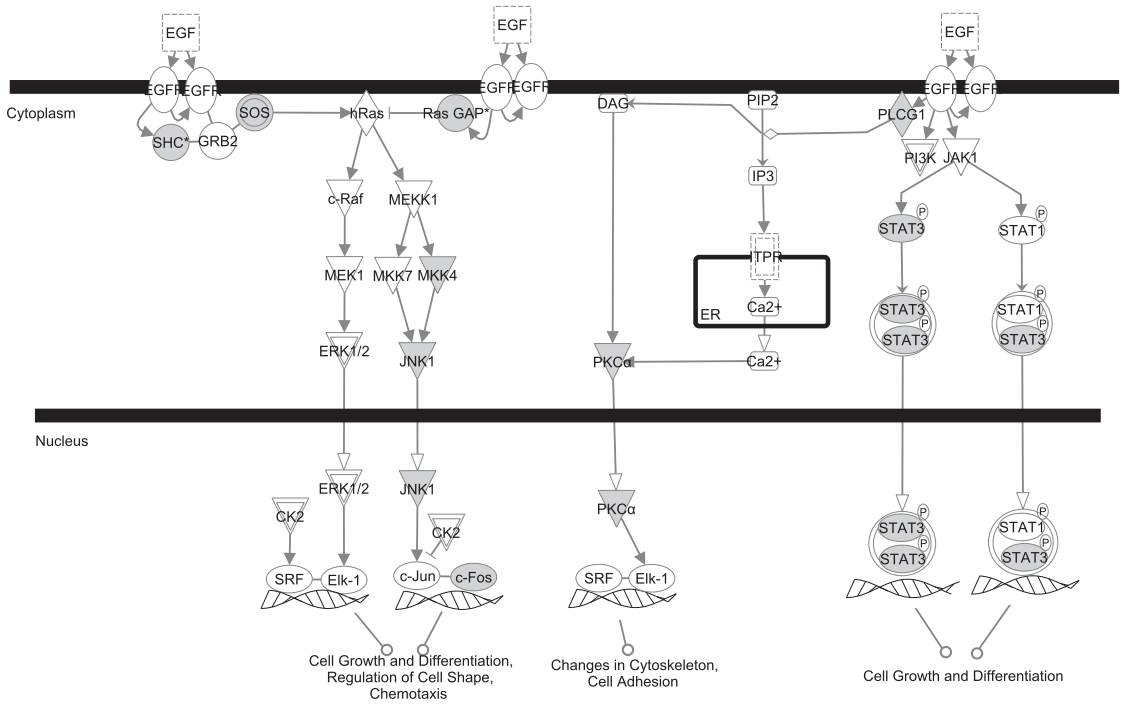

Figure 3
EGF signaling pathway generated by the Ingenuity Pathway Analysis (IPA) software. EGF and PDGF signaling pathways were the top two pathways in the top 10 functions/diseases associated with the high fertility spermatozoa (Table 2). Each node represents a protein; proteins in shaded nodes were found in the high fertility spermatozoa dataset (see additional file 1) while proteins in clear nodes were not found in the high fertility spermatozoa dataset.

proteins unique to the high and low fertility sperm as well as those common to both).

From proteome profiles of specific cells or tissues, one acquires large datasets that are inherently complex. As a result we consider it beneficial to model our bovine sperm proteome data sets using GO and IPA. From GO associations of differentially expressed proteins we found that there was a comparative up regulation of three biological processes in high fertility spermatozoa: metabolism, cell communication and cell motility (Fig 2).

Up regulation of metabolism is consistent with the fact that capacitation is coupled to a specific type of metabolism, that is glycolysis or oxidative respiration [30]. Pyruvate metabolism and glycolysis were the top most significant metabolic pathways represented in high fertility sperm proteome by IPA. In glycolysis, expression of pyruvate kinase (PKM2) was higher in high fertility spermatozoa. PKM2 catalyzes the production of pyruvate and ATP from phosphoenol pyruvate. Pyruvate formed in this process serves as an energy source for cells [31]. Impaired or lower pyruvate metabolism could limit the cell's ability to produce energy and this could be one of the reasons for reduced fertility in the low fertility group.

Expression of COX 3 and ATP5B involved in oxidative respiration was higher in high fertility spermatozoa compared to low fertility spermatozoa. COX3 is a member of the large transmembrane protein complex found in the mitochondrion and is the last protein in the electron transport chain. Coupling of electron transport to oxidative respiration maintains the high mitochondrial transmembrane potential required for mitochondrial ATP production [32]. ATP5B catalyzes the production of ATP from ADP in the presence of a proton gradient across the mitochondrial membrane and this ATP is utilized for the motility of sperm and capacitation [33].

Communication between sperm and oocyte is critical for successful fertilization. We found that there was up regulation of cell communication in the high fertility sperm proteome when compared to low fertility sperm proteome (Figure 2). To bring about cell to cell communication several signaling pathways are necessary. EGF signaling and PDGF signaling were the top two significant signaling pathways identified in high fertiliy spermatozoa. EGF and PDGF signaling pathways stimulate tyrosine phosphorylation of various MAP kinases and their upstream activators MEK1, MEK2 and MEKK [34,35]. EGF signaling has an important role in sperm capacitation as it stimulates tyrosine phosphorylation of many proteins [36]. In addition, EGF signaling also activates phospholipase C (PLC) [36] (Figure 3). PLC is important for the acrosome reaction (AR), fertilization and embryo development. PLC catalyzes the production of inositol 1, 4, 5-triphosphate (IP3) from phosphatidylinositol 4, 5-biphosphate. IP3 generated by PLC activates the extra cellular calcium influx required for the AR via binding to the IP3 receptor (IP3R) gated calcium channel located on the acrosome membrane [37]. Mutations in mouse PLCB1 reduced the AR rate, fertilization rate and embryo development [38]. EGF signaling was specific to high fertility bull sperm. Defects in EGF signaling in low fertility spermatozoa may prevent capacitation.

Expression of casein kinase 2 (CKII) prime poly peptide in EGF signaling was higher in high fertility spermatozoa compared to low fertility spermatozoa (Table 1). CKII is preferentially expressed in late stages of spermatogeneis and is involved in sperm chromatin decondensation after sperm oocyte fusion [39,40]. CKII deficient mice are infertile with oligospermia and globozoospermia[40]. EGF signaling also induces actin polymerization in bovine sperm capcitation [41]. Actin polymerization is essential for incorporation of sperm into egg cytoplasm [42] and for sperm nuclei decondensation [43].

Comparing the proteome profiles of bull sperm of high and low fertility showed some molecular features associated with low fertility. Cell cycle: G2/M DNA damage check point regulation was the topmost significant signaling pathway followed by integrin signaling in low fertility bull sperm (Table 3). The G2/M DNA damage checkpoint could help in maintaining the integrity of the genome during different stages of development. Progression through different phases of the cell cycle requires the sequential activation of various cyclin dependent kinases and these kinases in turn are regulated by integrin signaling. Integrin signals are necessary for cells to traverse the cell division cycle [44]. These two pathways may be a compensatory response for reproductive system disease function which was identified only in low fertility sperm (Table 3).

In addition to differences in signaling and metabolic pathways between high and low fertility spermatozoa, we identified differences in protein expression that had implications in sperm motility. Expression of A-kinase anchor protein-4 (AKAP4) was significantly higher in high fertility spermatozoa (Table 1). AKAP4 is a major fibrous sheath protein of the principal piece of the sperm flagellum. AKAP4 recruits Protein kinase A to the fibrous sheath and facilitates local phosphorylation to regulate flagellar function in humans [45]. It also serves as a scaffolding protein for signaling proteins and proteins involved in metabolism. Higher expression of AKAP4 in the high fertility group sperm could result in higher motility.

Conclusion

In summary, this is the first comprehensive description of the spermatozoa proteome of bovine. Comparative proteomic analysis of high fertility and low fertility bulls, in the context of protein interaction networks identified putative molecular markers associated with high fertility phenotype. We observed marked differences in signaling and metabolic pathways between high fertility and low fertility spermatozoa that have implications in sperm capacitation, acrosomal reaction and sperm-oocyte communication.

Methods

Selection of high and low fertility bulls

Frozen semen samples and bull fertility data (see additional file 2) from six mature and progeny tested Holstein bulls with satisfactory semen quality were provided by Alta Genetics (Watertown, WI).

Sample and Data Sources

The fertility data were established by a progeny testing program named Alta Advantage®, which is the industry's most reliable source of fertility information. It consisted of insemination records collected from 180 well managed partner dairy farms located in different geographical regions across the United States. This breeding program provided the advantages of DNA verification of the paternity of the offspring, and diagnosed pregnancies by veterinary palpation, instead of just relying on non-return rates 60–90 days after breeding.

Bull Fertility Prediction

To predict fertility of the bulls from the given source, a sub-set of data were generated consisting of 962,135 insemination records from 934 bulls with an average of 1,030 breedings ranging from 300 to 15,194. The environmental and herd management factors that influence fertility performance of sires were adjusted using threshold models which were similar to previously published models by Zwald et al [46,47]. Parameters estimation and fer-

tility prediction were obtained using Probit.F90 software developed by Y. M. Chang [48].

Therefore, for the definition of fertility, instead of relying only on the number of pregnant cows (verified using palpation by a veterinarian or ultrasound examination) divided by the total number of cows examined for pregnancy, we considered the outcome of each breeding event and adjusted the environmental factors such as the effects of herd-year-month, parity, cow, days in milk, sire proven status (young, proven, colored) in order to rank the bulls based on their breeding values for fertility. Further, the fertility of each bull was calculated and expressed as the percent deviation of its conception from the average conception of all bulls having at least 300 breeding in the data set.

Selection of high and low fertility bulls

For this study, we used an arbitrary threshold for classifying high and low fertility bulls. However, the bulls scoring highest and lowest fertility deviation from average with highest reliability (>1,000 breeding/bull) were selected for this study. The differences in the average fertility indexes between high and low fertility groups were 5.46% which was obtained from bulls having adequate records for higher reliability. While three bulls which were scored 5.3% above the average were considered high fertile, three bulls which were scored 10.76% below the average were defined as low fertility (see additional file 2). Two separated pools of sperm cells (3×10^8) were constituted by mixing equal amounts of sperm cells from either three low or three high fertility bulls. The experiment was replicated three times.

Isolation of pure sperm cells

Spermatozoa were collected from high and low fertility bulls and frozen in 0.25 ml straws. For each bull, the total spermatozoa collected were purified by Percoll gradient centrifugation: 90% Percoll solution in water was prepared with DL-Lactate (19 μM), $CaCl_2$ (2 μM), $NaHCO_3$ (25 mM), $MgCl_2$ (400 μM), KCl (3 μM), NaH_2PO_4 (310 μM), NaCl (2 mM) and Hepes (10 mM). 90% Percoll solution was diluted to 45% with sperm diluent medium (1 mM pyruvate, 10 mM Hepes, 0.021 mM DL-Lactate in Tyrode's salt solution, pH 7.4). A density gradient of Percoll was prepared in an Eppendorf tube (0.1 ml of 90% fraction under 1 ml of the 45% fraction). Spermatozoa were thawed at 35°C for 1 min and layered on top of the percoll gradient. The spermatozoa were pelleted by centrifugation (956 *g*; 15 min) followed by two washes in phosphate-buffered solution (PBS) (956 *g*; 5 min,). The total sperm count was obtained using an Improved Neubauer Hemacytometer and 10^8 sperm cells were aliquoted and stored at -80°C.

Protein extraction by DDF

DDF sequentially extracts proteins from different cellular compartments using a series of detergents and this off-line pre-fractionation step in sample preparation increases the proteome coverage. Another advantage of using DDF is that based on the DDF fractions from which proteins are identified, proteins can be found in different cellular locations. Proteins were isolated using DDF as previously described [18]. Cytosolic proteins were extracted by six sequential incubations in a buffer containing digitonin (10 min each); next a fraction containing predominantly membrane proteins was isolated by incubating the cells in 10% Triton X-100 buffer for 30 min and then removing the soluble protein. Nuclear DDF buffer containing deoxycholate (DOC) was then added to the remaining insoluble material and subjected to freeze-thawing to disrupt the nucleus. Nuclear proteins were collected from the resulting soluble fraction and the sample was then aspirated through an 18 g needle and treated with a mixture of DNase I (50U, Invitrogen, Carlsbad CA;) and RNase A (50 mg; Sigma-Aldrich, St Louis, MO) at 37°C for 1 h) to digest nucleic acids. Any remaining pellet, containing the least soluble proteins, was treated with a buffer containing 5% SDS.

Proteomics

Proteomic analysis was carried out with triplicate samples of spermatozoa from the high fertility group and low fertility group spermatozoa as described [19]. Proteins were precipitated with 25% tricholoroacetic acid to remove salts and detergents. Protein pellets were resuspended in 0.1 M ammonium bicarbonate with 5% HPLC grade acetonitrile (ACN), reduced (5 mM, 65°C, 5 min), alkylated (iodoacetamide, 10 mM, 30°C, 30 min) and then trypsin digested until there was no visible pellet (sequencing grade modified trypsin, Promega; 1:50 w/w 37°C, 16 h). Peptides were desalted using a peptide macrotrap (Michrom BioResources, Inc., Auburn, CA) and eluted using a 0.1% trifluoroacetic acid, 95% ACN solution. Desalted peptides were dried in a vacuum centrifuge and resuspended in 20 μL of 0.1% formic acid and 5% ACN. LC analysis was accomplished by strong cation exchange(SCX) followed by reverse phase liquid chromatography (RP-LC) coupled directly in line with an ESI ion trap mass spectrometer (LCQ Deca XP Plus; ThermoElectron Corporation; San Jose, CA). Samples were loaded into a LC gradient ion exchange system (Thermo Separations P4000 quaternary gradient pump coupled with a 0.32 × 100 mm BioBasic strong cation exchange column). A flow rate of 3 μL/min was used for both SCX and RP columns.

A salt gradient was applied in steps of 0, 10, 15, 20, 25, 30, 35, 40, 45, 50, 57, 64, 90, and 700 mM ammonium acetate in 5% ACN, 0.1% formic acid, and the resultant pep-

tides were loaded directly into the sample loop of a 0.18 × 100 mm BioBasic C18 reverse phase liquid chromatography column of a Proteome X workstation (ThermoElectron). The reverse phase gradient used 0.1% formic acid in ACN and increased the ACN concentration in a linear gradient from 5% to 30% in 20 min and then 30% to 95% in 7 min, followed by 5% for 10 min for 0, 10, 15, 25, 30, 45, 64, 90, and 700 mM salt gradient steps. For 20, 35, 40, 50 and 57 mM salt gradient steps ACN concentration was increased in a linear gradient from 5% to 40% in 65 min 95% for 15 min and 5% for 20 min.

The mass spectrometer was configured to optimize the duty cycle length with the quality of data acquired by alternating between a single full MS scan followed by three tandem MS scans on the three most intense precursor masses (as determined by Xcalibur software in real time) from the full scan. The collision energy was normalized to 35%. Dynamic mass exclusion windows were set at 2 min, and all of the spectra were measured with an overall mass/charge (m/z) ratio range of 300–1700.

All searches were done using TurboSEQUEST™ (Bioworks Browser 3.2; ThermoElectron). Mass spectra and tandem mass spectra were searched against an *in silico* trypsin-digested database of bovine RefSeq proteins downloaded from the National Center for Biotechnology Institute [NCBI; 12/26/2006; 24,853 entries]. Trypsin digestion including mass changes due to cysteine carbamidomethylation (C, 57.02 Da) and methionine mono- and di-oxidation (15.99 Da and 32 Da), was included in the search criteria. The peptide (MS precursor ion) mass tolerance was set to 1.5 Da and the fragment ion (MS^2) mass tolerance was set to 1.0 Da. Rsp Value less than 5.

As a primary filter we first limited our Sequest search output to include only peptides ≥ 6 amino acids long, with ΔCn ≥ 0.08 and Sequest cross correlation (Xcorr) scores of 1.5, 2.0 and 2.5 for +1, +2, and +3 charge states, respectively. We next used a decoy database search strategy [49] (using the same primary filter for the real database search) to calculate P values for peptide identifications as this allows us to assign the probability of a false identification based on the real data from the experiment itself [49-52]. Since the accuracy of peptide identification depends on the charge state we calculated P values for +1, +2, and +3 charge states separately. The probability that peptide identification from the original database is really a random match (P value) is estimated based on the probability that a match against the decoy database will achieve the same Xcorr [51,53]. Protein probabilities were calculated exactly as described [54,55] using only peptides with a $P < 0.05$ and only those proteins were used for further modeling. All protein identifications and their associated MS data have been submitted to the PRoteomics IDEntifications database (PRIDE ;[56]) and PRIDE accession numbers are 1883–1888.

Differential protein expression

Label free quantification approaches design to quantify relative protein abundances directly from high throughput proteomic analyses with out labeling techniques. Here, we used *ProtQuant* [29], a java based tool for label free quantification that uses a spectral counting method with increased specificity (and thus decreased false positive i.e. type I errors). This increased specificity is achieved by incorporating the quantitative aspects of the Sequest cross correlation (XCorr) into the spectral counting method. *ProtQuant* also computes the statistical significance of differential expression of control and treatment for each protein using one-way ANOVA ($\alpha \leq 0.05$). This method requires at least 3 peptides for each protein from the combination of the control and treatment before to calculate a p-value.

Gene Ontology Annotation

We used Gene Ontology (GO) resources and tools available at AgBase [57] to identify the molecular functions and biological processes represented in differentially expressed proteins in our datasets. We used *GORetriver* tool to obtain all existing GO annotations available for known proteins in our datasets. We first GO-annotated differentially expressed proteins in our datasets using existing annotations from probable orthologs with ≥90% sequence identity using the UniRef 90 database. Proteins without annotation at UniRef 90, but between 70–90% sequence identities to presumptive orthologs with GO annotation were GO-annotated using *GOanna tool* [22]. Biological process annotations for these proteins were grouped into more generalized categories using *GOSlim viewer* [22].

Modeling using Ingenuity pathway analysis

To gain insights into the biological pathways and networks that are significantly represented in our proteomic datasets we used Ingenuity Pathways Analysis (IPA; Ingenuity Systems, California). Currently IPA accepts gene/protein accession numbers from human, mouse, and rats only. Therefore, to use IPA, we mapped bovine proteins from our datasets to their corresponding human orthologs by identifying reciprocal-best-BLAST hits and uploaded these accession numbers into IPA. IPA selects "focus genes" to be used for generating biological networks. Focus genes are based on proteins from our datasets that are mapped to corresponding gene objects in the Ingenuity Pathways Knowledgebase (IPKB) and are known to interact with other genes based on published, peer reviewed content in the IPKB. Based on these interactions IPA builds networks with a size of no more than 35 genes or proteins. A P-value for each network and canon-

ical pathway is calculated according to the fit of the user's set of significant genes/proteins. IPA computes a score for each network from P-value and indicates the likelihood of the focus genes in a network being found together due to chance. We selected networks scoring ≥ 2, which have > 99% confidence of not being generated by chance [58,59].

Biological functions are assigned to each network by using annotations from scientific literature and stored in the IPKB. Fisher exact test is used to calculate the P-value determining the probability of each biological function/disease or pathway being assigned by chance. We used $P \leq 0.05$ to select highly significant biological functions and pathways represented in our proteomic datasets [58].

Authors' contributions

DP performed the proteomics sample preparation, data generation, analyzed and interpreted proteomic data, systems biology modeling and analysis and wrote the draft of the manuscript. BN developed the biomarker discovery computational tools, participated in design of this study and helped to interpret the systems biology modeling. AK and JF did sample collection and pre-proteomic sample preparation. EM facilitated sample collection, contributed to design of the study, provided expert knowledge and interpretation in reproductive biology and helped to draft the manuscript. SCB conceived of the study, participated in its design and coordination, helped analyze and interpret the statistical analysis of the proteomics data and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Proteins identified by DDF-MudPIT and their distribution in high or low fertility group spermatozoa. Column A show the GI numbers of the identified proteins, Column B indicates the corresponding protein names (assigned by NCBI). Column C shows the protein distribution in high or low fertility group spermatozoa or common to both (HF: High fertility group spermatozoa; LF: Low fertility group spermatozoa; C: common to both). For each protein we provided the information about number of peptides, Sequest cross correlation score (∑Xcorr) and DDF fraction information (DDF1, 2, 3 and 4). DDF sequentially extracts proteins from different sub cellular locations. DDF1, 2, 3, 4 corresponds to cytosolic, membrane, nuclear and cytoskeletal fractions respectively [18, 60]. We identified few proteins in more than one DDF fraction. This may be due to membrane proteins identified in all DDF fractions with increasing number of transmembrane domains in each DF fraction. Many of the proteins that function in the nucleus at some stage may be present in the cytoplasm and can thus be found in all the fractions [18].
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-2-19-S1.XLS]

Additional file 2

Fertility data of bulls whose sperm samples were used for this study. For each bull we provided the information about bull number, number of services, percent difference from average breeding rate and standard deviation. Sperm samples from three high fertility (HF) bulls were pooled as HF group, and Sperm from three low fertility (LF) were pooled as LF group.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1752-0509-2-19-S2.XLS]

Acknowledgements

This study was the result of collaboration between the Laboratories of Drs. Burgess and Memili.

This study was funded by the Life Sciences and Biology Institute & Mississippi Agricultural and Forestry Experiment Station, Mississippi State University (Manuscript number: J-11138), and by Alta Genetics, Inc. We would like to acknowledge Dr. Susan M Bridges and Bryce Magee for helping us with Biomarker discovery tool and Dr. Tibor Pechan for MS/MS analysis.

References

1. Saacke RG, Dalton JC, Nadir S, Nebel RL, Bame JH: **Relationship of seminal traits and insemination time to fertilization rate and embryo quality.** *Animal reproduction science* 2000, **60–61:**663-677.
2. Killian GJ: **High – Fertility Proteins Enhance Reproduction Rates in Dairy Cattle.** 1999.
3. Watson PF: **The causes of reduced fertility with cryopreserved semen.** *Animal reproduction science* 2000, **60–61:**481-492.
4. DeJarnette JM, Saacke RG, Bame J, Vogler CJ: **Accessory sperm: their importance to fertility and embryo quality, and attempts to alter their numbers in artificially inseminated cattle.** *Journal of animal science* 1992, **70(2):**484-491.
5. Dejarnette JM: **The effect of semen quality on reproductive efficiency.** *The Veterinary clinics of North America* 2005, **21(2):**409-418.
6. Love CC: **The sperm chromatin structure assay: a review of clinical applications.** *Animal reproduction science* 2005, **89(1–4):**39-45.
7. Evenson DP: **Loss of livestock breeding efficiency due to uncompensable sperm nuclear defects.** *Reproduction, fertility, and development* 1999, **11(1):**1-15.
8. Braundmeier AG, Miller DJ: **The search is on: finding accurate molecular markers of male fertility.** *Journal of dairy science* 2001, **84(9):**1915-1925.
9. Saacke RG, DeJarnette JM, Bame JH, Karabinus DS, Whitman SS: **Can spermatozoa with abnormal heads gain access to the ovum in artificially inseminated super- and single-ovulating cattle?** *Theriogenology* 1998, **50(1):**117-128.
10. Ballachey BE, Evenson DP, Saacke RG: **The sperm chromatin structure assay. Relationship with alternate tests of semen quality and heterospermic performance of bulls.** *Journal of andrology* 1988, **9(2):**109-115.
11. Eid LN, Lorton SP, Parrish JJ: **Paternal influence on S-phase in the first cell cycle of the bovine embryo.** *Biology of reproduction* 1994, **51(6):**1232-1237.
12. Aitken RJ: **Sperm function tests and fertility.** *International journal of andrology* 2006, **29(1):**69-75. discussion 105–108.
13. Pixton KL, Deeks ED, Flesch FM, Moseley FL, Bjorndahl L, Ashton PR, Barratt CL, Brewis IA: **Sperm proteome mapping of a patient who experienced failed fertilization at IVF reveals altered expression of at least 20 proteins compared with fertile donors: case report.** *Human reproduction (Oxford, England)* 2004, **19(6):**1438-1447.
14. Moura AA, Koc H, Chapman DA, Killian GJ: **Identification of proteins in the accessory sex gland fluid associated with fertility indexes of dairy bulls: a proteomic approach.** *Journal of andrology* 2006, **27(2):**201-211.

15. McCauley TC, Zhang H, Bellin ME, Ax RL: **Purification and characterization of fertility-associated antigen (FAA) in bovine seminal fluid.** *Molecular reproduction and development* 1999, **54(2):**145-153.
16. Gerena RL, Irikura D, Urade Y, Eguchi N, Chapman DA, Killian GJ: **Identification of a fertility-associated protein in bull seminal plasma as lipocalin-type prostaglandin D synthase.** *Biology of reproduction* 1998, **58(3):**826-833.
17. Henault MA, Killian GJ: **Effect of homologous and heterologous seminal plasma on the fertilizing ability of ejaculated bull spermatozoa assessed by penetration of zona-free bovine oocytes.** *Journal of reproduction and fertility* 1996, **108(2):**199-204.
18. McCarthy FM, Burgess SC, van den Berg BH, Koter MD, Pharr GT: **Differential detergent fractionation for non-electrophoretic eukaryote cell proteomics.** *Journal of proteome research* 2005, **4(2):**316-324.
19. McCarthy FM, Cooksey AM, Wang N, Bridges SM, Pharr GT, Burgess SC: **Modeling a whole organ using proteomics: the avian bursa of Fabricius.** *Proteomics* 2006, **6(9):**2759-2771.
20. Lubec G, Afjehi-Sadat L, Yang JW, John JP: **Searching for hypothetical proteins: theory and practice based upon original data and literature.** *Progress in neurobiology* 2005, **77(1–2):**90-127.
21. Bhattacharyya AK, Kanjilal S: **Assessment of sperm functional competence and sperm-egg interaction.** *Molecular and cellular biochemistry* 2003, **253(1–2):**255-261.
22. McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, *et al.*: **AgBase: a functional genomics resource for agriculture.** *BMC genomics* 2006, **7:**229.
23. Dikic I: **Mechanisms controlling EGF receptor endocytosis and degradation.** *Biochemical Society transactions* 2003, **31(Pt 6):**1178-1181.
24. Lalancette C, Faure RL, Leclerc P: **Identification of the proteins present in the bull sperm cytosolic fraction enriched in tyrosine kinase activity: a proteomic approach.** *Proteomics* 2006, **6(16):**4523-4540.
25. Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S: **Proteome mapping of mature pollen of Arabidopsis thaliana.** *Proteomics* 2005, **5(18):**4864-4884.
26. Hozumi A, Satouh Y, Ishibe D, Kaizu M, Konno A, Ushimaru Y, Toda T, Inaba K: **Local database and the search program for proteomic analysis of sperm proteins in the ascidian Ciona intestinalis.** *Biochemical and biophysical research communications* 2004, **319(4):**1241-1246.
27. Martinez-Heredia J, Estanyol JM, Ballesca JL, Oliva R: **Proteomic identification of human sperm proteins.** *Proteomics* 2006, **6(15):**4356-4369.
28. Tan Y, Fan L, Luo K, Zhu W, Lu G: **[Establishment of the two-dimensional gel electrophoretic protein map of the human sperm head].** *Zhonghua nan ke xue = National journal of andrology* 2004, **10(12):**886-889.
29. Bridges SM, Bryce GB, Wang N, Williams WP, Burgess SC, Nanduri B: **ProtQuant: a tool for the label-free quantification of mudPIT proteomics data.** *BMC Bioinformatics* 2007, **8(Supple 7):**S24.
30. Ferrandi B, Lange Consiglio A, Chiara F, Uber E, Marchini M, Baglioni A, Carnevali A, Cremonesi F, Porcelli F: **Cytochemical study on human spermatozoa metabolism during in vitro capacitation.** *Andrologia* 1987, **19(Spec No):**278-283.
31. Pithukpakorn M: **Disorders of pyruvate metabolism and the tricarboxylic acid cycle.** *Molecular genetics and metabolism* 2005, **85(4):**243-246.
32. Guthrie HD, Welch GR: **Determination of intracellular reactive oxygen species and high mitochondrial membrane potential in Percoll-treated viable boar sperm using fluorescence-activated flow cytometry.** *Journal of animal science* 2006, **84(8):**2089-2100.
33. Peterson RN, Freund M: **ATP synthesis and oxidative metabolism in human spermatozoa.** *Biology of reproduction* 1970, **3(1):**47-54.
34. Hunter T: **When is a lipid kinase not a lipid kinase? When it is a protein kinase.** *Cell* 1995, **83(1):**1-4.
35. Payne DM, Rossomando AJ, Martino P, Erickson AK, Her JH, Shabanowitz J, Hunt DF, Weber MJ, Sturgill TW: **Identification of the regulatory phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase).** *The EMBO journal* 1991, **10(4):**885-892.
36. Breitbart H, Naor Z: **Protein kinases in mammalian sperm capacitation and the acrosome reaction.** *Reviews of reproduction* 1999, **4(3):**151-159.
37. Walensky LD, Snyder SH: **Inositol 1,4,5-trisphosphate receptors selectively localized to the acrosomes of mammalian sperm.** *The Journal of cell biology* 1995, **130(4):**857-869.
38. Choi D, Lee E, Hwang S, Jun K, Kim D, Yoon BK, Shin HS, Lee JH: **The biological significance of phospholipase C beta 1 gene mutation in mouse sperm in the acrosome reaction, fertilization, and embryo development.** *Journal of assisted reproduction and genetics* 2001, **18(5):**305-310.
39. Mudgal P, Anand SR: **Casein kinase II activity of buffalo sperm chromatin.** *Molecular reproduction and development* 1998, **50(2):**178-184.
40. Xu X, Toselli PA, Russell LD, Seldin DC: **Globozoospermia in mice lacking the casein kinase II alpha' catalytic subunit.** *Nature genetics* 1999, **23(1):**118-121.
41. Breitbart H, Cohen G, Rubinstein S: **Role of actin cytoskeleton in mammalian sperm capacitation and the acrosome reaction.** *Reproduction (Cambridge, England)* 2005, **129(3):**263-268.
42. Sanchez-Gutierrez M, Contreras RG, Mujica A: **Cytochalasin-D retards sperm incorporation deep into the egg cytoplasm but not membrane fusion with the egg plasma membrane.** *Molecular reproduction and development* 2002, **63(4):**518-528.
43. Kumakiri J, Oda S, Kinoshita K, Miyazaki S: **Involvement of Rho family G protein in the cell signaling for sperm incorporation during fertilization of mouse eggs: inhibition by Clostridium difficile toxin B.** *Developmental biology* 2003, **260(2):**522-535.
44. Giancotti FG, Ruoslahti E: **Integrin signaling.** *Science (New York, NY)* 1999, **285(5430):**1028-1032.
45. Huang Z, Somanath PR, Chakrabarti R, Eddy EM, Vijayaraghavan S: **Changes in intracellular distribution and activity of protein phosphatase PP1gamma2 and its regulating proteins in spermatozoa lacking AKAP4.** *Biology of reproduction* 2005, **72(2):**384-392.
46. Zwald NR, Weigel KA, Chang YM, Welper RD, Clay JS: **Genetic selection for health traits using producer-recorded data. II. Genetic correlations, disease probabilities, and relationships with existing traits.** *Journal of dairy science* 2004, **87(12):**4295-4302.
47. Zwald NR, Weigel KA, Chang YM, Welper RD, Clay JS: **Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values.** *Journal of dairy science* 2004, **87(12):**4287-4294.
48. Chang YM, Gianola D, Heringstad B, Klemetsdal G: **Effects of trait definition on genetic parameter estimates and sire evaluation for clinical mastitis with threshold models.** *Animal science* 2004, **79:**355-364.
49. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.** *Nature methods* 2007, **4(3):**207-214.
50. Elias JE, Gibbons FD, King OD, Roth FP, Gygi SP: **Intensity-based protein identification by machine learning from a library of tandem mass spectra.** *Nature biotechnology* 2004, **22(2):**214-219.
51. Park GW, Kwon KH, Kim JY, Lee JH, Yun SH, Kim SI, Park YM, Cho SY, Paik YK, Yoo JS: **Human plasma proteome analysis by reversed sequence database search and molecular weight correlation based on a bacterial proteome analysis.** *Proteomics* 2006, **6(4):**1121-1132.
52. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP: **Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome.** *Journal of proteome research* 2003, **2(1):**43-50.
53. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG 2nd, Smith RD: **Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome.** *Journal of proteome research* 2005, **4(1):**53-62.
54. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Analytical chemistry* 2003, **75(17):**4646-4658.
55. MacCoss MJ, Wu CC, Yates JR 3rd: **Probability-based validation of protein identifications using a modified SEQUEST algorithm.** *Analytical chemistry* 2002, **74(21):**5593-5599.

56. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, Gevaert K, Vandekerckhove J, Apweiler R: **PRIDE: the proteomics identifications database.** *Proteomics* 2005, **5(13):**3537-3545.
57. McCarthy FM, Bridges SM, Wang N, Magee GB, Williams WP, Luthe DS, Burgess SC: **AgBase: a unified resource for functional analysis in agriculture.** *Nucleic acids research* 2007:D599-603.
58. Gerling IC, Singh S, Lenchik NI, Marshall DR, Wu J: **New data analysis and mining approaches identify unique proteome and transcriptome markers of susceptibility to autoimmune diabetes.** *Mol Cell Proteomics* 2006, **5(2):**293-305.
59. Huang Y, Yan J, Lubet R, Kensler TW, Sutter TR: **Identification of novel transcriptional networks in response to treatment with the anticarcinogen 3H-1,2-dithiole-3-thione.** *Physiological genomics* 2006, **24(2):**144-153.
60. Fazal MA, Palmer VR, Dovichi NJ: **Analysis of differential detergent fractions of an AtT-20 cellular homogenate using one- and two-dimensional capillary electrophoresis.** *Journal of chromatography* 2006, **1130(2):**182-189.

CHAPTER V

CONCLUSIONS

Fertilization is a well-orchestrated, stepwise process during which the participating male and female gametes undergo irreversible changes, losing some of their structural components while contributing others to the resultant zygote. As the use of immature oocytes is becoming an option in clinical *in-vitro* fertilization (IVF), understanding the factors that regulate oocyte maturation and developmental competence is likely to help in the optimization of the *in-vitro* maturation (IVM) conditions. The oocyte and its surrounding cumulus cells (CC) are constantly growing and differentiating in a mutually dependent manner throughout the course of folliculogenesis. This is necessary to promote oocyte nuclear and cytoplasmic maturation and hence the acquisition of developmental competence. Oocyte IVM is thus an emergent process and *in-vitro* techniques that even subtly alter this process can affect post-IVM oocyte developmental potential. Over the past decade, science has gained significant new insights into the oocyte developmental biology by examining the extremely intricate cumulus-oocyte complex as a complete functional and dynamic unit. Although studies emanating from multiple laboratories is leading to a better understanding of oocyte and cumulus cell communication, proteins and pathways involved in this process remained to be characterized at the beginning of my work. A spermatozoon must reach the site of fertilization and be capacitated for successful fertilization to occur. A subsequent step is

the acrosome reaction, characterized by fusion of the outer acrosomal membrane of spermatozoon with the overlying plasma membrane. The molecular mechanisms and signal transduction pathways mediating the processes of capacitation and acrosome reaction have been partially defined. At present, the molecular nature of sperm fertility defects or biomarkers for accurate fertility prediction is not known.

Together, the work in this doctoral dissertation provides the first comprehensive proteome analysis of the bovine germinal vesicle (GV) oocyte, its surrounding CCs, and sperm. The work in chapter 2, my first aim, established methods that provide a foundation for high-throughput proteomics approaches of bovine oocyte and cumulus cell biology, and allowed me to model the intricate cross communication between oocyte and cumulus cells using computational systems biology approaches. The work in chapter 3, my second aim, provided more comprehensive information on the biological processes associated with bovine GV oocyte and cumulus cell function. Not only does this improve data quality and allow improved functional modeling, but we almost doubled the GO Annotation Quality (GAQ) score (a quantitative measure of the GO annotation) of the proteins in our dataset. By adding manual literature biocuration, this study also captured the quantitative positive and negative regulatory roles that these GV oocyte proteins play in the various biological processes. Overall this study could identify and quantify proteins agonistic or antagonistic to biological processes including signal transduction, cell cycle regulation, DNA transcription, apoptosis regulation, protein metabolism and modification, generation of precursor metabolites and energy, cytoskeleton organization and biogenesis, and response to stress in GV oocyte, all of which may have putative roles in oocyte competence and maturation. This type of quantitative modeling is not possible

with the available pathway analysis tools. Furthermore, I analyzed the protein datasets identified in this experiment in the context of protein interaction networks and pathways using Ingenuity Pathway Analysis. Proteomics based systems biology modeling of oocytes and cumulus cells in the context of GO and canonical protein interaction networks identified ninty-one proteins involved in the cell-to-cell signaling biological process. Most of the proteins involved in the cell-to-cell signaling biological process are the components of integrin, actin cytoskeleton, mitogen activated protein kinase (MAPK) and Phosphoinositide 3-kinase (PI3K) signaling pathways which have implications in various reproductive processes such as oocyte development and maturation. This new perspective on studying oocyte-CC interactions at the proteomic level improves our knowledge of the processes regulating oocyte quality, providing new options for the treatment of female infertility. A systems biology model of potential interacting bovine oocyte and cumulus cell proteomes at the GV stage not only provides a foundation for understanding signaling and cell physiology at the GV stage of oocyte development, but is also valuable for comparative studies of other stages of oocyte development at the molecular level. The proteomes and biological processes identified in this study can also serve as reference points for further comparative studies on immature and abnormal oocytes and cumulus cells, elucidating the underlying molecular mechanisms involved in normal and pathological oocyte-cumulus cell communication. Furthermore, some of the proteins involved in cell-to-cell signaling may have value as molecular biomarkers which could be useful for assessing oocyte quality.

The last part of this dissertation describes the first non-electrophoretic proteomic study of bull sperm proteome of varying fertility rates. This study identified 125 proteins

differentially expressed between high and low fertility bull spermatozoa**.** Next, using quantitative systems biology modeling and canonical protein interaction pathways and networks, this study showed that high fertility spermatozoa differ from low fertility spermatozoa in four main processes: energy metabolism, cell communication, spermatogenesis, and cell motility. This study also identified marked differences in signaling and metabolic pathways between high fertility and low fertility sperm which have implications in sperm capacitation, acrosomal reaction, and sperm-oocyte communication. For example, epidermal growth factor (EGF) signaling and platelet derived growth factor (PDGF) signaling were significantly represented in high fertility sperm only. These two pathways had an important role in bovine sperm capacitation as they stimulate tyrosine phosphorylation of various proteins. My results also suggest that low fertility sperm DNA integrity may be compromised because cell cycle: G2/M DNA damage checkpoint regulation was the most significant signaling pathway identified in low fertility spermatozoa.

The work in this dissertation represents the foundational stages fundamental to an iterative process of hypothesis development, testing, and refinement that will ultimately yield a global understanding of mechanisms of oocyte and spermatozoon development. My research has also established the foundations necessary for further structural and functional annotation of the bovine genome aimed at identifying markers for developmental competency that are essential for selecting oocytes and spermatozoa for mammalian reproduction. The outcome of the proposed research is highly relevant impacting different areas of US agriculture, including embryos, clones, and transgenic animal production. Production of these animals with increased efficiencies will provide

significant economic benefits both to consumers and producers. As proteomic profiles can only show one step in cell phenotype and function control, transcriptomic analysis could complement this study by providing a more complete picture of the regulation of oocyte and spermatozoa development. GO information for a proportion of the identified proteins in this dissertation is still incomplete due to lack of available literature and orthologs. Thus, for some of the proteins cellular component annotation is known, but the biological process and/or its molecular function is not documented. Further improvement in the functional annotation of the bovine genome may greatly facilitate global proteomic and gene expression studies in the bovine species. Future studies should involve validation of the differentially expressed proteins identified in the bull fertility study, and oocyte-CC communication study at both transcription level and translation level using reverse-transcriptase PCR (RT-PCR) and western blotting techniques. Molecular biomarkers identified in this dissertation can also be evaluated by a comparative functional genomics studies for other species including pigs and humans. Further studies analyzing protein patterns throughout the bovine oocyte development (MI, and MII) will aid in identification of proteins that have an important role in oocyte competence and maturation. The research in this dissertation complements the transcription analyses, and together the two methods may be used in the future for systems biology modeling of early mammalian development.