Inter - Species Analysis and Likely Functions of Sperm RNA

Stefanie Elvira Nadj

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Medicine
Leeds Institute of Cardiovascular and Metabolic Medicine *(*LICAMM*)*

November 2016

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to thank my supervisor Dr. David Miller for the opportunity to work on this project, his guidance, support, ideas and incredible patience throughout this project and for proofreading this manuscript. Many thanks to Dr. John Huntriss for all his helpful input into this project as my co-supervisor, Dr. David Iles for his bioinformatical guidance and help, valuable approaches and especially patience and Riitta Partanen for giving me helpful advices and a hand at all immunocytochemistry, who also organised the cells needed and the good company she provided. I would like to thank all former and current members of the Division of Reproduction and Early Development I worked with, for providing me needed tissue, primers sequences and for teaching me all necessary skills for my future. These thanks go especially to Dr. Karen Forbes, Dr. Nadir Ciray, Dr. Paul McKeegan and Dr. Niamh Forde for their emotional support, good shared memories and constructive input throughout this work and my time in Leeds.

I am very grateful of all the lovely people I had the chance to meet in this institute and who had always an open ear for everything and were helpful throughout. Herby, I would like to thank Kay White and the core technicians who made my life easier every day, including Prof. Dr. David J. Beech and his group for giving me this incredible chance.

I had the honour to work together with some incredible scientists. This includes Prof. Dr. Stephen Krawetz, who gave me the opportunity to learn new techniques in his laboratory in Detroit and the sequencing Unit at St James's in Leeds, Dr. Ian Carr, Catherine Daly and Dr. Sally Fairweather who gave me heaps of technical advice and made it possible to perform good quality science.

I would like to express thanks to the European Union, which supported this study financially by the ITN Reprotrain Network, along with all Reprotrain Fellows I had the honour to meet.

I could not have made it without my friends and the ECG who gave me a fantastic time in the UK. Special thanks to Arron who proofread parts of this manuscript.

A huge thank you goes also to Dr. Gerd Scherer, who supported my early career and always belief in my good working abilities and thought me early to perform good quality science and supported my critical thinking.

I am deeply grateful, honoured and feel incredibly loved by all the support, strength, motivation, encouragement, good times and belief that Steven Quayle has given me and not to forget for all the proofreading he has undertaken for this manuscript.

Finally I would like to thank my family, especially my sister and mother who never lost their faith in my abilities and gave me this wonderful opportunity, who strengthened my future and always believed in me.

This work is dedicated to my family who always believed in my abilities, taught me to be curious and gave me their endless love and support throughout.

Abstract

The spermatozoon is a highly complex cell which was thought to be transcriptionally inactive and solely delivers the paternal genome, centrioles and proteins to the oocyte. During the past few years, a large number of different sperm RNA species have been discovered, however, the functionality and the gene expression networks they may be involved remain unclear to date. Different studies showed that sperm RNAs provide a historical record of gene activity during spermatogenesis and that they are delivered to the oocyte. The studies reported herein had the following foci: Firstly, the spermatozoal transcriptomes of bovine, ovine, porcine and human were compared to their respective testis transcriptomes. Secondly, the spermatozoal transcriptome of each species was analysed and compared to identify shared gene expression networks that may be indicative of common functionality in the processes of spermatogenesis, fertilisation and reproduction. The results revealed common functional pathways, indicating a possible post-fertilisation role in the developing embryo. In addition, bioinformatic analysis revealed 23 mutually shared transcripts. To further characterise the potential transfer to the oocyte and to investigate the potential function and stability of these paternal transcripts, 16 transcripts were selected and followed in the developing bovine embryo. No clear potential functions for spermatozoal RNAs post-fertilisation and during embryo development could be derived from these experimental data.

Assisted reproductive processes and the breeding industry rely on freezing spermatozoa and choosing the best candidate spermatozoon. Therefore as an additional focus, RNA profiles of frozen and fresh spermatozoa from the same donor were analysed and did not reveal major changes in the spermatozoal transcriptome caused by cryopreservation processes. Additionally as a last focus, the transcriptome of hyaluronic acid (HA) selected spermatozoa was compared to unselected spermatozoa and revealed no significant differences in RNA expression. However, a trend towards an increased expression of *MOSPD3* was observed and investigated as a potential fertility marker. Expression of MOSPD3 protein was seen to be upregulated in motile spermatozoa compared to less motile by both Western Blot analysis and immunocytochemistry. The present data supports the suggestion that sperm binding to HA may represent more viable populations and that MOSPD3 is a potential marker of spermatozoa viability that could be developed into a diagnostic tool.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

18S rRNA	18S ribosomal RNA
28S rRNA	28S ribosomal RNA
A	Adenine
APS	Ammonium Persulfate
ART	Assisted Reproduction Technology
ATP	Adenosine Triphosphate
BSA	Bovine Serum Albumin
bp	Base pairs
B. taurus	*Bos taurus*
$C_{24}H_{39}NaO_4$	Sodium Deoxycholate
C	Celsius
C	Cytosine
Ca^{2+}	Calcium Ion
cDNA	Complementary DNA
CC	Cumulus Cells
$CaCl_2$	Calcium Chloride
CHAPS	3-[(3-Cholamidopropyl)Dimethylammonio]-1-Propanesulfonate
Ct	Threshold Cycle
CO_2	Carbon Dioxide
COC	Cumulus-Oocyte Complex
Da	Dalton
DAPI	4,6-Diamidino-2-Phenylindole
DAVID	The Database for Annotation, Visualization and Integrated Discovery
ddH_2O	Double-Distilled Water
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
Dnase	Deoxyribonuclease
dNTPs	deoxynucleotide Triphosphates
DMSO	Dimethyl Sulfoxide
dH_2O	Distilled Water
DTT	Dithiothreitol
ds	Double stranded
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor

ET water	Embryo Tested Water
EtOH	Ethanol
FCS	Foetal Calf Serum
FIM	Follicle Isolation Medium
FSH	Follicle Stimulating Hormone
g	Grams
x g	Times Gravity
GAPDH	*Glyceraldehyde 3-phosphate dehydrogenase*
gDNA	Genomic DNA
h	Hour
H_2O	Water
HA	Hyaluronic Acid
HCl	Hydrochloric acid
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HM	Holding Media
HSOF	Hepes-SOF
H. sapiens	*Homo sapiens*
ICSI	Intracytoplasmatic Sperm Injection
IVF	*In Vitro* Fertilisation
IVM	*In Vitro* Maturation
k	Kilos
kb	Kilobase pairs
KCl	Potassium Chloride
l	Litres
LH	Luteinising Hormone
lncRNA	Long non-coding RNA
m	Metres
M	Molar
MII	Metaphase II
mol	Mole
MEM	Minimum Essential Medium
mg	Milligrams
$MgCl_2$	Magnesium Chloride
min	Minutes
ml	Millilitres
mm	Millimetres
Mole	~ 6.023×10^{23}
mRNA	Messenger RNA

mtRNA	mitochondrial RNA
n	Nano
N	Nitrogen
NaCl	Sodium Chloride
NaH_2PO_4	Monosodium Phosphate
$NaHCO_3$	Sodium Bicarbonate
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NP-40	Nonyl Phenoxypolyethoxylethanol 40
OCC	Oocyte-Cumulus Complex
O_2	Oxygen
O. aries	*Ovis aries*
PA	Percoll™ Additives
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
Pen/Strep	Penicillin/Streptomycin
PMSF	Phenylmethylsulfonyl Fluoride
PRM 1	*Protamine 1*
PRM 2	*Protamine 2*
qPCR	Quantitative Polymerase Chain Reaction
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNase	Ribonuclease
RT	Room Temperature
RT-PCR	Reverse Transcription-PCR
s	Seconds
SDS	Sodium dodecyl sulfate
snc	Small non-coding
SOF	Synthetic Oviduct Fluid
ssDNA	Single stranded DNA
S. scrofa	*Sus scrofa*
T	Thymine
Ta	Annealing temperature
TALP	Tyrode's Albumin Lactate Phosphate
Taq	*Thermus aquaticus*

TBE	Tris-Borate-Ethylenediaminetetraacetic Buffer
TEMED	Tetramethylethylenediamine
T_m	Melting Temperature
Tris	Trizma Base
U	Units
UCSC	University of California, Santa Cruz
V	Volts
v/v	Volume per volume
w/v	Weight per Volume
µ	Micro

Chapter 1: General Introduction

A spermatozoon is a highly differentiated and complex cell, with the single function of providing the oocyte with DNA, paternal centrioles, proteins and RNA (García-Herrero *et al.*, 2010; Lalancette *et al.*, 2008b; Ostermeier *et al.*, 2004). Bhargava *et al.* (1959) and Pessot *et al.* (1989) first described the presence of RNAs in bovine spermatozoa, rat epididymal spermatozoa and human ejaculated spermatozoa in 1989. Spermatozoal transcripts match those from the testis and can be considered as untranslated stored remnants of spermatogenesis and might reflect past events of spermatogenesis (Moldenhauer *et al.*, 2003; Miller *et al.*, 1999; Lambard *et al.*, 2004). It is widely accepted that spermatozoa are considered to be 'silent cells' and are transcriptionally and translationally inactive (Gilbert *et al.*, 2007; Miller and Ostermeier, 2006a). Some evidence of transcriptional and translational activities in human spermatozoa during capacitation and acrosome reaction have been described which challenge the understanding to date (Naz, 1998; Gur and Breitbart, 2006). Although other groups have indicated that the translational activity is most likely bacterial in origin, raising the question of whether spermatozoal RNA and protein can be synthesised *de novo* (Premkumar and Bhargava, 1972; Abraham and Bhargava, 1963; Díez-Sánchez et al., 2003b). Since this first RNA discovery in spermatozoa, different populations of RNA have been found within the spermatozoon and it is assumed that these RNAs are delivered to the zygote alongside the paternal genome (Gilbert *et al.*, 2007; Ostermeier *et al.*, 2004; Kempisty *et al.*, 2008; Anderson, 2013). Despite the fact that the role and function of those RNAs remain mainly unknown, Ostermeier *et al.* (2004) and Liu *et al.* (2012) observed that spermatozoal mRNA seems to be required from the first embryonic cleavage until the activation of the embryonic genome. Some RNAs have been found only in the spermatozoon and the zygote and not in the oocyte, providing evidence for a unique paternal contribution (Ostermeier *et al.*, 2004; Kempisty *et al.*, 2008; Yao *et al.*, 2010). In addition, recent studies revealed that paternal sncRNAs can affect phenotypic traits in the resulting offspring (Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Sharma *et al.*, 2016; Chen *et al.*, 2016). Spermatozoal examination, especially of RNA functional profiles and the understanding of the transcripts' significance may be used as a simple diagnostic tool, alongside semen and therefore may provide a unique source for the discovery of male fertility biomarkers (Lambard *et al.*, 2004; Yatsenko *et al.*, 2006; Platts *et al.*, 2007; Garrido *et al.*, 2009; García-Herrero *et al.*, 2011; Bonache *et al.*, 2012; Malcher *et al.*, 2013; Johnson *et al.*, 2015; Lima-Souza *et al.*, 2012). Successful fertilisation and embryo development is dependent on the quality of the

paternal and maternal gamete. An examination at a molecular level may simplify the diagnosis of infertility and prove useful in the comparison of the mechanisms which result in function or dysfunction in spermatozoa and fertility (Miller *et al.*, 1999; Miller and Ostermeier, 2006b; Moldenhauer *et al.*, 2003).

1.1 Spermatogenesis

Spermatogenesis is the maturation process that spermatogonial stem cells undergo in order to be functional spermatozoa and be able to fertilise the oocyte and contribute the paternal genome to the zygote. This process occurs in the seminiferous tubules located in the testis (Eddy and O'Brien, 1994). During the early development of the male foetus, Sertoli cells associate with germ cells to form testis cords, while Sertoli cells accumulate around the germ cells forming the adluminal and the basal compartment (Yazama, 2008). The formation of tight junctions between Sertoli cells produces the blood-testis barrier, which protects the adluminal compartment against toxic substances and antibodies present, and maintains the germ cells in an immune privileged microenvironment (Mital *et al.*, 2011; Mruk and Cheng, 2004; Singh *et al.*, 2003). The basal compartment stores spermatocytes and reserve spermatogonia which are generated throughout the adult life time (Mital *et al.*, 2011). Spermatogenesis is a process organised into four developmental phases (Figure 1-1): spermatocytogenesis (proliferative), spermatidogenesis (meiotic), spermiogenic (condensation) and spermiation (Figure 1-1). During the first phase spermatogonial stem cells undergo a clonal expansion, which is also a self-renewal process through continuous mitosis to maintain the pool of germ cells. This spermatogonial population is known as type A (de Rooij, 2001). A second spermatogonial population, known as type B undergo a differentiation process, to produce spermatozoa and to allow constant repopulation of the testis with developing germ cells. Type B spermatogonial (2n:2c) function is spermatocytogenesis where they undergo a single round of DNA replication (2n:4c), and differentiate into primary spermatocytes (2n:4c). This is followed by the first meiotic division (leptotene, zygotene and pachytene stages) to become haploid secondary spermatocytes (1n:2c). It is noteworthy that a significant increase of transcription and translation of early-pachytene spermatocytes is observed and a dramatic decrease in mid-pachytene spermatocytes (Figure 1-1), storing all the RNAs that the mature spermatozoa will retain (Heller and Clermont, 1963; Berndston, 1977; Johnson, 1986; Braun, 2000; Grootegoed *et al.*, 2000). Additionally a second meiotic division takes place, during which, the spermatocytes transform into round spermatids (1n:1c). After meiosis the spermatozoa enter the adluminal compartment (Anderson, 2013; Braun, 2000; Hamatani, 2012; Holstein *et al.*, 2003). During these stages, spermatids become

transcriptionally inactive Figure 1-1). Upon entering the next phase, spermiogenesis (post-meiotic) phase, spermatids undergo chromatin condensation and a final cell differentiation to complete chromatin reorganization, where the nucleus becomes condensed, histones are predominantly replaced by sperm-specific protamines, and the transcriptional process is essentially shut down (Figure 1-1) (Kretser, 1989; Sutovsky, 2003; Jodar *et al.*, 2016). In this phase, the acrosome is synthesised from the Golgi apparatus and the flagellum extended by elongation. Remaining cytoplasm, containing organelles which are not integrated into the spermatozoon form the residual body which is phagocytosed by Sertoli cells (Anderson, 2013). In the last phase (spermiation), mature spermatozoa are released into the epididymis where they pass their way through to the caudal epididymis and remain temporarily stored until ejaculation (O'Donnell *et al.*, 2011). The process of spermatogenesis in an adult human takes approximately 64 days, while the epididymal transit takes an additional 10 days, and is highly sensitive to fluctuations in the intra testicular environment (Heller and Clermont, 1963; Bujan, 1998; Stutz *et al.*, 2004).

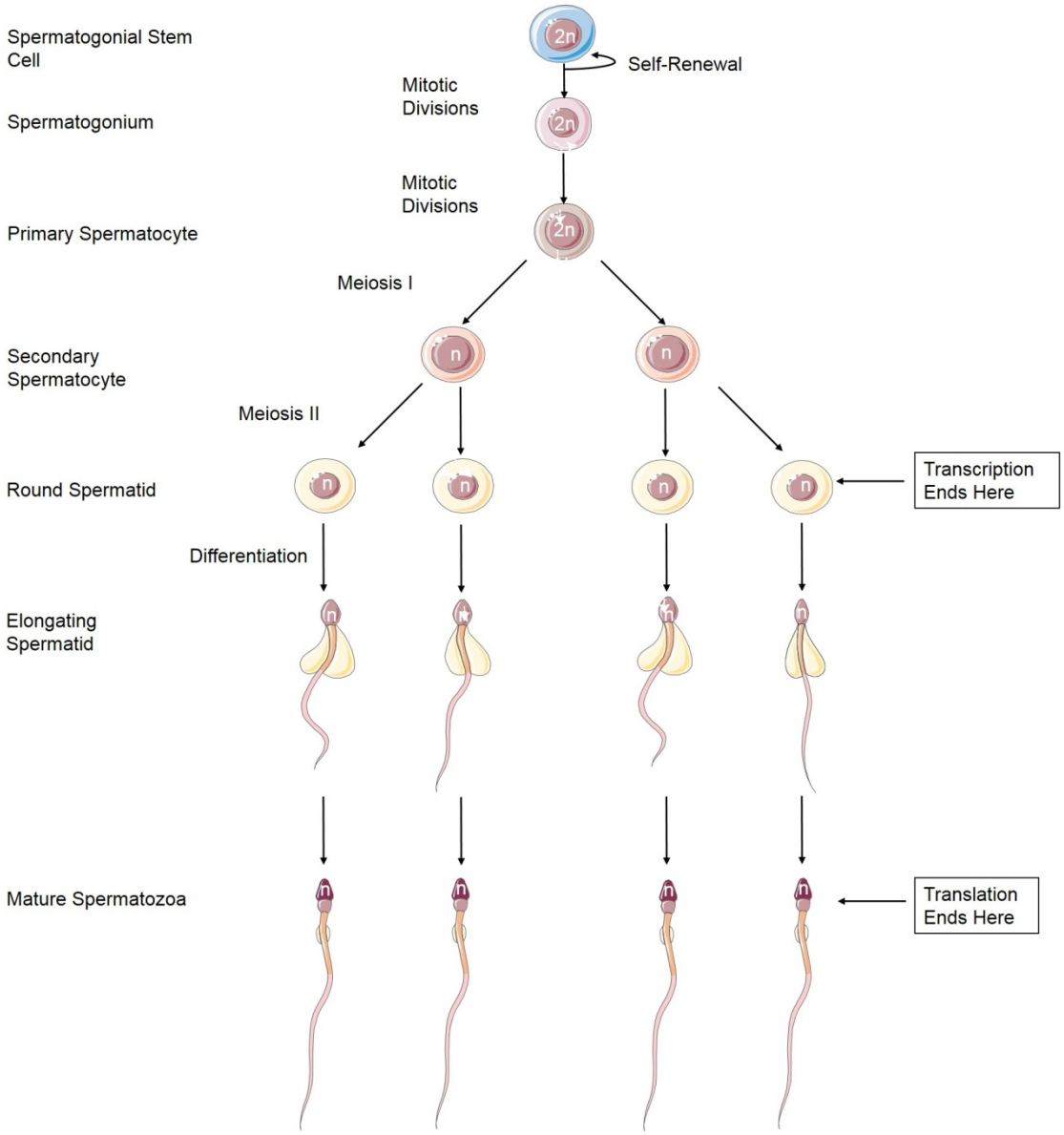


Figure 1-1: The process of spermatogenesis. Spermatozoa production from spermatogonial stem cells (2n) over mitosis of spermatogonia, meiotic division of spermatocytes (n) forming spermatids (n), which differentiate to round spermatids (n) and form elongated spermatids (n) in the seminiferouse tubules in the testis. Spermatozoa (n) mature afterwards in the epididymis. Picture adapted from Anderson (2013), images used from “Servier Medical Art Powerpoint Image Bank”

1.2 Molecular regulation in spermatozoa

1.2.1 Gene expression in spermatozoa

The spermatozoon is a highly differentiated cell type with minimal cytoplasm and a compact nucleus, designed to deliver the paternal genome to the oocyte (Miller and Ostermeier, 2006b). In contrast to former misconceptions that a spermatozoon only delivers the paternal genome and proteins and centrioles, it has more recently been shown that a spermatozoon also delivers RNAs to the oocyte, (García-Herrero *et al.*, 2010; Lalancette *et al.*, 2008b; Dias and Ressler, 2014). Bhargava *et al.* (1959) first described the existence of RNA in bovine spermatozoa in 1959 whereas the first mRNA for mature human spermatozoa was found for *c-MYC* by Kumar *et al.* (1993) in 1993. A single human spermatozoon contains ~12.5-50 fg, bovine ~10 fg and porcine ~5 fg total RNAs (Goodrich *et al.*, 2013; Card *et al.*, 2013; Jodar *et al.*, 2013) and smaller quantities of small RNAs (Hamatani, 2012; Card *et al.*, 2013; Krawetz *et al.*, 2011). Jodar *et al.* (2013) described the composition of the spermatozoal RNA populations with the most abundant being ribosomal RNA (rRNA), followed by mitochondrial RNA (mtRNAs), annotated coding messenger RNA (mRNAs), small non-coding (snc)RNAs, and then intronic retained elements. Mature sperm also display transcript variants that seem to differ in a variety of ways from those found in whole testis (Freiman, 2009; Jodar *et al.*, 2013; Das *et al.*, 2013). Therefore, transcript variants may likely occur only in the final transcriptionally active stages of spermatogenesis (Figure 1-1) (Jodar *et al.*, 2013). One quarter of the spermatozoal transcripts found reveal alternative sites of polyadenylation, which retain the integrity of the coding region but contain an abbreviated 3′ untranslated region (UTR) (Jodar *et al.*, 2013). Di Giammartino *et al.* (2011) suggest a model where modification may impact translation and affect the ability of different regulatory proteins and miRNAs to bind to the alternative UTR. Recent spermatozoal RNA-sequencing data showed many abundant transcripts, which are either not observed or less abundant in either somatic cells or testes, being an indicator for unique spermatozoal transcripts (Jodar *et al.*, 2015; Jodar *et al.*, 2013; Krawetz *et al.*, 2011; Georgiadis *et al.*, 2015; Das *et al.*, 2013). Spermatozoal RNA quantity is far lower than in any other cells types, which contain around ~50 fg of long RNAs (>200 nt) and 0.3 fg of snc-RNAs in human in contrast to 100-200 times more in somatic cells (Goodrich *et al.*, 2013; Jodar *et al.*, 2013; Cappallo-Obermann and Spiess, 2016). Spermatozoa are considered to be “silent cells” and translationally inactive due to the cytoplasm being largely removed during spermatogenesis (Gilbert *et al.*, 2007; Grunewald *et al.*, 2005; Miller and Ostermeier, 2006a). *De novo* protein synthesis may be due to mitochondrial translation in spermatozoa, which was confirmed using mitochondrial translational inhibitors to completely knock out mitochondrial protein translation (Hecht and Williams, 1978; Gur

and Breitbart, 2006). However, cytoplasmic translation has not been shown to date (Premkumar and Bhargava, 1972; MacLaughlin and Terner, 1973; Gur and Breitbart, 2006; Cappallo-Obermann *et al.*, 2011). The consensual view that spermatozoa have no intact cytoplasmic translational apparatus is supported by the absence of intact 18S and 28S rRNA, which is now being used to check for the absence of RNA arising from somatic cells in spermatozoal RNA preparations (Goodrich *et al.*, 2013; Johnson *et al.*, 2011). The evidence for translation in spermatozoa is complicated by the finding of bacteria in semen, which may have been responsible for translational processes detected and raises the question of whether spermatozoal RNA and proteins can be synthesised *de novo* (Premkumar and Bhargava, 1972; Abraham and Bhargava, 1963; Díez-Sánchez et al., 2003b).

Cappallo-Obermann *et al.* (2011) showed that spermatozoa contain incomplete 18S and 28S rRNA, which are components of the 80S ribosomes, necessary for protein translation. The 18S rRNA is present in all spermatozoa irrespective of somatic cell contamination and seems to be an 'inherent' characteristic in pure populations of spermatozoa according to Cappallo-Obermann *et al.* (2011). Different studies confirmed these partial and total absence of 18S and 28S rRNAs visible in Bioanalyzer traces of spermatozoal RNA, additionally showing further peaks deriving possibly through mitochondrial rRNA (12S and 16S) and degraded spermatozoal RNA (Gilbert *et al.*, 2007; Goodrich *et al.*, 2007; Bissonnette *et al.*, 2009; Das *et al.*, 2010; Cappallo-Obermann *et al.*, 2011; Jodar *et al.*, 2012; Sendler *et al.*, 2013; Ing *et al.*, 2014). Cappallo-Obermann *et al.* (2011) hypothesised that the large ribosomal subunits are degraded or not seen in Bioanalyzer traces, because non-functional rRNA may undergo RNA degradation in the exosome (LaRiviere *et al.*, 2006; Doma and Parker, 2006; Cole *et al.*, 2009). As paternal mitochondria, paternal ribosomes may be degraded ensuring that no paternal transcripts which may be lethal for post-fertilisation processes are delivered to the oocyte (Cappallo-Obermann *et al.*, 2011). Cappallo-Obermann *et al.* (2011) and Lalancette *et al.* (2008b) observed that ejaculates with abundant 18S rRNA profiles led to either sudden pregnancy termination in humans or a low return rate in less fertile bulls.

Given that the rRNA subunits are incomplete in spermatozoa, it is generally considered that they are not functional. This was also confirmed by Gur and Breitbart (2006), who additionally showed that mitochondrial like ribosomes are involved in *de novo* protein synthesis.

Taken together, these findings show that a large pool of RNA exists in spermatozoa, which is at least partially degraded of intact rRNAs and other RNAs, and may have functions, perhaps in post-fertilisation regulatory processes requiring further analysis.

1.2.2 RNA subtypes

In general, around 5-10% of the human genome encodes for transcribed sequences, however, the majority does not encode for proteins (Ponting *et al.*, 2009). Only 1% of the genome encodes proteins which leaves between 4% and 9% of coding RNAs with an unknown function to date (Ponjavic and Ponting, 2007; Ponting *et al.*, 2009). Miller *et al.* (1999) and others demonstrated that spermatozoa contain a wide range of RNA species, including specific coding and non-coding, sncRNAs, antisense equivalents of several protein-encoding sense mRNAs, putative small RNA candidates and RNA from transposable elements (see also Figure 1-2) (Chiang *et al.*, 1994; Kumar *et al.*, 1993; Lai *et al.*, 1996; Krawetz *et al.*, 2011; Ostermeier *et al.*, 2005b; Jodar *et al.*, 2013). Most of the RNAs are degraded and their functions are neither known nor annotated; however, they appear uniquely in spermatozoa (Jodar *et al.*, 2013; Bissonnette *et al.*, 2009; Sendler *et al.*, 2013; Ing *et al.*, 2014). Most of the coding genes are well-conserved and fulfil equivalent functions between species (Werner and Swan, 2010). It is speculated that RNA molecules might play an unknown role in fertilisation and early embryo development, and may also be useful as markers for male fertility (see section 1.4).

1.2.2.1 Coding RNAs

The function of coding RNA is simply to code for protein synthesis in the cytoplasm through the ribosome machinery and tRNAs (Caspersson and Schultz, 1939; Caspersson, 1941; Brachet and Chantrenne, 1956; Brenner *et al.*, 1961; Crick, 1958). Most coding RNAs are messenger RNAs (mRNA) which contain a polyadenylic acid (poly(A^+)) sequence that not only controls the translational efficiency but is also involved in the stability and transport of the RNA, in contrast to some mRNAs without a poly(A^+) sequence like histone mRNA (Adesnik *et al.*, 1972; Gallie, 1991; Colgan and Manley, 1997; Grunewald *et al.*, 2005). Untranslated regions (UTRs) flank both sides of the mRNA transcript. The 3´UTR contains regulatory elements and may be modified, therefore leading to alternative splicing and consequently to several isoforms of the transcript, altering not just the sequence of the mRNA but also its function (Kleene, 2005; Liu *et al.*, 2007). The first described spermatozoal transcript in mature spermatozoa was *c-MYC*, however, with the development of better techniques, more transcripts like the most abundant spermatozoal transcripts *PRM1* and *PRM2* and *β-actin* have been detected using RT-PCR, *in situ* hybridisation, microarrays and RNA-sequencing (Kumar *et al.*, 1993; Miller *et al.*, 1994; Miller *et al.*, 1999; Wykes *et al.*, 1997; Huszar and Vigue, 1990; Parrington *et al.*, 2002; Sendler *et al.*, 2013). Due to the cost decrease in microarray and Next-generation sequencing analysis and technology improvements, examination of patient data could contribute into the global understanding of spermatozoal mRNA. Patterns in infertile, teratozoospermic and cryptorchidism patients and smoking men vs. fertile men could be analysed and were indicative of a decrease in

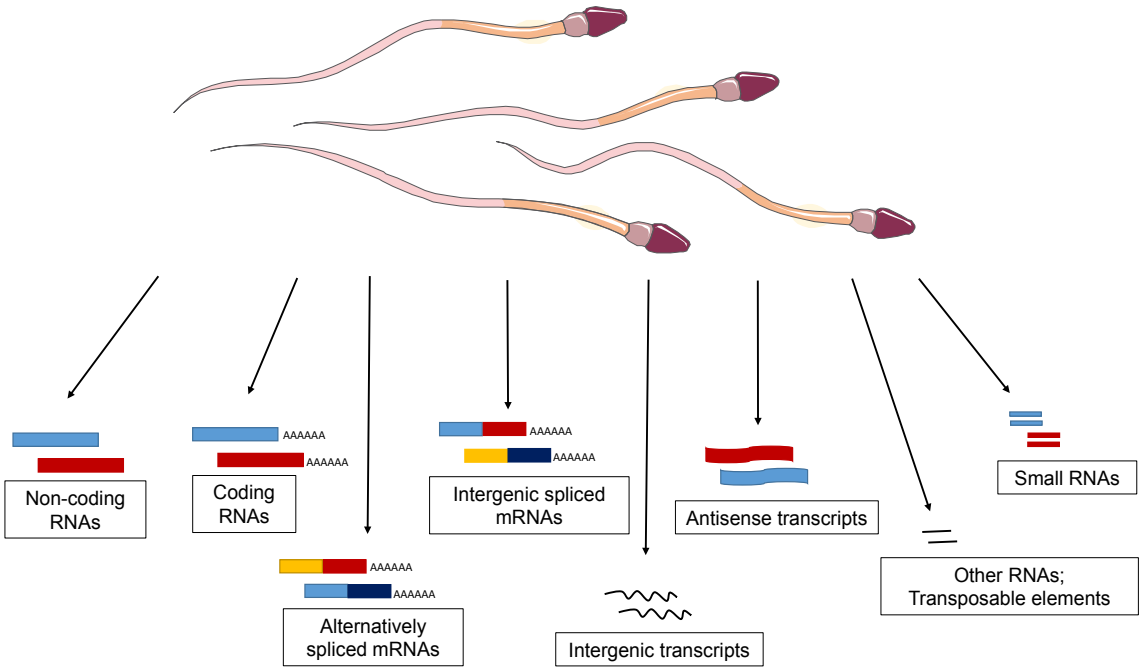


Figure 1-2: Spermatozoal RNA subtypes. Spermatozoa contain a complex repertoire of different RNA species: Coding, non-coding, alternatively spliced mRNAs, intergenic spliced mRNAs, intergenic transcripts, antisense transcripts, small RNAs and transposable elements are found to date. Images used from "Servier Medical Art Powerpoint Image Bank"

genes involved in spermatogenesis, spermatozoal motility and germ cell anti-apoptotic processes (*PRM2*, *SPZ-1*, *SPATA-4*, *NEA-1* and *CREM*), DNA repair (*NIPBL*), oxidative stress regulation (*PARK7*) and histone modification (*DDX3X*, *JMJD1A*) compared to normozoospermic men (Platts *et al.*, 2007; Lalancette *et al.*, 2008a; Garrido *et al.*, 2009; Linschooten *et al.*, 2009; Nguyen *et al.*, 2009; García-Herrero *et al.*, 2010; García-Herrero *et al.*, 2011; Hamatani, 2012; Montjean *et al.*, 2012). Sequencing profiles of normozoospermic men showed increased *PRM1, eNOS and nNOS* mRNA in low density fractions compared to high density fractions of the same donor (Lambard *et al.*, 2004). The right ratio of *PRM1, PRM2* and *TNP1* is important for the histone protamine exchange in the post-meiotic phase of spermatogenesis leading to chromatin condensation (Steger *et al.*, 2001).

Interestingly, one quarter of coding spermatozoal RNAs show alternative sites of polyadenylation (APA) which keeps the integrity of the coding region but exposes an abbreviated 3´UTR, which Liu *et al.* (2007) described in testis and might modulate transcript stability, localisation and/or transport of the coding RNA. Different regulatory proteins and miRNAs might bind to the alternative UTR through the impact of translational processes introduced through modifications in APAs (Di Giammartino *et al.*, 2011). Fischer *et al.* (2012) suggests that a substantial proportion of genes relevant for

spermatogenesis and found enriched in testis expression profiles are not included in the spermatozoal transcriptome e.g. open reading frames (ORF) are low in testis and absent in somatic cells but have been found to be abundant in spermatozoa (Jodar *et al.*, 2013). Due to the shutdown of gene expression from the X chromosome during meiosis and to accommodate the switch from transcriptional to translation control of protein synthesis, alternative splice-variants have been found in spermatozoa, assuming that these transcripts arise during the final transcriptional stage of spermatogenesis or are mainly involved in spermatozoal functions like motility, acrosome reaction or capacitive functions e.g.: *PRM1* and *PRM2*, *TNP1*, *GAPSDH*, *ODF1,* oestrogen receptors, integrins, L-type calcium channels, N-cadherin, aromatases (Miller, 1997; Wykes *et al.*, 1997; Ostermeier *et al.*, 2002; Kempisty *et al.*, 2007; Depa-Martynów *et al.*, 2007; Jodar *et al.*, 2013). See more about spermatozoal coding RNAs in section 1.5 and Chapter 3.

1.2.2.2 Non-coding RNAs

It is a misconception that RNA operates only as a messenger between DNA and proteins. Approximately 98% of transcribed sequences lie in noncoding RNAs that are derived from intergenic, intronic and UTR sequences (Mattick, 2001; Babak *et al.*, 2007). The first non-coding RNA discovered was a tRNA in yeast in 1965 (Holley *et al.*, 1965). Several types of non-coding RNAs have been discovered to date: lncRNA, sncRNA, snorRNA, miRNA, piwiRNA, rRNA, tRNA and the known function lies either in repression or expression regulation or protein synthesis (Ponting *et al.*, 2009; Askew and Xu, 1999; Eddy, 1999; Eddy, 2001). Non-coding RNAs can be divided into housekeeping non-coding and regulatory non-coding RNAs. The former group include mitochondrial, ribosomal, transfer, small nuclear and small nucleolar RNAs, in contrast to the latter group regulating micro RNAs, small interfering RNAs and Piwi-associated RNAs (Ponting *et al.*, 2009). Both groups can be placed into five broad categories: “(1) sense, or (2) antisense, when overlapping one or more exons of another transcript on the same, or opposite, strand, respectively; (3) bidirectional, when the expression of it and a neighbouring coding transcript on the opposite strand is initiated in close genomic proximity, (4) intronic, when it is derived wholly from within an intron (although these may sometimes represent pre-mRNA sequences), or (5) intergenic, when it lies within the genomic interval between two genes” (Eddy, 1999; Eddy, 2001; Mattick and Makunin, 2006; Mercer *et al.*, 2009). A former misunderstanding was that non-coding RNAs have no functionality, since they seemed poorly conserved (Pang *et al.*, 2006). However, non-coding RNAs indeed show functionality, and are moreover also associated with protein-coding genes (Ponjavic *et al.*, 2007). Kapranov *et al.* (2007) found transcription of ncRNAs concentrated at the promoter, initial 5’ exons and introns of genes. In addition, it has also been described that non-coding RNA genes do not form large homologous

families and also that RNA relics of transcribed transposable element insertions are very common (Ravasi *et al.*, 2006).

1.2.2.3 Long Non-Coding (lnc) RNA

lncRNAs can be found throughout the genome and vary between ~200 nt to 1000 nt. The localisation varies, as it has also been observed for protein-coding genes (Mercer *et al.*, 2008). It is assumed that the transcription of long non-coding RNAs promotes the accessibility of protein-coding genes to RNA polymerases and that they function through the binding of DNA or protein (Hirota *et al.*, 2008; Ponting *et al.*, 2009). Studies in somatic cells have shown that long non-coding RNAs can modulate transcription through the interaction with an associated promoter region, which is either a *cis*-acting regulation (derived from or near protein coding loci) close to the genomic proximity or a *trans* – acting regulation, which targets transcriptional activators or repressors and post-transcriptionally during splicing (Martianov *et al.*, 2007; Ponting *et al.*, 2009; Yoon *et al.*, 2012). Carninci *et al.* (2005) hypothesised that long non-coding RNAs may be highly conserved in promoter regions, compared to sequences in their transcripts. Another long non-coding RNA family CAR (Chromatin-associated RNAs) was also previously identified playing a role in gene expression or genomic architecture in *cis* and *trans* position (Mondal *et al.*, 2010; Rodríguez-Campos and Azorín, 2007). Long non-coding RNAs also interact with repressors, modulating the translation of RNAs or with general interference, regulating expression levels in *cis*, with the initiation complex (Kindler *et al.*, 2005; Ponting *et al.*, 2009; Yoon *et al.*, 2013; Yoon *et al.*, 2012). Katayama *et al.* (2005) described up to 72% bi-directional transcription in the mouse genome and similar percentages in other species, which involves a class of long non-coding RNA called natural antisense transcripts (NAT). NATs are derived from the reverse strand, are fully processed, are not evenly distributed and are mainly found enriched 250 nucleotides upstream of the transcription start side and 1.5 kb downstream of sense genes (Ostermeier *et al.*, 2005b; Sendler *et al.*, 2013; Werner and Swan, 2010; Jodar *et al.*, 2013; Preker *et al.*, 2008). Antisense RNA can either encode for proteins or also act as a non-coding RNA, whereas the most abundant form found is a non-protein-coding antisense RNA partner of a protein-coding RNA (Faghihi *et al.*, 2008; Faghihi and Wahlestedt, 2009; Preker *et al.*, 2008; Uchida *et al.*, 2004). Werner and Swan (2010) reported the association of antisense transcripts with many active promoters and the protection of the sense transcript against nuclease degradation. Furthermore, the transcription of antisense RNA strongly correlates with the expression of the sense transcripts as was previously described (Preker *et al.*, 2008; Werner and Swan, 2010). It has been reported that NATs accumulate locally and trigger DNA or chromatin modifications and their influence can extend to neighboring genes despite the fact that these genes are not always related to the NAT target (Faghihi and Wahlestedt, 2009).

Furthermore, Faghihi and Wahlestedt (2009) suggest that these antisense transcripts are able to silence miRNA binding sites. Taken together, NATs seem to be involved in gene silencing, changes at the sense strand, selective transcript editing, promoter inactivation and alteration, genomic rearrangements and epigenetic modification through DNA methylation and demethylation, chromatin modifications and monoallelic expression (Faghihi and Wahlestedt, 2009; Lapidot and Pilpel, 2006; Lavorgna *et al.*, 2004; Werner *et al.*, 2009; Werner and Swan, 2010). Interestingly, NATs were reported to be abundant in haploid spermatids found in mouse testis and seem to play a role in spermatogenesis compared to other tissues where the expression of NATs is at low levels (Carlile *et al.*, 2009; Werner and Swan, 2010). Werner and Swan (2010) suggested gene silencing as a selection factor for a specific spermatozoal population and therefore 'unfit' cells for survival of the fittest. Ostermeier *et al.* (2005b) assumed NAT RNAs play a role in early fertilisation processes, since they are also often associated with imprinted genes (81%) (Preker *et al.*, 2008). In this regard, strand orientations cannot be directly determined. Nolasco *et al.* (2012) reported long non-coding RNAs involved in the regulation of spermatogenesis and speculated that some might also be antisense. Antisense transcripts have been described as frequently functional, using diverse transcriptional and post-transcriptional gene regulatory mechanisms, carrying out a variety of biological roles (Faghihi and Wahlestedt, 2009). Interestingly, antisense transcripts found in the testis are abundant and might contribute to the regulation and function of gene expression during spermatogenesis (Lee *et al.*, 2009). Furthermore, it is thought that the antisense transcripts might play a role in early embryonic development, providing control that helps to establish imprints during the maternal / paternal to zygotic genome transition (Ostermeier *et al.*, 2005b). The product of long non-coding RNAs or the long non-coding RNAs itself may be important for epigenetic gene silencing of imprinted genes (Ponting *et al.*, 2009). Imprinting clusters also contain sncRNAs, which might be derived from long non-coding RNAs. DNA methylations and mRNA alterations were seen in imprinted vs. non-imprinted genes comparing patients with abnormal semen parameters, however, promoter associated DNA methylations profiles were correlated with spermatozoal mRNA content. Additionally, hypermethylation and low levels of imprinted and critical epigenetic regulatory genes seem associated with reduced motility of spermatozoa (Pacheco *et al.*, 2011). An abundance of nested and overlapping gene structures in coding and non-coding transcripts are seen in mammalian transcriptomes (Preker *et al.*, 2008). Several abundant spermatozoal non-coding transcripts, between 100-300 nt in size, were found overlapping either the coding or UTR section of less abundant transcript (Jodar *et al.*, 2013).

To this end, transcription and translation also seems to be mediated by lncRNAs, which might play an important role in spermatogenesis, gametogenesis, fertilisation and embryogenesis, and this has to be further explored.

1.2.2.3.1 Small non-coding RNAs

Small-non coding RNAs (sncRNAs) are required for gene activation and/or repression and are therefore important for the regulation of gene expression. sncRNAs are derived from double stranded endogenous pre-cursors and have a size between 20-24 nt, in contrast to piRNAs, which are derived from single-stranded precursors, with a size of 26-32 nt (Werner and Swan, 2010). It has been described that sncRNAs play an important role in cellular differentiation, apoptosis and proliferation, and might be derived from transposable elements and pseudogenes (Preker *et al.*, 2008). Transposable elements may be highly active in spermatogenesis, might be altered through sncRNAs during translational regulation and are likely repressed through Piwi RNAs (piRNAs) (McIver *et al.*, 2012; Aravin *et al.*, 2007). Maternally derived sncRNAs are stable for various cell divisions and play a suspected role in gene and transposon regulation in the early embryo (Li *et al.*, 2013; Nodine and Bartel, 2012; Olszańska and Borgul, 1993).

If maternal sncRNAs or repetitive elements are once paired with paternal RNA, then their partner genes maybe activate or suppress modifications in the epigenome of the zygote (Suh and Blelloch, 2011; Lippman *et al.*, 2004). Small non-coding RNAs may play a role in non-Mendelian inheritance of traits or phenotypes obtained throughout life (Gapp *et al.*, 2014; Kawano *et al.*, 2012; Krawetz *et al.*, 2011). Krawetz *et al.* (2011) described different small RNA populations in human spermatozoa using the ejaculate of three different fertile men. Looking at the small RNAs, using the RNA next-generation sequencing approach, the fertile ejaculates showed 7% of miRNAs, 17% of piRNAs, 65% of repeat sequences (including retrotransposons) and finally 11% of short RNAs within the transcription start site/promoter fraction, revealing enriched genes involved mainly in early embryonic development. Although some sncRNA research in spermatozoa was performed recently, the functions are poorly understood and performed studies are of a preliminary nature.

1.2.2.3.2 miRNAs and siRNAs

miRNAs are of 21-22 nt in length and their function lies mainly in the regulation of gene expression. Many are located in the intronic region of protein-coding sequences, whereas it is not certain if miRNAs are co-regulated with their host genes, miRNAs might have their own promoters (Kim, 2006; Ying and Lin, 2005). miRNAs bind to 3´UTR of their complementary target RNA and are involved in its stability and/or translation, which might be downregulated through degradation or translational suppression of the target,

something which is still not clearly understood (Bartel, 2009). However, it is known that each miRNA has many targets and the mRNA might be targeted by more than one miRNA. It is also speculated that miRNAs are involved in histone modifications and hence the regulation of chromatin structure. Therefore, it is thought that miRNAs are involved in epigenetic changes (Chuang and Jones, 2007).

Ostermeier *et al.* (2005b) hypothesised that spermatozoal miRNAs might be derived from testis miRNAs, since the spermatozoal transcripts reported were equal to the transcripts found in testis. An altered pattern of miRNAs was found in non-obstructive azoospermia (see explanation at Table 1-1) patients. In addition, the testes of infertile patients showed an increase in miRNA levels compared to fertile patients (Lian *et al.*, 2009; Abu-Halima *et al.*, 2014). Several groups described the correspondence between increased miRNA levels and increasing gene transcription at the pachytene spermatocyte and post meiotic stages during spermatid development in mice (Ro *et al.*, 2007; Yan *et al.*, 2007). Marcon *et al.* (2008) examined miRNAs in elongating spermatids and pure spermatozoa using microarrays, observing miRNA mainly expressed in the meiotic germ cells. It is assumed that paternal sncRNAs are influencing fertility rates and additionally seem to be mediators in epigenetical non-Mendelian inheritance in the offspring resulting in potential health risks (Rassoulzadegan *et al.*, 2006; Krawetz *et al.*, 2011; Liu *et al.*, 2012; Du *et al.*, 2014; Sendler *et al.*, 2013; Jodar *et al.*, 2013; Kawano *et al.*, 2012; Stowe *et al.*, 2014; García-López *et al.*, 2015; Fagerlind *et al.*, 2015; Govindaraju *et al.*, 2012; Curry *et al.*, 2011; Rodgers *et al.*, 2015; Chen *et al.*, 2016; Sharma *et al.*, 2016; Jimenez-Chillaron *et al.*, 2016).

Differences in the miRNA content may be therefore significant in embryo development, sudden termination of pregnancy, premature birth and in health risks passed on to the offspring (Rodgers *et al.*, 2013; Gapp *et al.*, 2014; Dias and Ressler, 2014). It is shown that sperm-borne miRNA once repressed leads to a termination of embryo growth (Liu *et al.*, 2012). Additionally to all human findings to date, Du *et al.* (2014) found that bovine miRNAs may be involved in fertilization events. Some groups confirmed paternal non-Mendelian transgenerational inheritance of diseases or disorders found in the progeny (Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Sharma *et al.*, 2016; Chen *et al.*, 2016). These are presumably derived from small non-coding RNAs. The function and modes of action of these miRNAs is nonetheless unknown to date. Injecting spermatozoal heads and anti-miRNAs into oocytes for miRNA suppression revealed no interruption in pronuclear activation or preimplantation development and suggests no important role of miRNAs in mammalian fertilisation (Amanai et al., 2006; Boerke et al., 2007). In addition, Amanai *et al.* (2006) showed that the miRNA levels in spermatozoa are far lower than in oocytes. Later research compared oocytes to one-cell zygotes and

found a highly conserved spermatozoal mir-34c which may play a role not only in spermatogenesis- being involved in the negative control of the cell cycle, but also in terms of being responsible for the early cleavage division in mice (Bouhallier *et al.*, 2010; Liu *et al.*, 2012; Pantano *et al.*, 2015). In addition, sperm-borne mir-34c was found to be involved in the first cleavage of the developing embryo (Liu *et al.*, 2012).

In spermatogenesis and during the first stages of early embryo development, little transcription takes place; one miRNA recently found in spermatozoa (miR-10a) activates and enhances rather than represses translation and could potentially be given to the developing embryo to support translation (Pantano *et al.*, 2015; Ørom *et al.*, 2008).

More recent research showed 182 miRNA among two fertile individuals. This research found some piRNAs arrived from pseudogenes (Pantano *et al.*, 2015). Only 37 out of the 182 transcripts were previously reported (Pantano *et al.*, 2015; Krawetz *et al.*, 2011). To this end, it seems that spermatozoal miRNA is transmitted to the zygote and may play roles in fertilisation, epigenetic regulations in the developing embryo and gene-expression regulation during pre-implantation development (Amanai *et al.*, 2006; Wagner *et al.*, 2008; Liu *et al.*, 2012; Pantano *et al.*, 2015).

The origin of endogenous small interfering RNA (siRNA) are mRNAs and their corresponding antisense transcripts (Preker *et al.*, 2008). Normally siRNAs consist of 21 nucleotides and are processed from a long dsRNA. Studies in mice found formations of endogenous siRNAs which regulate gene expression (Watanabe *et al.*, 2008). Findings of Watanabe *et al.* (2008) strengthen the cytoplasmic origin of siRNA, which corresponded to mRNA or retrotransposons in growing oocytes and derived from the overlapping region of the sense and antisense RNA. Sense and antisense RNA is regulated by the RNAi pathway, therefore endogenous siRNAs could regulate sense and antisense transcripts, including the suppression of retrotransposons through silencing (Watanabe *et al.*, 2008; Werner and Swan, 2010).

The former conviction that the oocyte may solely be responsible for embryonic development and the sperm only carries the required paternal genome has been challenged in the recent years and it has been shown that the spermatozoon is involved in carrying an epigenetic mark, which is involved in embryo development and in the health of the offspring (Rassoulzadegan et al., 2006; Rodgers et al., 2013; Dias and Ressler, 2014; Gapp et al., 2014; Chen et al., 2016).

1.2.2.3.3 Transposable and repeat elements

Krawetz *et al.* (2011) describe a large quantity of sncRNAs reads mapping to repetitive elements where LTR (long terminal repeat), SINE (short interspersed nuclear elements), ALU (*Arthrobacter luteus*) and LINE (Long interspersed nuclear elements) are the most

abundant forms found and play a possible role in early embryo function and altering the transcriptional machinery promoting their own expression together with endogenous retrovirus- mammalian apparent LTR retrotransposons (ERVL-MaLR) which have been described by Inoue *et al.* (2012) and Kigami *et al.* (2003) (Fadloun *et al.*, 2013). Transposable element sequences often seem to randomly coincide with transcription start sides and are very likely to contribute to changes in gene transcript repertoires (Conley *et al.*, 2008). Fadloun *et al.* (2013) address the regulation of repetitive elements through RNA and examine reprogramming after mammalian fertilisation enclosing transcriptional activation of retrotransposons. Repeat elements involved in epigenetic inheritance frequently originate from transposons and the high frequency is normally related to imprinted genes in mammals (Rando, 2012). However, light has to be shed on the questions regarding the role of transposable elements derived from the spermatozoon in terms of how they might modulate gene expression in the germline and in early embryogenesis. This remains a controversial topic to this day (Beraldi *et al.*, 2006; Georgiou *et al.*, 2009; van der Heijden and Bortvin, 2009; Jodar *et al.*, 2013).

1.3 Origin, localisation and hypothesized roles of RNA in spermatozoa

Miller *et al.* (1999) and others demonstrated that spermatozoa contain a wide range of RNA species including specific coding and non-coding RNAs, small RNAs, antisense equivalents of several protein-encoding sense mRNAs and putative siRNA candidates (Chiang *et al.*, 1994; Kumar *et al.*, 1993; Lai *et al.*, 1996; Krawetz *et al.*, 2011; Ostermeier *et al.*, 2005b; Jodar *et al.*, 2013). It was originally suggested that these RNAs were remnants of untranslated stores left over from spermatogenesis, giving a historic record of spermatogenesis (Ostermeier *et al.*, 2002). Little is known about their roles to date.

According to Pessot *et al.* (1989), spermatozoal RNA is localised in the periphery of the nucleus (perinuclear theca), close to the nuclear envelope and post-acrosomal sheath, whereas other studies reported localisation at the mid- and principle piece of the tail (see Figure 1-3) (Kumar *et al.*, 1993; Modi *et al.*, 2005; Miller and Ostermeier, 2006a). Furthermore, spermatozoal RNA is probably co-localised or closely associated with chromatin in the nucleus and might play a role within the spermatozoal nuclear matrix (Miller, 2014; Pessot *et al.*, 1989; Rejon *et al.*, 1988; Martins and Krawetz, 2005). The axoneme and the fibrous sheath are regions of the spermatozoa which may also contain RNA. Some transcripts (e.g. *SMCY*) are also found at the spermatozoal surface, and may play a role in fertilisation or spermatozoal survival (Anderson, 2013).

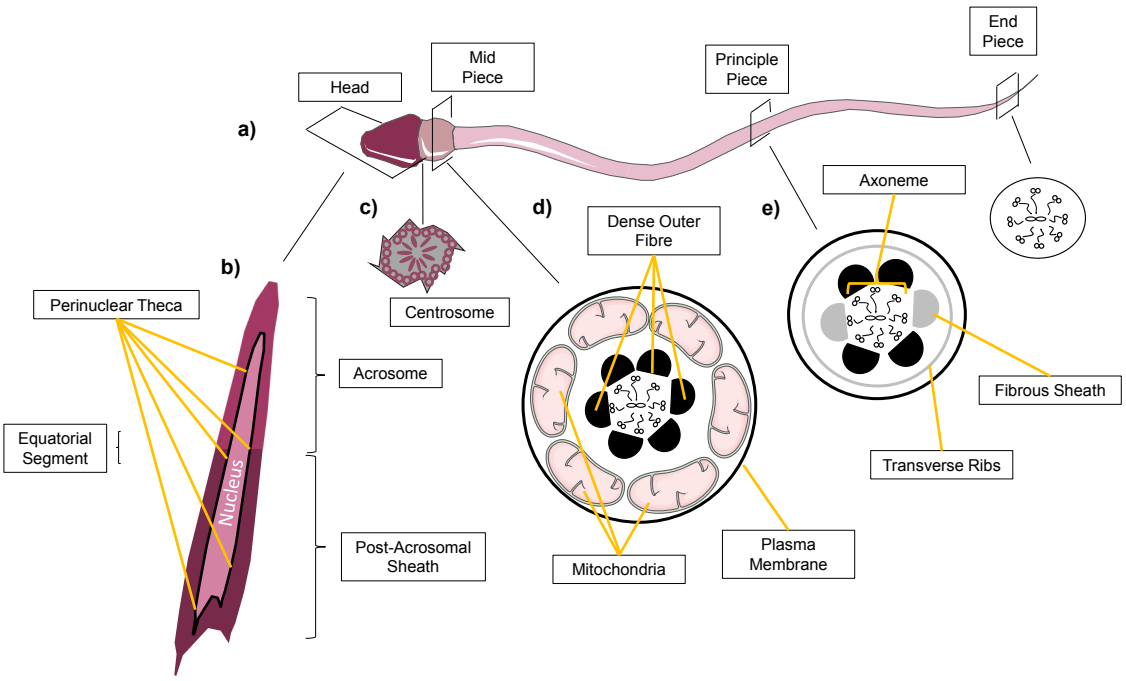


Figure 1-3: Structure and RNA localisation of a spermatozoon. a) shows the four main segments of a spermatozoon: head, mid piece, principal piece and the end piece; b) till e) shows possible locations for spermatozoal RNA, like the nucleus, perinuclear teca and post-acrosomal sheath as well the centrosome; c) and in d) the mitochondria; e) the fibrous sheath and the axoneme might be possible spermatozoal RNA locations. Spermatozoal RNA in the perinuclear theca and the midpiece have been shown to date. Picture adapted from Miller and Ostermeier (2006a); Images used from "Servier Medical Art Powerpoint Image Bank"

One possible role for spermatozoal RNA is *de novo* protein synthesis in order to replace proteins, degraded during capacitation but restricted to mitochondrial translational pathways (Gur and Breitbart, 2006). Like maternal RNAs, paternal RNA may be required from the first embryonic cleavages until the activation of the embryonic genome occurs (Liu *et al.*, 2012). It is also possible that spermatozoal RNA transfers male benefit traits and selfish genetic elements (SGE) to secure the successful transmission of the paternal genome (Miller, 2006; Miller and Ostermeier, 2006a; Hosken and Hodgson, 2014; Miller, 2015). Selfish genetic elements are elements enhancing their own transmission relative to the rest of an individual's genome and are passed on at a higher frequency to ensure accumulation within the genome (Miller and Ostermeier, 2006a; Miller, 2015). SGE could be small inhibitory RNAs, antisense RNAs or episomal elements (Wedell, 2013; Werren, 2011). Furthermore, spermatozoal RNA might be reverse transcribed into cDNA through an endogenous reverse transcriptase which could be involved in fertilisation events (Sciamanna *et al.*, 2003; Miller and Ostermeier, 2006b).

Although evidence for transcriptional and translational activity in the cytoplasm of spermatozoa has not been found, transcriptional and translational activity is present in the mitochondria of mature spermatozoa (Hamatani, 2012; MacLaughlin and Terner, 1973; Miller and Ostermeier, 2006a; Gur and Breitbart, 2006). Miller and Ostermeier (2006a), hypothesised that these RNAs might also play a role in epigenetic re-programming of spermatozoal chromatin. Since the RNA is a component of the nuclear envelope, the hypothesis suggested that the RNA might stabilise an interaction between the envelope and histone-bound DNA. Another role of the RNA could be in chromatin re-organisation or marking DNA sequences for histone packaging (Hamatani, 2012). The delivery of these spermatozoal RNAs into the oocyte during fertilisation may facilitate a procedure for checking, through recognition and interaction of ncRNAs of a 'foreign' genome for consolidation and acceptance by the oocyte (Miller, 2014). New studies support the hypothesis of paternal environmental effects transmitted through changes in the spermatozoal epigenome e.g. the paternal influence on the offspring phenotype which could happen through modified chromatin methylation derived from chromatin modifications through epigenetic RNAs (Rando, 2012; Jodar *et al.*, 2013; Khraiwesh *et al.*, 2010; Krawetz *et al.*, 2011; Taft *et al.*, 2011).

Dadoune (2009) also suggested the possibility that spermatozoal RNA mediated epigenetic changes and delivery to the oocyte could therefore influence embryonic development. Additionally, Sone *et al.* (2005) found that the injection of *Phospholipase C, zeta* 1 (*PLC- ζ*) RNA into mouse oocytes caused Ca^{2+} oscillations and oocyte activation. Microinjections of brain and testis mRNA into fertilized oocytes can induce a heritable mutant phenotype (Rassoulzadegan *et al.*, 2006). Rassoulzadegan *et al.* (2006) demonstrated that RNA and miRNA injection into mouse embryos led to litters born with an introduced paramutation, visible in the phenotype. Paramutations are introduced through RNA-silencing, which was firstly described in maize affecting plant colour (Brink, 1956). Dias and Ressler (2014), examined the inheritance of paternal traumatic exposure in mice. These male mice were subjected to a special kind of odour and fear-conditioned. The F1 and F2 generations of these mice were found to be as behaviorally sensitive to the odour as the male F0 conditioned founders. Furthermore, they showed that these transgenerational effects were inherited via the spermatozoon. It seems that the somatic stress is transferred to the germline, possibly via epigenetic changes and possibly through miRNAs and tRNAs (Rodgers et al., 2013; Miller, 2014; Dias and Ressler, 2014; Gapp et al., 2014; Chen et al., 2016).

The influence of the phenotype is strengthening the hypothesis that the spermatozoon does not simply transfer the paternal genome. Spermatozoal RNA introduced to the oocyte through fertilization may introduce a heritable epigenetic change in the zygote.

While it is widely accepted that RNA can modify the epigenome and alter gene expression, RNA transmission through the germ line is still a matter of debate and the function of spermatozoal RNA has to be further explored, particularly in relation to maturation, fertilization, early embryo development and phenotype after birth (Rassoulzadegan *et al.*, 2006). It has to be considered that spermatozoa are not the only component in the ejaculate which might play a role in influencing the offspring phenotype (Rassoulzadegan and Cuzin, 2010). Seminal fluid can influence the kinetics of embryo development in mammals, although the exact process by which it achieves this is currently unclear (Rando, 2012).

1.4 Spermatozoal dysfunction

Infertility is a worldwide problem which affects 10-15% of couples (De Kretser, 1997; Evers, 2002). The question of infertility arises if frequent unprotected intercourse attempts failed to conceive a child after 12 month (WHO, 2010). When investigating fertility, both low female fertility rates and possible semen quality deterioration should be taken into account (Jensen *et al.*, 2002). Semen quality plays a crucial role in human reproduction (Sun *et al.*, 1997). 7% of men worldwide are estimated to be affected by infertility and one quarter are diagnosed with idiopathic infertility (Yatsenko *et al.*, 2006; Roy *et al.*, 2007). A male patient is infertile if the spermatozoa he produces are not able to fertilise an oocyte or if a miscarriage arises because of a defect in the spermatozoon. The reason for male infertility is not well understood and it is thought that 15%-30% of cases originate through genetic contributions (Ferlin *et al.*, 2007; Moldenhauer *et al.*, 2003; Yatsenko *et al.*, 2006). The analysis of visible infertility parameters includes semen quality evaluation, where spermatozoal concentration, motility and morphology are examined using the World Health Organization guidelines (see Table 1-1 for categories and nomenclature) (WHO, 2010). Spermatozoal concentration in semen tends to be associated with pregnancy rates (Anton and Krawetz, 2012). A significant decrease in the number of spermatozoa in the semen and severe oligozoospermia (low concentration of spermatozoa, see Table 1-1) or azoospermia (men with abnormally low or no numbers of spermatozoa in their semen, see Table 1-1) is mostly found in idiopathically infertile men (Wu *et al.*, 2012; WHO, 2010). Gandini *et al.* (2000) showed that the shape of spermatozoa is correlated with their function, but morphologically abnormal spermatozoa are not necessarily genomically abnormal (Yanagimachi, 2005). Yatsenko *et al.* (2006) reports the rising concern for the epigenetic effects on offspring born from assisted reproductive techniques (ART), ~1-4% of children were conceived through ART in developed countries in 2006 (Yatsenko *et al.*, 2006).

Infertility could have many causes and spermatozoal quality can be affected by several factors including viral or bacterial infection of the testes and accessory glands. Additionally, infections in the genital tract could also lead to infertility. Almost 15% of male infertility results from infections of the male genitourinary tract (Moretti *et al.*, 2009). Many microorganisms are involved in various ways and not only the testes are affected (Pellati *et al.*, 2008). Some infections caused by genital ureaplasma and/or mycoplasmas originate from the urinary tract and contaminate the semen during ejaculation (Moretti *et al.*, 2009). The ejaculate is an excellent medium for supporting bacteriological activity (Miller and Ostermeier, 2006a). Many different reasons that cause male reproductive problems could be a result of viral (e.g. peritubal mumps) or bacterial (e.g. *Streptococcus agalactiae*, *Staphylococcus epidermidis*, *Streptococcus anginosus*, *Ureaplasma urealyticum*, *Enterococcus faecalis*) infection of semen, which lead to spermatozoal necrosis.

Other factors include genital tract obstructions, unilaterally/bilaterally undescended testis, varicocele and genital malformations. Furthermore, mutations in regulatory elements of spermatogenesis or Y-chromosome deletions, increases in gross chromosomal abnormalities and DNA fragmentation may also affect both spermatozoal quantity and quality (Ahmadi and Ng, 1999; Braun, 1998; Miller and Ostermeier, 2006a; Moldenhauer *et al.*, 2003; Singh *et al.*, 2003; Sun *et al.*, 1997). Semen quality also seems to be sensitive to environmental toxicants, which can impair spermatogenesis through reactive oxygen species (ROS). Sun *et al.* (1997) showed that infections could increase the concentration of ROS, which then reduces semen quality. These ROS could be a result of subsequent inflammation either in the testes where spermatozoa are formed or through passage and storage in the epididymides and downregulation of DNA repair during the late stages of spermatogenesis (Ahmadi and Ng, 1999; Lewis and Aitken, 2005). A normal ROS level is needed for regulating normal testicular function. An uncontrolled level of ROS has damaging effects, which can result in DNA strand breaks.

Interestingly, Jensen *et al.* (2009) published data that suggested a decrease in mortality among men with good semen quality was linked to a decrease in a wide range of diseases. In addition, the frequency of ejaculation and also the countries in which men are living may play a role. Additionally, a significant decrease in the mean seminal volume was also found (Ahmadi and Ng, 1999; Mathur and D'Cruz, 2011). However, it seems that geographical differences exist and that these geographical differences are also correlated with a perceived severe decrease in semen quality, and correlations of testis abnormalities are assumed (Carlsen *et al.*, 1992; Skakkebék *et al.*, 1998).
Similarly, endocrine and metabolic diseases, drug use and abuse, psychiatric conditions and age may also affect fertility (Ahmadi and Ng, 1999; Braun, 1998; Delbes *et al.*, 2007;

Fraga *et al.*, 1996; Moldenhauer *et al.*, 2003; Pellati *et al.*, 2008; Singh *et al.*, 2003; Sun *et al.*, 1997) Skakkebék *et al.* (1998) hypothesized that there may be a connection between germ cell cancer and a generalized disorder of spermatogenesis. Germ cell cancer is the most common malignancy in young men (aged 20-45 years). Over the past 50 - 60 years, testicular cancer has increased dramatically, equating to a fourfold increase- especially in Western countries. Men with genetic predisposition of cancer, Hodgkin`s lymphoma or congenital malformations of the male reproductive tract such as cryptorchidism should be especially aware of the difficulties of fathering a child (Gandini *et al.*, 2003). Testicular malignancy may arise during foetal life, and could be associated with infertility in later life. In most cases, an impairment or disturbance of spermatozoal production is seen. Fertile spermatozoa can be damaged through cancer/Hodgkin's lymphoma irradiation and chemotherapy, which can cause azoospermia (see Table 1-1 for explanation) due to loss of spermatogonia, even at low doses. Drug treatment results in an increase in the number of DNA strand breaks in spermatozoa (Delbes *et al.*, 2007). These strand breaks are correlated with infertility (Sakkas *et al.*, 2003). Furthermore, these strand breaks could impair pregnancy outcome, resulting in higher mortality rates at birth or higher morbidity (Delbes *et al.*, 2007). Some disease treatments require whole body irradiation, which can lead to permanent loss of spermatogonia. The quality of spermatozoa gives an indication of testicular health, which is linked with the germ cell production. Testicular and prostate cancer therapy treatment often includes surgical orchidectomy, which patients may find psychologically challenging especially if the patient wishes to father a child in future. Data after unilateral orchidectomy indicate that a 50% reduction in spermatozoa concentration occurs. Unfortunately, there is also the effect that 10% of men undergoing cancer therapy become azoospermic before the testicle has to be removed (Chan *et al.*, 2012; Delbes *et al.*, 2007). Furthermore, 8% of males who undergo a surgical procedure do not produce spermatozoa in the contralateral testis. Loss of spermatogonia will most probably result in infertility (Giwercman and Petersen, 2000). Even if some spermatogonia survive the treatment, they might fail to proliferate and/or differentiate. At higher doses, Leydig cells, which are important for the male hormone balance, may be affected. In general, Leydig cells are more resistant to cancer treatment than spermatozoa. The dosage and the treatment procedure determine the level of impairment. Before the treatment is started, there is still the option of trying to preserve fertility using cryopreservation. Cryopreserved spermatozoa can be used for Intrauterine insemination (IUI) or Intracytoplasmic sperm injection (ICSI) during IVF treatment in IVF clinics. One of the treatments used to increase the chance of pregnancy among infertile patients involves the use of testicular and epididymal spermatozoa aspiration and extraction (TESA) for patients diagnosed with aspermia, which today is attained through testicular biopsy. Testicular biopsy always

carries a health risk for the patient and there is a possibility that the tissue sample will be too small to reveal the actual cause of infertility and in addition to be able to provide spermatozoa capable of fertilising oocytes (Linschooten *et al.*, 2009; Funaro and Paduch, 2014; Goldstein, 2002). A non-invasive and cost-effective approach is needed as an alternative for invasive approaches.

Nomenclature	Condition
aspermia	No semen (no or retrograde ejaculation)
asthenozoospermia	Percentage of progressively motile (PR) spermatozoa below the lower reference limit
asthenoteratozoospermia	Percentages of both progressively motile (PR) and morphologically normal spermatozoa below the lower reference limits The lower reference limit for total motility (progressive + non-progressive) is 40%. The lower reference limit for morphological normal forms is 4%.
azoospermia	No spermatozoa in the ejaculate (given as the limit of quantification for the assessment method employed)
cryptozoospermia	Spermatozoa absent from fresh preparations but observed in a centrifuged pellet
necrozoospermia	Low percentage of live, and high percentage of immotile, spermatozoa in the ejaculate The lower reference limit for vitality (membrane-intact spermatozoa) is 58%.
normozoospermia	Total number (or concentration, depending on outcome reported) of spermatozoa, and percentages of progressively motile (PR) and morphologically normal spermatozoa, equal to or above the lower reference limits. The lower reference limit for sperm concentration is 39×10^6 spermatozoa per ml. The lower reference limit for total motility (progressive + non-progressive) is 40%.

Nomenclature	Condition
oligoasthenozoospermia	Total number (or concentration, depending on outcome reported) of spermatozoa, and percentage of progressively motile (PR) spermatozoa, below the lower reference limits. The lower reference limit for sperm concentration is 39×10^6 spermatozoa per ml. The lower reference limit for total motility (progressive + non-progressive) is 40%
oligoasthenoteratozoospermia	Total number (or concentration, depending on outcome reported) of spermatozoa, and percentages of both progressively motile (PR) and morphologically normal spermatozoa, below the lower reference limits. The lower reference limit for sperm concentration is 39×10^6 spermatozoa per ml. The lower reference limit for total motility (progressive + non-progressive) is 40%
oligoteratozoospermia	Total number (or concentration, depending on outcome reported) of spermatozoa, and percentage of morphologically normal spermatozoa, below the lower reference limits. The lower reference limit for sperm concentration is 39×10^6 spermatozoa per ml.

Nomenclature	Condition
oligozoospermia	Total number (or concentration, depending on outcome reported) of spermatozoa below the lower reference limit The lower reference limit for sperm concentration is 15×10^6 spermatozoa per ml.
teratozoospermia	Percentage of morphologically normal spermatozoa below the lower reference limit. The lower reference limit for normal forms is 4%.

Table 1-1: Nomenclature related to semen condition. Table adapted from WHO (2010)

1.5 The utility of spermatozoal RNA as a fertility assay and biomarker

Male infertility can be classified into testicular dysfunctions or idiopathic infertility. Testicular dysfunctions can be dived into: 1. pre-testicular, caused by genetic disorders represented in several forms of mutation in either autosomal or sex chromosome abnormalities like deletions on the Y-chromosome. Another example is hypogonadotrophic hypogonadism, which is associated with delayed puberty and impaired spermatogenesis. 2. testicular dysfunctions: like meiotic arrest of the spermatocytes or varicocele, which is a physical abnormality of the testis in growth and development leading to impaired spermatogenesis and therefore to reduced fertility. The 3. is post-testicular dysfunctions, e.g. obstructive azoospermia (OA) is the absence of spermatozoa in the semen caused by congenital absence of the vas deferens or functional obstructions of spermatozoa transport down the genital tract or post inflammatory obstructions of the ejaculatory duct (Jungwirth *et al.*, 2012). Idiopathic infertility is likely caused by the combination of a genetic predisposition and environmental factors and results in either idiopathic oligoasthenoteratozoospermia (low number of spermatozoa in the semen, poor movement and abnormal shape) or idiopathic infertility azoospermia (AZO) (count is equal or less than 15 million/ml in the semen) (WHO, 2010). See Table 1-1 for categories and nomenclature.

How can the testicular dysfunctions or idiopathic functions, which impair male fertility, be examined and understood, especially on a genetic and molecular level and how can

fathering a healthy child be ensured? Figure 1-4 shows a workflow used to diagnose different types of infertility and does not include tests on a molecular biological level. Before histopathological signs of infertility are visible showing either type of impaired spermatozoal production (see Table 1-1), changes occur at the molecular level, which may help elucidate the mechanisms of testicular dysfunction. If the ejaculate is examined and the testicular health questioned, it should be taken into account that parts such as the urethra could be affected through bacterial infections and be responsible for infertility.

If these factors are excluded, some reproducible prognostic tools and biomarkers to evaluate male factor fertility and testicular health are available. Spermatozoal quality can be observed partially by using a microscope in terms of spermatozoal numbers, concentration, motility, morphology and vitality. To assess chromosomal aneuploidies, chromatin structures and spermatozoal DNA integrity, more wide-ranging assays are needed e.g. terminal deoxynucleotidyl transferase (TdT)-mediated dUTP-biotin end-labeling (TUNEL), Comet assay, Acridine orange test, spermatozoa chromatin structure assay (SCSA). Karyotype analysis is the most common genetic tool to date to diagnose chromosomal abnormality associated with infertility and only used in cases of complete azoospermia (Kovac and Lamb, 2014). All of the assays in use today have their limitations and there is a need for deeper insight into infertility and accurate diagnosis, as well as for more robust diagnostic tests (Lefièvre *et al.*, 2007; Evenson and Wixon, 2006; Anton and Krawetz, 2012). The emergence and wider adoption of more accessible and improved techniques for this purpose; for example, high throughput RNA sequencing, provides more detailed and refined data and stands to offer us a far better molecular understanding of spermatogenic dysfunction (Anton and Krawetz, 2012; Mardis, 2008).

With the growth of microarray technologies and next-generation sequencing, investigating spermatozoa at a molecular level may simplify the non-invasive diagnosis of (at least) testicular infertility and might prove to be useful in the comparison of the mechanisms which result in function or dysfunction in spermatozoa (Miller *et al.*, 1999; Miller and Ostermeier, 2006a; Moldenhauer *et al.*, 2003; Ostermeier *et al.*, 2005b). As a result of medical advances, a testicular biopsy is not always necessary. There are promising molecular alternatives available to date. Spermatozoal RNA can be isolated from the ejaculate, and then examined in relation to spermatozoal quality and to testicular health. Ostermeier *et al.* (2002) and Zhao *et al.* (2006) used spermatozoal mRNA to generate a RNA profile for normal fertile men, providing an insight into the developmental history, functional stability and the potential spermatozoal-delivered elements which may be essential for male fertility and the subsequent support of embryo development (Sendler *et al.*, 2013; Jodar *et al.*, 2015). Some mRNAs are stably

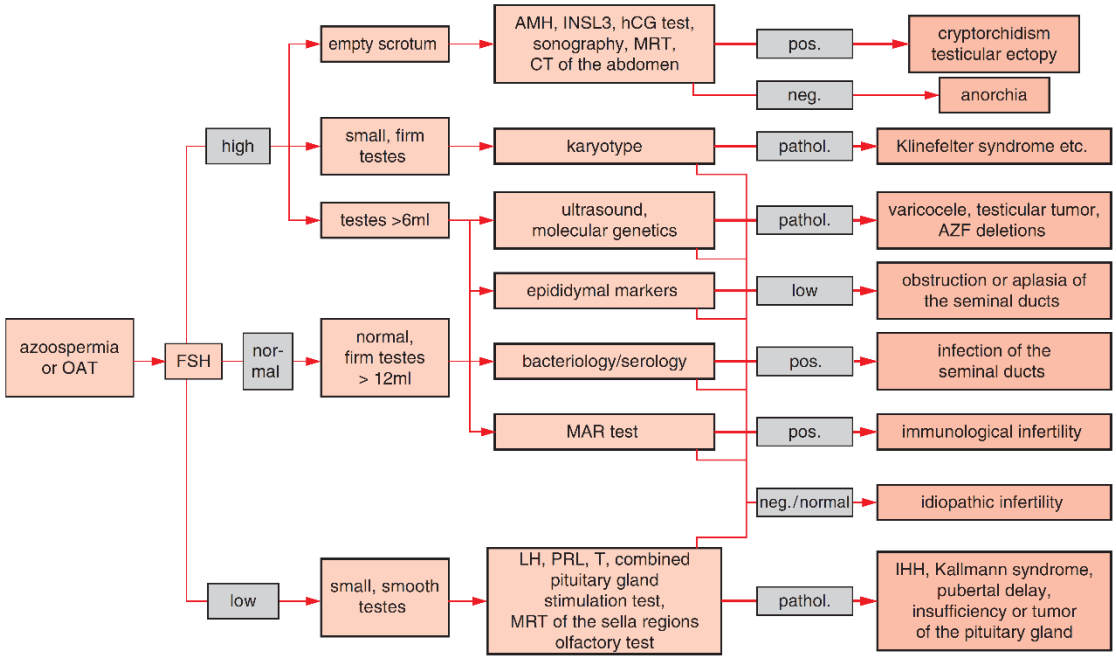


Figure 1-4: Diagnostic pathway to assess infertility. Figure adapted from Nieschlag *et al.* (2010).

regulated within and between individual males, making these a promising area for male fertility assessment and fertility assay development (Card *et al.*, 2013; Gilbert *et al.*, 2007; Lalancette *et al.*, 2008b).

Understanding the role of RNAs in spermatozoa may help to shed light upon factors regulating spermatogenesis that may be causes of male infertility (Jodar *et al.*, 2013). Both the miRNA and mRNA content can be isolated from semen samples, however, it is important to ensure that the samples are not contaminated with bacteria or somatic cells present in the ejaculate. mRNA provides a historical record of spermatogenesis which can be analyzed via microarray and Next-generation sequencing (NGS) based transcriptional profiling or qRT-PCR (Miller and Ostermeier, 2006a; Moldenhauer *et al.*, 2003). Montjean *et al.* (2012) showed that oligozoospermic infertile men have a downregulated pattern in genes important for spermatogenesis, spermatozoal motility, germ cell anti-apoptotic processes and genes involved in DNA repair, oxidative stress regulation and histone modification. Spermatozoal motility, which plays an important role in fertilisation was found either to be impaired or low in spermatozoa combined with specific RNAs e.g.: *PRM1* in human being upregulated in less motile spermatozoa (Carreau *et al.*, 2007). In contrast some miRNAs and coding RNAs were seen upregulated in porcine and bovine showing a high motility rate (Curry *et al.*, 2008; Ganguly *et al.*, 2013; Jodar *et al.*, 2013; Carreau *et al.*, 2007). Kempisty *et al.* (2008)

reported that the level of *Protamine 1* and *Protamine 2* were significantly lower in asthenozoospermic (reduced spermatozoal motility) men than in fertile controls and might be a potential diagnostic tool for male infertility. Additionally, microarray analysis examining long spermatozoal RNAs showed differences between fertile and infertile patients achieving pregnancy or not (Garrido *et al.*, 2009; García-Herrero *et al.*, 2011). García-Herrero *et al.* (2011), for example, used microarrays to compare spermatozoal transcriptomes obtained from samples which result in pregnancy after ICSI treatment and those that did not. Routine infertility investigation had failed to determine the infertility factors in the latter. These investigations revealed that the spermatozoal transcriptome differs significantly between the examined populations. Spermatozoal samples from the group which resulted in more successful pregnancies showed upregulation in genes involved in atrophy prevention in seminiferous tubules and support of spermatogenesis up to the pachytene stage of spermatocytes, genes involved in detoxication, which prevent tissue damage of testes by heavy metals (García-Herrero *et al.*, 2011). Platts *et al.* (2007) examined spermatozoal RNA profiles from teratozoospermic individuals and showed that the expression of RNAs involved in the ubiquitin-proteasome pathway, RNAs transcribed in late stages of spermatogenesis, acrosomal proteins and non-tubulin components of spermatozoal tails like *ODF 1-4* were reduced compared with spermatozoal samples from normozoospermic men. Wu *et al.* (2012) investigated the potential of seminal plasma as a biomarker and explored the occurrence of different miRNAs which play an important role in male fertility in patients with oligozoospermia and non-obstructive azoospermia compared to fertile controls. These miRNA profiles may offer an insight into spermatogemic potential of the testis and the post-testicular environment, including function of the accessory glands. Wu *et al.* (2012) suggested that a cluster of biomarkers for non-obstructive azoospermia would be a better diagnostic tool with much higher sensitivity, specificity and accuracy. They also described such biomarkers as “accessible through non-invasive protocols, inexpensive to quantify, specific to the disease of interest, and a reliable early indication of disease”. In this regard, non-coding RNAs and coding RNAs could both be effective and accessible biomarkers for this purpose. If validated, the establishment of these diagnostic methods may turn out to be an improvement and a step forward towards increasing the chances of successful pregnancy and predicting success in ICSI and IUI treatments, without using an invasive method.

Examination of spermatozoal RNA profiles may provide evidence of sample heterogeneity in human semen through their molecular ‘fingerprints’ (Miller and Ostermeier, 2006a). Such ‘fingerprints’ could be used to identify the causes of infertility (García-Herrero *et al.*, 2010).

As a biomarker of testicular health, spermatozoal RNA has significant implications for public health programs as a biomonitoring tool and may be used for identifying the origins, prognosis and potential treatment of various forms of infertility (Jodar *et al.*, 2013). Recently published work identified spermatozoal RNA elements which seem to be required to achieve live birth through intercourse or IUI, whereas these elements do not seem essential if pregnancies are induced through ART (Jodar *et al.*, 2015). These spermatozoal RNA elements could be used to assess male fertility and predict individual success rate for idiopathic infertile couples before undergoing assisted reproductive technologies, having the possibility to avoid invasive procedures for either male (testicular biopsy) or female (oocyte collection) partners (Jodar *et al.*, 2015).

In conclusion, the contents of the ejaculate are a good proxy not just for the germ cell component of the testis and the testicular health, but also to examine factors critical for fertilisation and successful embryo development and the epigenetic impact on the birth of healthy offspring (Miller and Ostermeier, 2006a; Jodar *et al.*, 2013; Anton and Krawetz, 2012).

1.6 Thesis Hypothesis

The study hypothesis is that profiling spermatozoa gene expression patterns (and networks) that are associated with a fertile phenotype across various species, will enable us to identify key genes that are most likely to be altered in infertile phenotypes. Furthermore, a comparison of RNA profiles (RNA types and abundance) from different species will assist the search for likely functions of spermatozoal RNA beyond spermatogenesis itself. These data could ultimately lead to novel, RNA based, diagnostics for infertility, and could identify potential therapeutic targets to reduce infertility.

1.7 Thesis Aims

To address the hypotheses, a next-generation RNA sequencing approach was used to explore and analyse the RNA isolated from bovine, ovine, porcine and human spermatozoa and also from testis tissues from these species.

In addition, experiments aim at following the fate of spermatozoal RNA in the embryo to strengthen the hypothesis of paternal RNA playing a role in the developing embryo.

Aim 1: To compare and characterise the total sperm RNA composition from bovine, ovine, porcine and human species using next-generation RNA sequencing and to look for common gene expression patterns and networks.

Aim 2: To explore whether spermatozoal RNA has potential functional roles after fertilisation, by examining whether selected transcripts, including the key fertility indicators identified in aim 1, are found in the early embryo. Specifically, the embryonic fate of sperm RNA in bovine embryos will be examined by RT-qPCR.

Aim 3: To examine whether methods that are routinely used in IVF laboratories, alter transcriptional profiles of human spermatozoa and are improving the selection of the right candidate spermatozoon. This will be done by i) comparing the total transcriptional profile of fresh and frozen human spermatozoa, and ii) examining whether different RNA profiles are found through hyaluronic selected spermatozoa and non-selected spermatozoa.

Aim 4: Determine if the spermatozoal protein MOSPD3, has the potential to be used as marker of fertility.

Chapter 2: General Materials and Methods

Chemicals, reagents, materials and all company addresses are detailed in Appendix IV. Buffer compositions and solutions can be found in Appendix II.

2.1 Spermatozoa handling

2.1.1 Spermatozoal suppliers and ethical approval

Frozen bovine semen was obtained from Genusbreeding (Nantwich, Cheshire, UK; www.genusbreeding.co.uk). Frozen porcine semen was bought from JSR Genetics (Southburn, Driffield, UK; http://www.jsrgenetics.com/) and frozen ovine semen from Innovis®Ltd (East Mains, Ormiston, East Lothian, Scotland; www.innovis.org.uk). Human spermatozoa were obtained from the IVF unit at Seacroft hospital, Leeds, UK, or from donors at the University of Leeds, UK, following standard ethical approval and consenting guidelines. The study was considered and nationally approved by the relevant UK Integrated Research Application System (IRAS) ethics committee (NRES 12_NE_0192) on 13th January 2013 and locally approved by the University of Leeds' School of Medicine Research Ethics Committee (SoMREC/13/017) on 28th November 2013.

2.1.2 Frozen storage of human spermatozoa

One volume of Quinns Advantage™Sperm Freezing Medium was added dropwise and very slowly over a 30-second period to one volume of liquefied semen sample. After each drop, the solution was mixed thoroughly to allow equilibration with the cells. 1 ml of the solution was placed into cryotubes and the mixture was equilibrated for 10 minutes before storing the tube above liquid nitrogen in the vapour phase for 30 minutes. After 30 minutes, the tubes were transferred quickly into liquid nitrogen for long-term storage.

2.1.3 Thawing of human and animal spermatozoa

Aliquots of semen were rapidly thawed and then maintained for 20 mins at 37°C to fully liquefy before usage.

2.1.4 Density gradient centrifugation of bovine, ovine and porcine spermatozoa

Two different layers of Percoll concentrations were placed into a 15 ml polypropylene tube as follows: 1.5 ml of 90% Percoll (see Appendix II, Table II-1 and Table II-2 for recipe) was placed on the bottom and carefully overlaid with 1.5 ml of 45% Percoll (90% Percoll diluted

in spTalp). After the semen had thawed, it was carefully and slowly pipetted above the gradient. Samples were pelleted by centrifugation at 4°C, 700 x g for 30 mins. Immediately after centrifugation the samples were placed on ice. To avoid somatic cell contamination, the supernatant was removed very carefully without touching the edges of the tube and leaving 200 µl supernatant behind. Furthermore, a new pipette tip was used to collect the pellet carefully. Pellets, which contain the better quality spermatozoa (motile), were washed with 1x spTalp and centrifuged twice at 900 x g for 10 mins each. Pelleted spermatozoa were re-suspended into 1 ml 1x spTalp (see Appendix II, Table II-3) and an aliquot counted using a Neubauer chamber to determine the concentration (millions/ml). The counted number was multiplied by the used dilution and 10^4 to calculate the concentration in 10^6/ml. Examples of the approximate amounts for 1 straw containing semen can be seen below in Table 2-1.

Species	**Concentration**
Bovine	2-7 x 10^6
Ovine	10 x 10^6
Porcine	5 x 10^6

Table 2-1: Spermatozoa concentration of one straw, after gradient centrifugation

2.1.5 Density gradient centrifugation of human spermatozoa

Human spermatozoa were processed using a two layered Supra Sperm gradient: 1.5 ml of 90% Supra Sperm was overlaid with a 1.5 ml layer of 45% Supra Sperm or with a three different layer Percoll gradient (90%, 60% and 45%) in a polypropylene tube before the liquefied ejaculate was placed carefully and slowly on top. The Supra Sperm gradient sample was pelleted by centrifugation at 300 x g for 20 mins and washed twice (centrifugation at 300 x g for 10 mins) with Sperm Preparation Medium (according to the manufacturer's protocol), whereas the Percoll gradient sample was centrifuged at 400 x g and washed with spTalp twice before the number of the cells was counted (see above). All centrifugation steps were carried out at 4°C.

2.2 *B. taurus* testis tissue homogenisation

50 mg of bovine testis (locally sourced from the abattoir, JC Penny and Sons, Rawdon, Leeds, UK) was homogenised using a sterile liquid nitrogen filled mortar and a liquid nitrogen chilled pestle. It was made sure that the mortar was always filled with liquid nitrogen and that the tissue was homogenised until the consistency turned to powder. The powder was filled into pre-chilled microcentrifuge tubes afterwards. The homogenized tissue was used directly for a Trizol® based RNA extraction or stored at -80°C.

2.3 Sample slide preparation for microscopic analysis

2.3.1 DNA staining with 4',6-diamidino-2-phenylindole (DAPI)

4',6-diamidino-2-phenylindole (DAPI) stains were prepared for each sample to exclude somatic cell contamination. DAPI was diluted in PBS following the manufacturer's instructions, to a concentration of 0.5 µg/ml. 100 µl of the diluted DAPI was added to the pelleted spermatozoa and mixed by pipetting. The sample was incubated at room temperature for 30 mins. After incubation 1 ml of 1 x PBS was added, and centrifuged for 10 mins at 1,500 x g. The supernatant was discarded and the pellet was dried. Finally, the pellet was suspended into the remaining liquid, put on a glass slide and covered with mounting media and a cover slip before examination under a fluorescence microscope was performed.

2.3.2 Whole cell staining with Giemsa

To monitor for somatic cell contamination, Giemsa stains were prepared for each sample. Slides for the Giemsa staining were prepared by centrifuging spermatozoa at 1500g for 15 minutes onto poly-l-lysine coated slides using a cytospin centrifuge (Cytospin3, Shandon) and dried overnight. Slides were washed briefly in methanol and a Giemsa stain diluted into distilled water (ratio: 1:10) was applied to the slides for 30 minutes. Washing with distilled water was performed afterwards and a freshly prepared Giemsa solution was applied for further 30 minutes. Another wash with distilled water took place before the slides were air dried covered with mounting media and a coverslip, before examination using a bright-field microscope took place.

2.4 General molecular analysis

All reagents were prepared in 0.1% Diethylpyrocarbonate (DEPC) water and all work was carried out followed the RNAse free guidelines using RNAseZap, filter tips and RNAse/DNAse free equipment (Nielsen, 2011; Ambion®, 2012).

2.4.1 RNA extraction methods

2.4.1.1 Total RNA extraction using modified Trizol®

1 ml Trizol® reagent and 10 µl β-mercaptoethanol were applied to a maximum of 100 x 10^6 cells after gradient centrifugation and counting. The solution was transferred to a safe lock tube containing 100 mg of 0.5 mm sterile stainless steel beads, however, to promote homogenization the solution was passed 10 times consecutively through a 26G needle and incubated on a heat block at 65°C for 45 minutes. Every 10 mins the tubes were vortexed with the Disrupter® Genie (Scientific Industries) shaker for 1 min. 0.2 ml (per 1 ml of Trizol) of chloroform was added and the solution was thoroughly shaken manually for 15 seconds until turning “pink and milky”. Additionally, an incubation at room temperature was performed for 3 minutes. The solution was then centrifuged (Centrifuge 5415R, Eppendorf) at maximum speed of 16,000 x g at 4°C for 15 min. 450 µl to 500 µl of the RNA-containing upper phase was carefully drawn off without touching any other phase and transferred into a new 1.5 ml reaction tube. An equal amount of isopropanol plus 1 µl of Glycoblue was added, the sample mixed by vortexing and, to promote RNA precipitation, the samples were stored at -20°C overnight. The next day, the RNA was pelleted by centrifugation (Centrifuge 5415R, Eppendorf, Hamburg) at 16,000 x g at 4°C for 15 minutes. The isopopanol was carefully removed and the pellet washed with 0.5 ml 70% RNAse-free ethanol (-20°C), centrifuged twice at 16,000 x g for 5 minutes at 4°C. Between the two centrifugation steps, the ethanol was carefully drained with a pipette. As much ethanol as possible was removed and the pellet was air-dried for 5 minutes. After drying, RNA was dissolved in RNAse-free water and stored at -80°C. Testis RNA concentrations were measured spectrophotometrically (Spectrophotometer Nano Drop ND-1000), spermatozoal RNA quantity and quality was determined with the RNA 6000 Pico Assay using the Agilent 2100 Bioanalyzer (Agilent Technologies). If needed, the RNA was concentrated or cleaned using RNeasy Mini Kit columns and the clean-up protocol (QIAGEN) according to the manufacturer’s instructions.

2.4.1.2 RNA extraction using a modified Trizol® and column based method

500 µl of RLT buffer (RNeasy MiniElute Kit, QIAGEN), 7.5 µl β-mercaptoethanol and 100 mg of stainless steel beads were added to a maximum of 100 x 10^6 pelleted spermatozoa

for homogenisation. The tube was placed in a Disruptor®Genie and vortexed for 5 minutes before 500 µl of Trizol® reagent was added and vortexed a second time with the Disrupto®Genie for 5 minutes. To promote homogenization, the solution was homogenized 10 times using a 26G needle, before 200 µl of chloroform was added for phase separation. The mixture was shaken briefly by hands for 15 seconds and sat at room temperature for 5 minutes before the solution was centrifuged (Centrifuge 5415R, Eppendorf) at 4°C at 12,000 x g for 20 minutes. The upper aqueous layer, containing RNA, was carefully removed without touching any other layer and placed into a 2 ml polystyrene tube. For each 500 µl recovered upper aqueous phase, 360 µl of 100% ethanol was added and mixed in. A maximum of 700 µl of the mixed solution was added to a RNeasy Mini Spin column and centrifuged for 30 seconds at 16,000 x g (Centrifuge 5415R, Eppendorf), this process was repeated until the whole aqueous phase ethanol mixture was centrifuged through the column and between each centrifugation the collection tube was emptied. Washing of the column took place adding 700 µl of RW1 buffer, centrifuging the column at for 1 minute afterwards. If required, a DNA digestion was added at this point and followed according to the manufacturer's protocol using the column based DNase I method. Two washes with 500µl RPE buffer and centrifugation at 16,000 x g for 2 minutes took place. After the wash, a drying step took place. The columns were centrifuged at 16,000 x g for 2 minutes, without adding any solution. At this step the collection column was replaced by a fresh collection tube for RNA elution. 25 µl or 100 µl of RNase free water was added to elute the RNA and the column was equilibrated for 5 minutes, before centrifugation at 16,000 x g for 1 minute took place.

2.4.2 RNA and cDNA quantification and quality control

2.4.2.1 Bioanalyzer

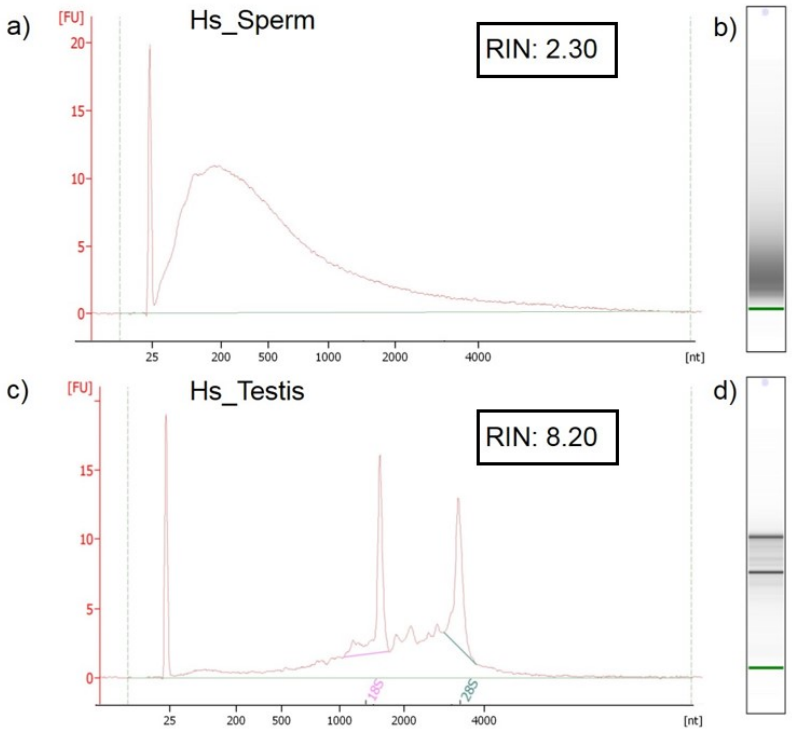


Figure 2-1: Bioanalyzer Electropherogram of Sperm and Testis RNA. a) human spermatozoal RNA, clustering under 500 nt indicates RNA degradation; b) gel picture of spermatozoal RNA. Spermatozoal RNA is visible in the lower range of the gel confirming the low molecular sizes of the RNA. If no 18S and 28 S rRNA was visible and the RIN was around 2, then the sample was counted as pure spermatozoal population without any somatic cell contamination and was used for further experiments; c) electropherogramm of human testis RNA; d) gel of human testis RNA, both showed 18S and 28S rRNA and a RIN of 8.20. A good quality RNA sample should show a RIN around 8 or more. FU=Fluorescent Unit, nt=nucleotides

A quality and somatic cell control was performed using the “Agilent Pico Chip 4000” which was run on an Agilent Bioanalyzer (Figure 2-1). If no 18S and 28S rRNAs were detected in spermatozoal RNA and the RNA Integrity Number (RIN) was around 2 or lower, then the sample was used for further analysis. No intact 18S and 28S rRNAs should be detected in spermatozoal RNA preparations, which contain degraded rRNA. The RIN is calculated using the 18S and 28S rRNA and a low RIN indicates good spermatozoal RNA quality (Cappallo-Obermann *et al.*, 2011; Shafeeque *et al.*, 2014). In contrast, good quality testis samples should show a RIN of 8 or more (Das *et al.*, 2010). RNA quantity was measured using a

Qubit. If the RNA passed all quality checkpoints, and the quantity was sufficient, NGS-RNA libraries were prepared.

The RNA qualification and quantification was carried out following the manufacturer's instructions (version: 2002) using the RNA Pico Kit 6000. cDNA was quantified and qualified using a High Sensitive DNA Kit (manufacturer's instructions 2009). The procedure was carried out following the manufacturer's protocol.

2.4.2.2 Nanodrop

Testis RNA was quantified applying 1 µl at the spectrophotometer and measuring the wavelength at 260 nm (Nano Drop ND-1000; Thermo Scientific).

2.4.2.3 Qubit

Quantification was achieved following the manufacturer's protocol using the 'Broad-Range' and 'High-Sensitive RNA' and 'High-Sensitive DNA' Assay. The RNA quantity was measured using a Qubit®2.0 'Fluorometer' (ThermoFisher Scientific) for low RNA quantity samples and cDNA and dsDNA for next-generation sequencing library creation

2.4.3 DNAse treatment

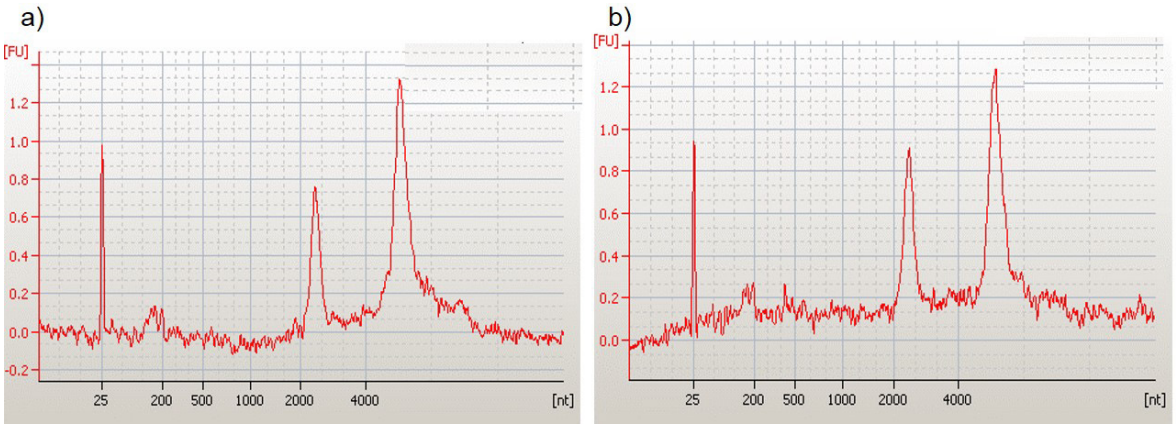


Figure 2-2: Bioanalyzer electropherogram of a) DNase treated bovine testis RNA b) untreated bovine testis RNA. If it was proven that the new DNase is RNase free, then residual DNA treatment using DNase was performed. FU=Fluorescent Unit, nt=nucleotides

All used DNase batches were tested for residual RNAse activity before usage, by incubating bovine testis RNA with every new batch of TURBO®DNase and running the digestion products on the Bioanalyzer to check the integrity of 18S and 28S rRNA. If both rRNA subunits were found intact, then the DNase was considered RNAse free and could be used according to the manufacturer's protocol to digest residual DNA (Figure 2-2). In brief: The 10 x buffer for the DNAse was adjusted to 1x buffer and added directly to the sample, before 1 µl of DNAse was added and the mixture was kept at room temperature for 15 minutes. For experiments reported in Chapter 3 and 5, the reaction was stopped by adding 15 mM EDTA and the reaction mix was incubated at 75°C for 10 minutes. cDNA was stored at -20°C for later follow up RT-PCR or RT-qPCR analysis. All steps were performed using a Veriti™ thermal cycler (ThermoFisher Scientific).

2.4.4 Reverse transcription for cDNA synthesis

RNA was reverse transcribed into cDNA using the protocol below.

Primers, dNTPs and the RNA (Table 2-2) were incubated at 65°C for 5 min, chilled on ice for 1 min and centrifuged at 16,000 x g before the second set (Table 2-3) of components was added. Since the RNA concentration was too low, no concentration could be measured and the maximum input volume of 9 µl was used for the RT reaction.

Component Set 1	**Volume per sample**
Oligo$(dT)_{27}$ (10 mM)	2 µl
Random Hexamer Primers (10 mM)	2 µl
dNTPs (10 mM)	1 µl
RNA	9 µl

Table 2-2: cDNA synthesis components set 1

Component Set 2	**Volume per sample**
5x Reaction Buffer	4 µl
DTT (0.1M)	1 µl
RNAse Block (40 U/µl)	1 µl
Total volume	20 µl

Table 2-3: cDNA synthesis component set 2

2.4.5 RT-PCR

gDNA contamination was analysed via RT-PCR for the first replicate set of NGS library samples in Chapter 3 and 5 using the protocol below (Table 2-4 and Table 2-5) according to Gilbert *et al.* (2007). The RNA concentration was too low to measure and therefore the maximum amount available was used in the RT-PCR reaction. For all PCRs a positive testis cDNA control according to the used species (bovine, ovine, porcine or human) and a negative water control was included.

Component	**Volume per sample**
cDNA	1 µl
RNAse/DNAse free water	17.55 µl
Reaction Buffer	2.5 µl
dNTPs (10 mM)	1 µl
$MgCl_2$	0.75 µl
Forward Primer (10 mM)*	1 µl
Reverse Primer (10 mM)*	1 µl
Taq Polymerase	0.2 µl

Table 2-4: PCR reaction set up. See primer sequences in Chapter 4

Step	**Time**	**Temperature**
	1 cylce	
One cycle melting	5 minutes	95°C
	35 cylces	
Melting Step	30 seconds	95°C
Annealing	30 seconds	60°C
Elongation	45 seconds	72°C
	1 cylce	
Final one Cycle	10 minutes	72°C

Table 2-5: PCR cycle set up

2.4.6 RT-qPCR

To control for gDNA contamination, RT-qPCR was used for the second replicate set of NGS library samples described in Chapter 3 and 5 and for all samples. RT-qPCR (see protocol in Table 2-6) is a more sensitive method than a PCR and ideal to use for gDNA contamination controls. The cycle number was adjusted from the usual 40 up to 50 cycles (Table 2-7), since gDNA contamination may be small and could appear in later cycles.

If no gDNA was detected, RNA was reverse transcribed into cDNA (see Figure 2-3a). cDNA in Chapter 5 was analysed via RT-qPCR in order to test the quality of the cDNA (see Figure 2-3 b). If a control gene (e.g.: *GAPDH* or *PRM1*) was amplified using cDNA, then the original RNA was used for further experiments.

Component	Volume per sample
SYBR Green	5 µl
Primer Forward and Reverse 10-20 mM	0.5 µl
cDNA /DNA	1 – 2.5 µl
RNAse/DNAse free Water	Up to 10 µl
Total Volume	10 µl

Table 2-6: qPCR components

Step	Time	Temperature
1 cycle		
One cycle melting	10 minutes	95°C
Melting Step	30 seconds	95°C
Annealing	1 minute	60°C
Elongation	45 seconds	72°C
Final one Cycle	10 minutes	72°C

Table 2-7: qPCR cycles

Time	Temperature
1 minute	50°C
1 minute	95°C
1 minute	60°C
15 seconds	95°C
1 minute	60°C
10 minutes	40°C

Table 2-8: Melting curve settings

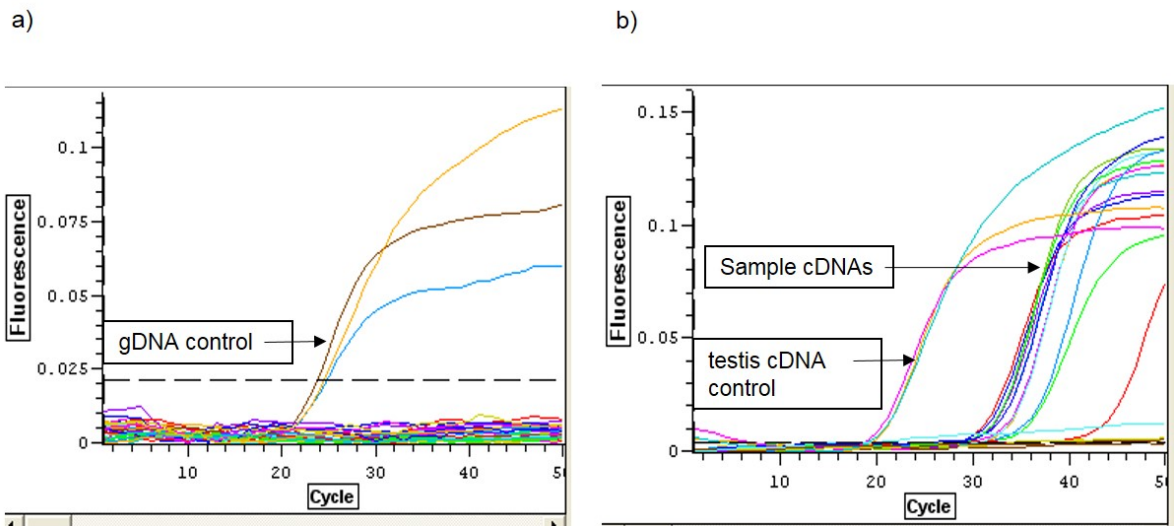


Figure 2-3: gDNA contamination and cDNA quality control: a) every RNA was checked for gDNA contamination, therefore up to 50 amplification cycles were run to control gDNA contamination. If no signal for gDNA contamination was detected then the RNA was converted into cDNA. The visible signal shows the positive control and confirms that the qPCR set up was working. If further signals were detected, the RNA was treated a second time using DNAse for DNA digestion b) the quality of every cDNA was controlled, using RT-qPCR. The signal at cycle 20 is showing amplification of the gene of interested (*PRM1* or *GAPDH*) using testis cDNA as positive control, whereas signals at a later amplification cycle show if the used spermatozoal cDNA was enough to amplify e.g. *PRM1*. If the cDNA showed a signal, then the original RNA was used for further experiments.

2.4.7 Agarose gel electrophoresis

A 2% (w/v) agarose gel was prepared to visualise either the RT-PCR or RT-qPCR product. The gel was mixed with Gel Red (1:100000) and the product was mixed with 1 µl of 10x loading buffer (Table II-4) and loaded onto the agarose gel. The gel was run in a Sub-Cell® GT connected to a PowerPac 300 (Bio-Rad Laboratories Ltd.), using a 1x trisaminomethane, borate, EDTA (TBE) electrophoresis buffer (Table II-5) at 90-100 V for 30-60 min. The gel was visualised in a Molecular Imager®Gel Doc™XR+ System (Bio-Rad Laboratories Ltd.).

2.5 RNA-Seq analysis using Bioconductor

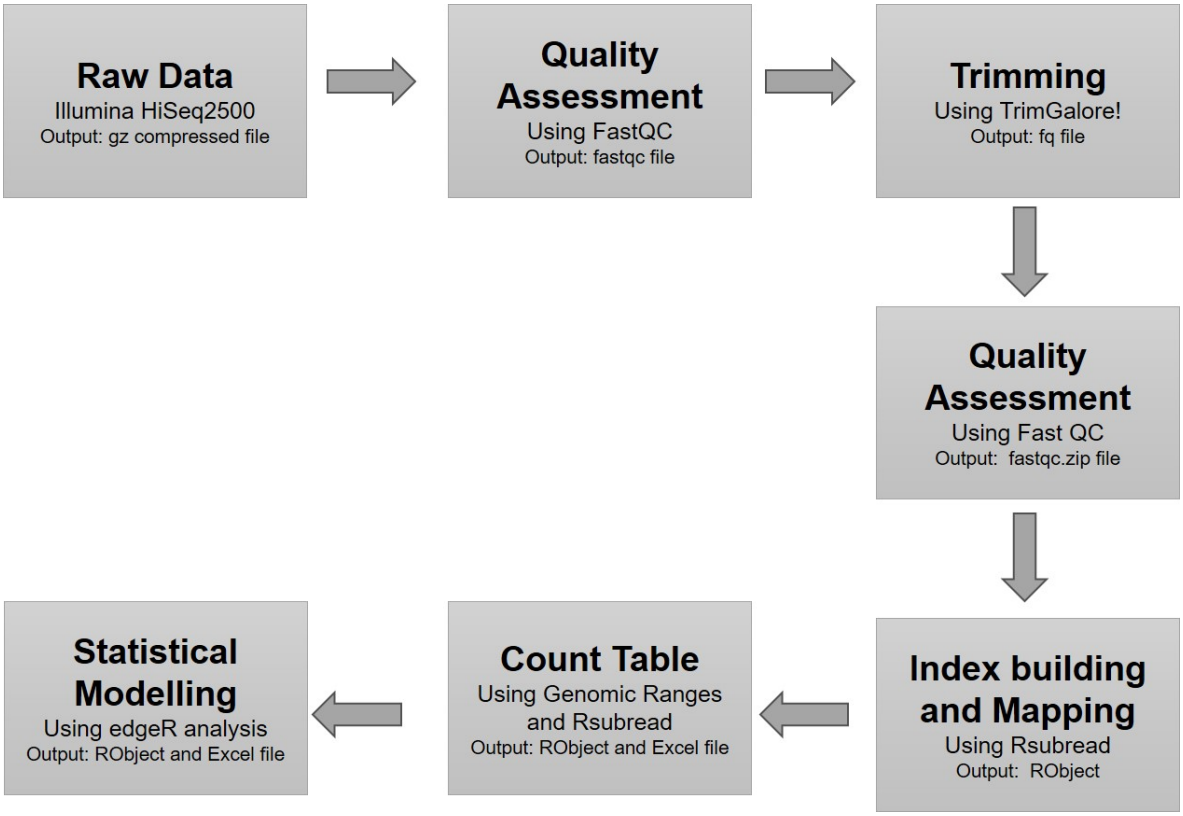


Figure 2-4: RNA-Seq analysis workflow

All steps from the initial sequencing to the final gene lists were performed for Chapter 3 and 5. For data processing, the R environment and Bioconductor packages were used throughout, except where stated otherwise (Gentleman *et al.*, 2004). R commands are provided in Appendix I. A general schematic of the process and tools used is shown in Figure 2-4 and explained for each step in more detail below.

2.5.1 Quality assessment and trimming

High throughput sequencing data output from the Illumina HiSeq2500 instrument needs to be quality controlled to verify in the first instance whether the sequencing run performed as expected. Obvious problems e.g.: PCR artefacts, presence of contaminating sequences/adapter sequences or any bias that might have been introduced through the usage of random hexamer primers, which may influence the uniformity of the reads along the transcriptome had to be controlled for (Hansen *et al.*, 2010). 'FastQC' (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which was used as a quality control tool, can be executed from the command line or as a graphical application (Andrews,

2010). All bars in the orange or red region indicate a bad sequencing run (Figure 2-5a), which can be improved through the usages of bioinformatical approaches e.g. FastQC. “In general it is common to see base calls falling into the orange region torwards the end of a read. The background of the graph divides the y axis into very good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red)” (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/3%20Analysis%20Modules/2%20Per%20Base%20Sequence%20Quality.html.).

In general, all the RNA-Seq data met the high quality standard required for downstream analysis, which indicates a successful run and library creation. However, as shown in Figure 2-5a) our paired end reads showed an expected adapter sequence bias at the first 13 bp, introduced through the usage of random hexamer primers and needed to be trimmed.

2.5.2 Trimming

‘TrimGalore!’ (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; Babraham Bioinformatics) was used to trim adapter sequences off using the command line. It is a script used to automate adapter trimming and quality control (Krueger, 2013). The program trims the first 13 bp of both end reads with Illumina adapter sequences by default. Compressed ‘gzip’ files can be used as an input and another ‘FastQC’ can be run on the resulting files. Figure 2-5b) shows the sequences after trimming the first 13 base pairs off and an improvement of the sequencing data, which is visible in the right panel of Figure 2-5b). All bars and error bars are within the green region (good) compared to the data before any trimming took place and the quality of the reads is improved.

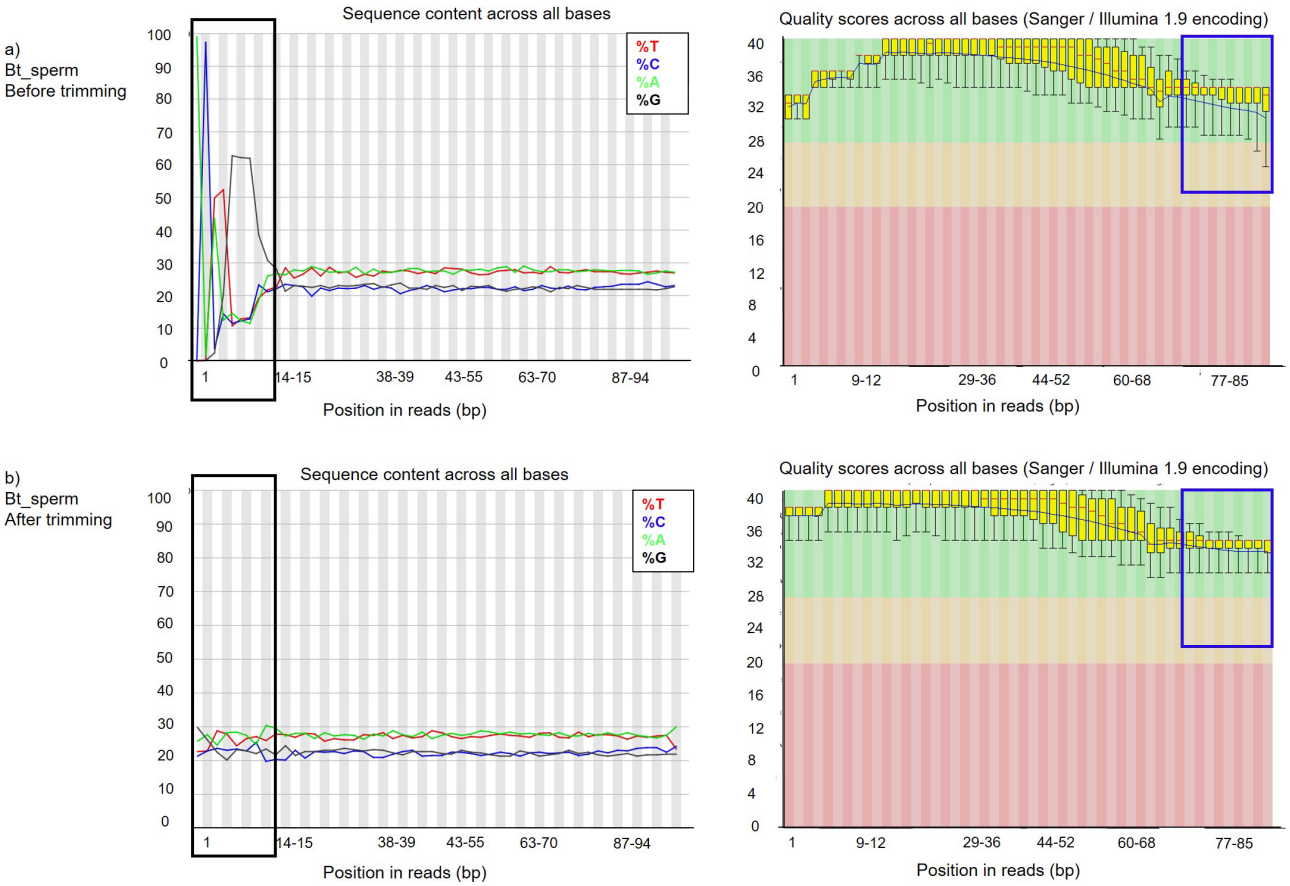


Figure 2-5: FastQC examples. a) Quality control directly after sequencing. Adapter sequences arising from random hexamer usage were detected (black box) and needed to be removed using TrimGalore! b) Quality control and improvement (blue boxes) after trimming the adaptors (black box). Good quality calls (green region), calls of reasonable quality (orange region), and calls of poor quality (red region).

2.5.3 Building indices using RSubread

Using 'RSubread' an index for the reference genome of each species can be built (all commands were used from the 2014 manual) (Liao *et al.*, 2013; Shi, 2014; Shi and Liao, 2016). The full genome indices data set can be extracted as a 'fasta' file from the University of California, Santa Cruz (UCSC) genome browser' (https://genome.ucsc.edu/) and is downloaded in order to build indices for alignments to the transcriptomic and genomic data of each species. The 'fasta files' include all reference sequences for each species (Shi and Liao, 2016). 16 bp mer sequences are extracted for each reference genome, using a 2 bp interval, building a hash table. "Keys in the hash table are unique 16 bp mers and values are their chromosomal location" (Liao et al., 2013; Shi and Liao, 2016). 16 bp mers were used since Shi and Liao (2016) reported it as an optimal range for sensitivity and accuracy. 'RSubread' has a reasonable specificity and to work powerfully all highly represented and uninformative subreads need to be removed. The '*buildindex'* functions, which implement the 'RSubread' index builder was used and the default was set to remove highly repetitive and uninformative 16 bp mer subreads (Liao *et al.*, 2013).

2.5.4 Gene mapping

Species	Genome Assembly	Released Year
H. sapiens	hg38	December 2013
B. taurus	bosTau8	June 2014
O. aries	Oar_v3.1	August 2012
S. scrofa	susScr3	August 2011

Table 2-9: Used genome assembly datasets, extracted from UCSC or !ensemble

After building the index for each species, all reads need to be mapped to their reference sequence. The 'RSubread' aligner was chosen the method using the '*align'* function to extract a number of overlapping seeds from each read, called subread. These subreads are used to 'vote' for the mapping location of each read. "The region receiving the largest number of 'votes' is selected as the final mapping location" (Liao *et al.*, 2013). 'RSubread' can generate a gene-level count summary. As paired end reads were considered, the former two paired-end 'fq files', for each sequencing direction were converted into one output BAM file aligned to the sequence of each respective species (see Table 2-9). To map the trimmed

sequences to the respective species the genome assembly datasets in Table 2-9 were used as a base.

2.5.5 Genomic Ranges

The tool ‘Genomic Ranges’ needed to add biological semantics to the sequencing metadata, including the treatment of sequence name and strand (Lawrence *et al.*, 2013). It serves as a basis for representing genomic ranges, genomic positions and groups (Carlson *et al.*, 2016). Using the ‘GenomicRanges’ function allows to place gene annotations and chromosome numbers in the correct order for generating the input file required for feature counts ('GeneID', 'Chr', 'Start', 'End', 'Strand'). Furthermore, ‘Genomic Ranges’ is needed to create a 100 bp tile path genome for the ‘*featureCounts’* input. The tiles are constructed so that they overlap and are distributed along the entire genome. The package and codes used were written by Aboyoun *et al.* (2014).

2.5.6 Counting mapped reads using RSubread´s feature counts

‘RSubread´s *featureCounts’* is used to “assign mapped reads to genomic features such as exons [features], genes [meta-features], promoters, gene bodies and ‘genomic bins’, within each bin all features are grouped into blocks” (Liao *et al.*, 2013). Using ‘*featureCounts’* requires a unique annotation file for the species of interest, including features for each species’ genome. Counting meta-features is useful for expression level analysis since reads for genes (not exons) are counted, which can be clustered. Multiple overlaps were allowed, which means that one read can overlap more than one feature. For the full human, bovine and porcine ‘RefSeq’ data, all chromosome and sequence information was downloaded from the ‘UCSC genome browser’, whereas ‘Ensembl’ (http://www.ensembl.org/index.html) was used for the ovine data. Since the structural annotations for the genomes of non-model organisms are incomplete to date, annotation files for bovine and porcine were built by combining the human and ‘xeno’ reference (available at UCSC genome browser) together with the most recent annotation of the respective species. The ovine annotation was completely re-built, using the available human and ‘xeno’ data combined with NGS data produced for this thesis and other sequencing projects in the group to determine and discriminate the best candidates for novel genes in the ovine. Regions with more than 1 read vote were put into 100 bp so called ‘bins’. All alignments were counted against 100 bp bins to determine the uniformity of coverage of each transcript. Duplicate reads and reads with 2 mismatches were ignored. The output is the likelihood that each bin represents a peak, with adjoined bins being combined to create regions. For each region, all the reads aligning in that region were identified and then the read 1 associated with each read 2 was

determined (Collins *et al.*, 2015). The output is a gene list, containing the expression level of each transcript, assigned to the correct gene. A table and a 'BED' file ca be generated, latter mentioned can be uploaded to a genome browser to visualise the counted reads paired with the right gene. Further ontological analysis (GO) and statistical analysis can be done with the output of feature counts.

2.5.7 edgeR anaylsis

The 'edgeR' package (http://bioconductor.org/packages/release/bioc/html/edgeR.html) is a framework for statistical analysis of read counts derived from digital gene expression technologies (Robinson *et al.*, 2010; McCarthy *et al.*, 2012). 'edgeR' compares read counts from a transcript under differing experimental conditions or different tissue types (Robinson *et al.*, 2010). The algorithm calculates the genewise distribution by conditional maximum probability, depending on the total count for a gene (Robinson *et al.*, 2010; Smyth and Verbyla, 1996). Pairwise comparison using 'edgeR' is possible (Robinson *et al.*, 2010; Smyth and Verbyla, 1996). The input for 'edgeR' is a 'featureCounts' generated table of counts, which includes the same gene annotation in both datasets. According to the questions being asked or the analysis to be carried out, either the classical linear model (GLM) or ANOVA can be used. The package is detailed in Chen *et al.* (2014).

2.5.7.1 Building a DGEList

To store and manipulate data in 'edgeR', a data object (DGEList) must be created using a data frame or matrix (Chen *et al.*, 2014).

2.5.7.2 Data exploration and multidimensional scaling (MDS) plot

The first analysis of the raw dataset should be a MDS to explore the data for outliers and their relationships to both the replicates and different treated samples. The data can be plotted using multidimensional scaling (MDS) to determine the 'distances' between samples. Ideally similar samples should show less variability than samples that differ markedly and should cluster accordingly (see Figure 2-6).

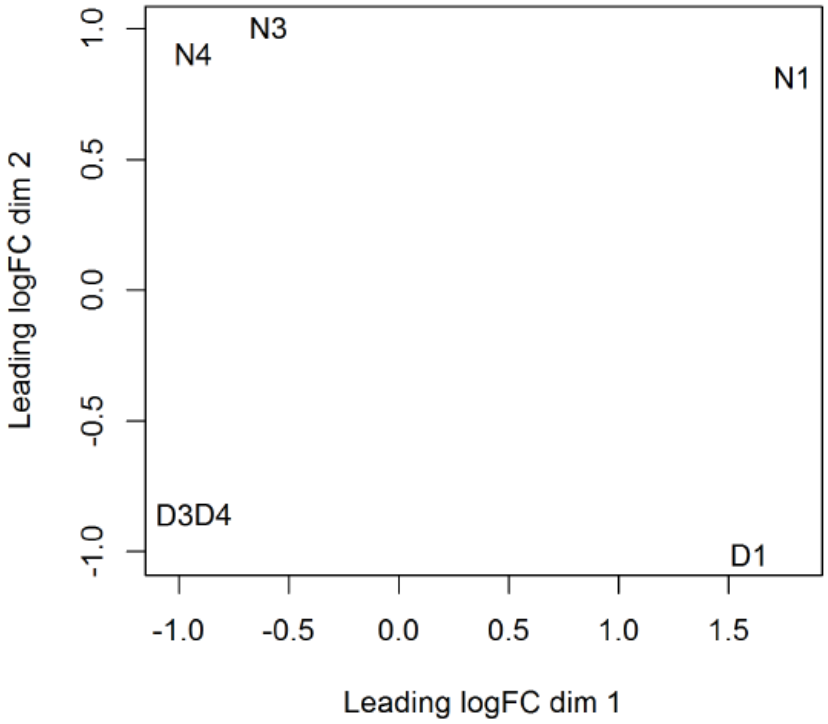


Figure 2-6: MDS plot. The closer samples cluster together the less biological or technical difference is seen between the replicates and the different used tissues. Figure adapted from the edgeR manual.

2.5.7.3 Differential Expression analysis

Having strong parallels with the Fisher's t-test, the classical approach for pairwise comparison between different conditions and groups using the exact test was used for our interspecies comparison (Chapter 3). The GLM approach for multiple groups having different treatments was used for human sample pairwise comparisons in Chapter 5. Differentially expressed (DE) genes can be graphically represented using a smear plot, where the $\log_2$-fold-change for each gene is plotted against average abundance. Significantly, differentially expressed genes, are plotted with a false discovery rate (FDR) of 5% and are highlighted in red in the smear blot. In the RNA expression comparison shown in Figure 2-7, all positive 'logFC' values, indicate upregulated RNAs for one group whereas all negative 'logFC' values indicate upregulated RNAs for the different treated group or tissue of comparison. RNAs not DE to both sources are shown in black.

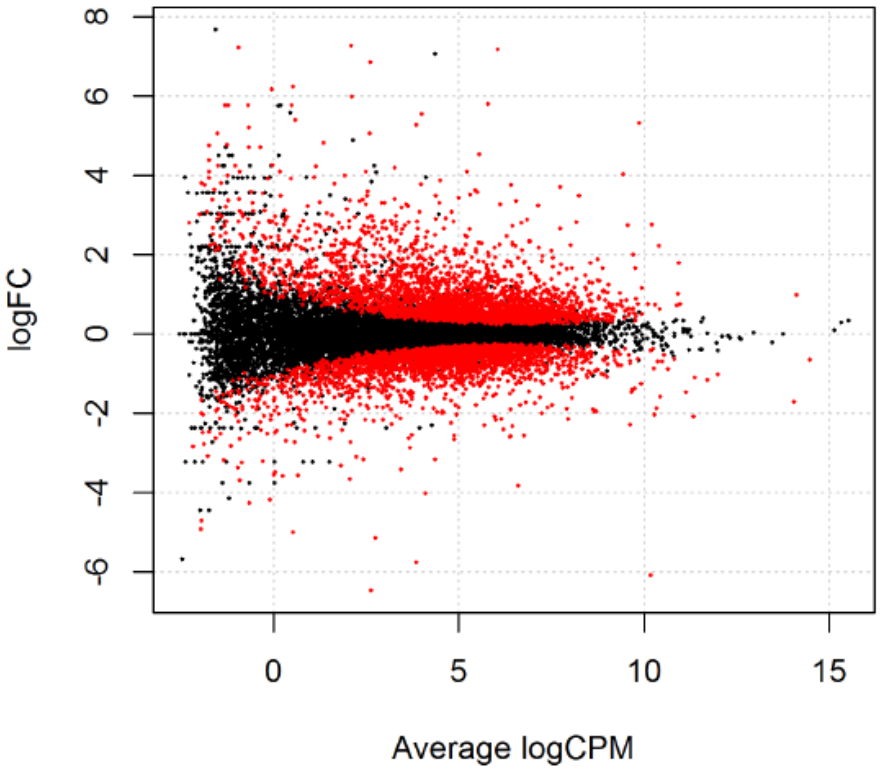


Figure 2-7: Smear Plot for DE transcripts. Red dots indicate significant differential expression in each tissue type or condition. Negative values are indicating the expression in one tissue and positive values in the other. Black dots are genes in common, but not significantly expressed. Figure adapted from the edgeR manual.

2.5.8 Ontological analysis

2.5.8.1 Cytoscape

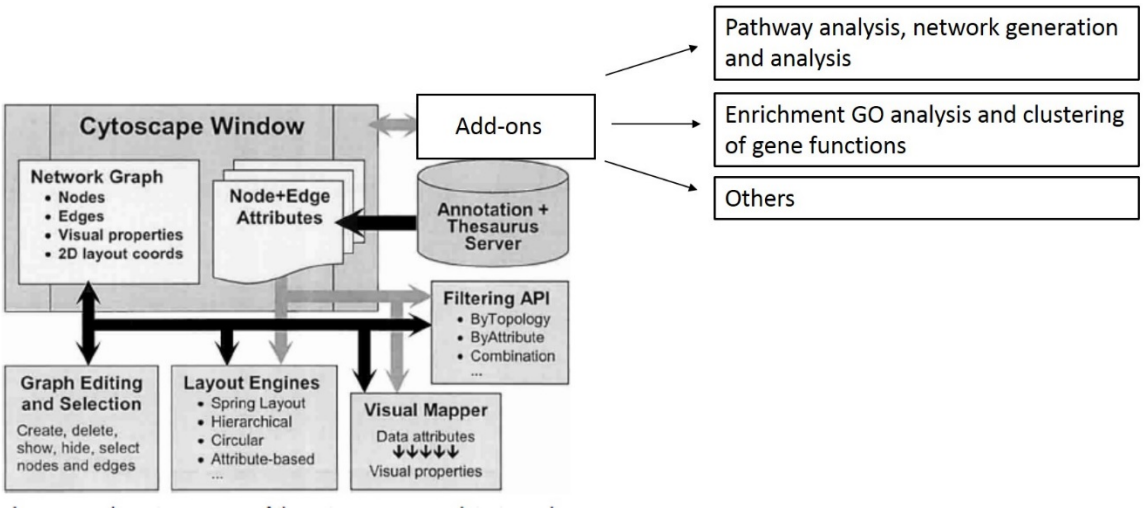


Figure 2-8: View of the Cytoscape architecture and capability adapted from Shannon *et al.* (2003)

Cytoscape (http://www.cytoscape.org/) is an open-source software tool that can create, visualise and analyse biomolecular interaction networks, using gene expression data outputs of feature counts and correspondingly, transcripts shown to be enriched by edgeR anaylsis (Shannon *et al.*, 2003; Bindea *et al.*, 2009). Cytoscape combines different online databases to improve the maximum output available leading to more reliable functional annotations. The gene ontology consortium approves all the databases used and is therefore on an 'up to date' base. Cytoscape links different databases using the same analysis pipeline regardless of species and combines gene and protein expression networks to provide output most suited to answering basic research questions. The software's functionality can be expanded using 'add-ons' (former known as plugins) allowing the user to use different network analysis approaches. Cytoscape version 3.4.0 was used for the analysis.

2.5.8.2 ClueGO

The 'add-on' 'ClueGO' (http://apps.cytoscape.org/apps/cluego) was used to group and annotate gene expression networks, accessible for bovine, ovine, porcine and human pathways, creating functionally organised GO/pathway term networks via visualising functionally clusters (Bindea *et al.*, 2009). Upregulated transcripts were annotated in biological terms, in a hierarchically structured way and enrichment analysis was done (Bindea *et al.*, 2009). Version 2.2.5 was used for the network analysis of this thesis.

2.5.9 Statistical analysis

2.5.9.1 The general linear model

The general linear model GLM is based on the t-test and predicts and compares differences in the mean of different groups, whereas it is not predicting the cause of the difference. The GLM was used for all edgeR analysis, showing strong parallels with the fisher's t-test.

2.5.9.2 The Benjamini-Hochberg

For the pathway analysis performed using 'ClueGO', Benjamini-Hochberg analysis was used as a less stringent integrated statistical tool to correct the *P*-values and in order to deal effectively with the limitations of the different and less well annotated species explored in this project (bovine, ovine and porcine), restriction criteria were set to general. The node size of a term shows the term enriched significance (Bindea *et al.*, 2009). ** was used if the term/group was PValue<0.001 (over significant); * if the PValue was 0.001<PValue<0.005 (significant) and no stars indicate 0.005<PValue<0.001.

2.4.9.3 Spearman's rank correlations

The nonparametric Spearman´s rank-order correlations (r_s) was used to calculate statistical relationship between two variables.

$0,0 \leq r_s \leq 0,2$ => None to weak correlation
$0,2 < r_s \leq 0,5$ => Weak to moderate correlation
$0,5 < r_s \leq 0,8$ => Distinct correlation
$0,8 < r_s \leq 1,0$ => High to perfect correlation

2.5.9.4 The paired one tailed Mann-Whitney U test

The paired one tailed Mann-Whitney U test was used for Western Blot analysis (see Chapter 6). A normal distribution and the requirements for a parametric test cannot be guaranteed with n = 3, therefore the t-test was not be used. The Mann-Whitney U ranks the mean for each condition and predicts how different two conditions are. The protein of interested (MOSPD3) was calculated against the control (GAPDH).

Chapter 3: Comparison and Characterisation of Total Spermatozoal and Testis RNA in: *Bos taurus, Ovis aries, Sus scrofa* and *Homo sapiens*

3.1. Introduction

10-15% of couples worldwide suffer from infertility (De Kretser, 1997; Evers, 2002). Assisted reproductive techniques (ART) have developed to help to overcome this condition. Since the birth of the first IVF child (1978), ART has led to ~ 3 million infertile couples achieving a live birth (Cohen, 1978; Carrell and Hammoud, 2010). Little, however, is known about factors influencing and causing infertility or the long term health consequences to the progeny born through ART (Schenker and Ezra, 1994; Schieve *et al.*, 2002). Therefore current research efforts are aimed at a better understanding of the causes of infertility and the development of tools (including NGS-RNA approaches) that can be used to diagnose and better understand the condition. The role of spermatozoal RNA in male infertility remains controversial, despite several reports suggesting that it may play a role in fertilization, early embryonic development and epigenetic inheritance in the progeny (Lalancette *et al.*, 2008a). Recent studies, for example, have shown that spermatozoal RNA is transferred stably to the oocyte and may be involved in post- fertilization events and even much later in the progeny's adulthood (Ostermeier *et al.*, 2004; García-Herrero *et al.*, 2011; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Schagdarsurengin and Steger, 2016).

3.1.1. Spermatozoal RNAs are involved in events of infertility

Previous investigations showed that spermatozoal RNA is involved in pre- and post-fertilisation events (Ostermeier *et al.*, 2004; Sendler *et al.*, 2013). PCR-based approaches have already indicated that apparently normal semen profiles in many idiopathically infertile patients (assessed by the WHO guidelines) may still harbour genetic abnormalities leading to distinctive mRNA profiles between fertile and infertile men (Garrido *et al.*, 2009). The rapid advancement of technology including the development of microarrays allowed the study of transcriptomic differences and first indicated alterations and pathway disruptions in teratozoospermic (morphologically normal spermatozoa below the lower reference limit) compared to normozoospermic (total number of motile and morphologically normal spermatozoa, equal to or above the lower reference limit) men (Platts *et al.*, 2007; WHO, 2010). New and more affordable NGS technologies have allowed more transcriptomic and genomic studies to be performed in recent years (Miller, 2014). These studies showed

consistent profiles of expression patterns in teratozoospermic patients, differences between fertile and infertile patients and different spermatozoal transcriptomic profiles in IUI/IVF/ICSI pregnancies giving rise to live births compared to miscarriages (Moldenhauer *et al.*, 2003; Platts *et al.*, 2007; Garrido *et al.*, 2009; García-Herrero *et al.*, 2011). An example of a potential spermatozoal RNA involvement in fertilisation events is *PLC- ζ* which leads to activation of the oocyte through Ca^{2+} signalling (Cox *et al.*, 2002). Increased *Protamine* (*PRM*) mRNA levels have been detected by RT-PCR in the less motile sperm population (40% layer), however, additionally a difference in the PRM1 and PRM2 ratio in immotile spermatozoa has also been recognised (Lambard *et al.*, 2004; Oliva, 2006). A higher level of *PRM* mRNAs in immotile spermatozoa may indicate a translational failure to replace degraded proteins (Lambard *et al.*, 2004; Carreau *et al.*, 2007; Galeraud-Denis *et al.*, 2007). With the improvement of molecular techniques, more clinical approaches are likely to be developed using the ejaculate as a non-invasive technique and a good surrogate throughout of the testis health compared to a testis biopsy (Miller, 2014; Yatsenko *et al.*, 2006).

3.1.2. Paternal influences in embryogenesis and the progeny

Several reports suggest that the paternal contribution to the embryo and to the progeny may go beyond simply providing the paternal genome (Miller and Ostermeier, 2006a; Lalancette *et al.*, 2008a; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Chen *et al.*, 2016; Schagdarsurengin and Steger, 2016; Sharma *et al.*, 2016). In the past few years, the hypothesis that spermatozoal RNA plays a role during fertilization events and in early embryogenesis has replaced the idea that the RNA is only a remnant of spermatogenesis. This hypothesis, however, needs to be confirmed (Ostermeier *et al.*, 2004; Miller and Ostermeier, 2006a; Lalancette *et al.*, 2008a). Meanwhile, there is evidence of the spermatozoon 1) functioning as a vector for genetic information; 2) transmitting RNA to the oocyte and 3) acting as a conduit for transgenerational, epigenetic effects on the offspring (Sciamanna *et al.*, 2003; Jimenez-Chillaron *et al.*, 2016; Hammoud *et al.*, 2014). Firstly, Sciamanna *et al.* (2003) showed that the spermatozoon can function as an exogenous RNA vector by incubating mouse spermatozoa with a RNA containing a β - galactosidase vector. After taking up the RNA and presumably transcribing the RNA into cDNA, the cDNA was transmitted to the progeny as shown by expression of the β - galactosidase in adult tissues. Secondly, several paternal RNAs have been reported to be transferred to the oocyte, but the function remains unknown in most cases (Ostermeier et al., 2004; Kempisty et al., 2008; Liu et al., 2012; Anderson, 2013). Some RNAs have been seen fulfilling a function e.g.: Liu *et al.* (2012) reported the inhibition of miRNAs transmitted by the spermatozoon to the oocyte leads to significant developmental delays. Thirdly, there is evidence of epigenetic alterations

influencing embryogenesis, the offspring and also transferring paternal environmental changes to the adult offspring (Anderson *et al.*, 2006; Carone *et al.*, 2010; Martínez *et al.*, 2014; Ng *et al.*, 2010). Alterations in the spermatozoal epigenome (histone modifications and DNA methylations) may be linked to the control of embryogenesis and may have effects on the offspring arising from ART (Carrell and Hammoud, 2010). ART offspring carry increased risks of preterm birth, low birthweight, congenital anomalies and higher perinatal mortality (Hansen *et al.*, 2002; Kalra and Molinaro, 2008). Transgenerational inheritance of acquired characteristics may involve the transmission of spermatozoal RNAs to the offspring that play a functional role in epigenetical programming (Dias and Ressler, 2014; Gapp *et al.*, 2014; Govindaraju *et al.*, 2012; Rodgers *et al.*, 2013). Inheritance under non-Mendelian rules is going back to Lamarck and is now considered possible by direct effects on paternal and maternal gametes that may lead to possible health risk factors to the offspring, particularly following ART (Lamarck, 1809; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Chen *et al.*, 2016; Sharma *et al.*, 2016). The hypothesis is supported by the discovery of a paternal heritable epigenetic modification (paramutation) for a c*kit* mediated phenotype on fur colour (Rassoulzadegan *et al.*, 2006). Furthermore, exposing male mice to chronic stress before breeding resulted in significantly decreased hypothalamic–pituitary–adrenal (HPA) stress axis response in the offspring (Rassoulzadegan *et al.*, 2006; Toth, 2015; Rodgers *et al.*, 2013). Consequences of paternal stress to the offspring are considered by several groups (Bohacek *et al.*, 2015; Dias and Ressler, 2014; Gapp *et al.*, 2014; Govindaraju *et al.*, 2012; Rassoulzadegan *et al.*, 2006; Schagdarsurengin and Steger, 2016). Diseases, stress and altered brain synaptic plasticity in the offspring have been associated with paternal stress factors including malnutrition, infections or advanced age (Bohacek *et al.*, 2015; Toth, 2015). Non-mendelian inheritance may involve paternally-derived miRNAs and tRNAs that are involved in epigenetic regulation and transgenerational programming like histone modifications and DNA methylation (Dias and Ressler, 2014; Toth, 2015; Chen et al., 2016). A further role of spermatozoal miRNAs was hypothesised by Liu *et al.* (2012), microRNA-34c was found to regulate the first cellular division of the mouse embryo and therefore plays a crucial role in an early stage of development. However, the relying mechanisms of the role miRNAs may play in epigenetic changes remain unanswered to date. The work of this chapter focuses on a closer investigation of the composition of spermatozoal RNAs from a number of mammalian species by NGS RNA-seq. We may get closer to understand their roles by revealing and describing the gene expression networks these RNAs could be involved in. Improved NGS techniques and a better understanding of epigenetic factors may help to understand the functional role of spermatozoal RNA.

3.2 Aims

1. Inter-species characterisation and comparison of spermatozoal and testis RNAs using NGS.
2. Determination of common transcripts and the construction of gene expression networks and pathways monitoring past events in spermatogenesis and potential future events in fertilisation, early embryogenesis and offspring.

3.3 Material and methods

3.3.1 Workflow

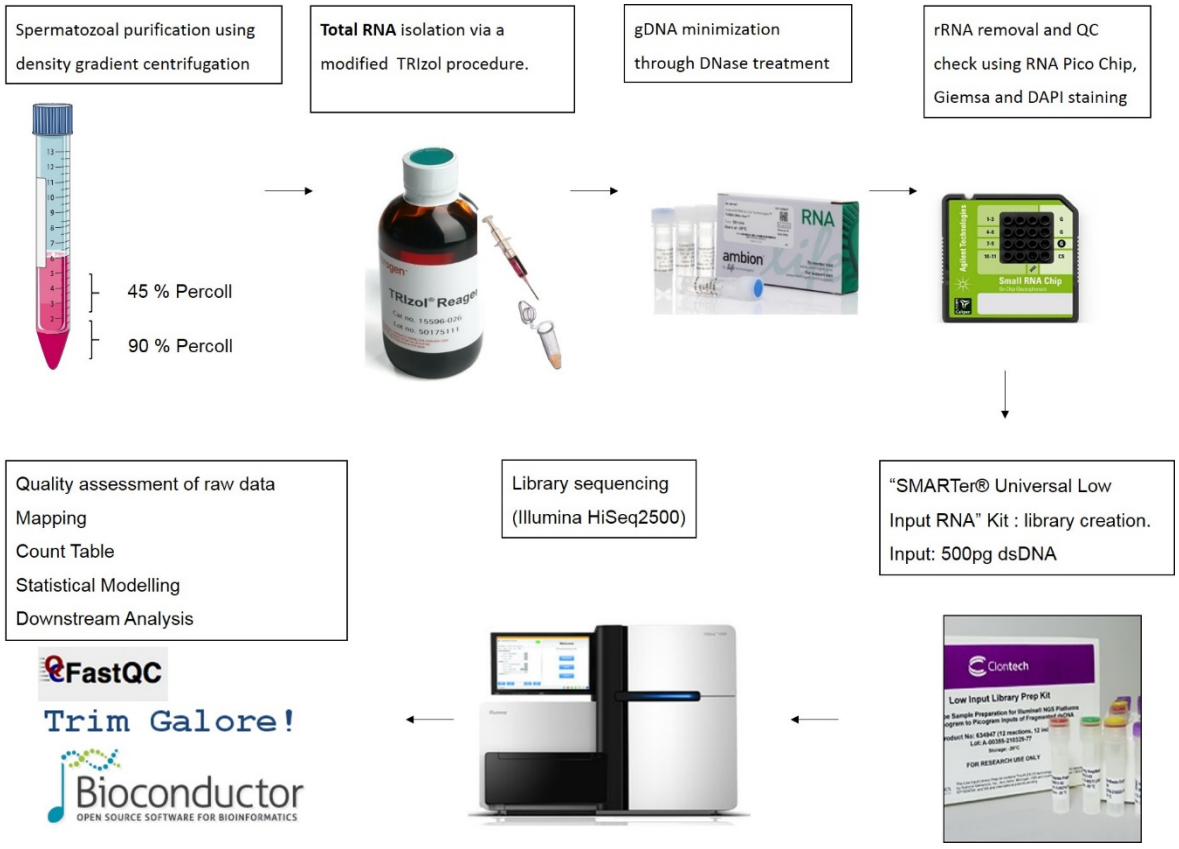


Figure 3-1: Workflow and experimental design.

3.3.2 Collection, demography and sample processing

Frozen bovine, ovine and porcine spermatozoa were thawed and processed using a density gradient centrifugation as indicated throughout in Chapter 2. The processing of human spermatozoa is described in 2.1.4 and the demography can be seen in Table 3-1.

Sample	Volunteer age (years)	Sperm concentration (million/ml)	Total sperm count ($\times 10^6$)	Semen volume (ml)	Sperm motility (%)
D2	21.00	30.00	165.00	5.50	52.00
D71	22.00	28.00	140.00	5.00	61.50
D81	21.00	101.00	404.00	4.00	55.00
D86	19.00	52.00	208.00	4.00	83.00
D88	22.00	35.00	175.00	5.00	67.00
D104	24.00	63.00	189.00	3.00	85.00
D105	21.00	59.00	218.30	3.70	60.00
D106	21.00	159.00	239.00	1.50	56.00
D107	19.00	251.00	1207.00	4.80	87.00
D109	21.00	78.00	304.00	3.90	23.00
Average	**21.10**	**85.60**	**324.93**	**4.04**	**62.95**
STD	**1.45**	**70.20**	**319.36**	**1.17**	**19.21**

Table 3-1: Donor demography.

3.3.3 Visual Quality control and RNA extraction

DAPI and Giemsa stains were prepared to check for the exclusion of somatic cells. See explanation in Chapter 2. Many procedures for RNA extraction were tested to maximise the RNA yield. For the particular experiment reported in this chapter and Chapter 5 a modified Trizol® procedure based on Gilbert *et al.* (2007) was used to extract total RNA, (see Chapter 2). To concentrate the samples and to yield a higher RNA concentration, QIAGEN MinElute Columns were used according to the manufacturer's instructions and DNase digestion took place as described in Chapter 2. For the second set of replicates an additional step was introduced and gDNA contamination was controlled using a 50 cycle RT-qPCR step (Chapter 2). Additionally, a quality and somatic cell contamination control was performed using the 'Agilent Pico Chip 4000' which was run on an Agilent Bioanalyzer (Chapter 2).

3.3.4 Library construction and equimolar pooling for NGS

3.3.4.1 rRNA depletion

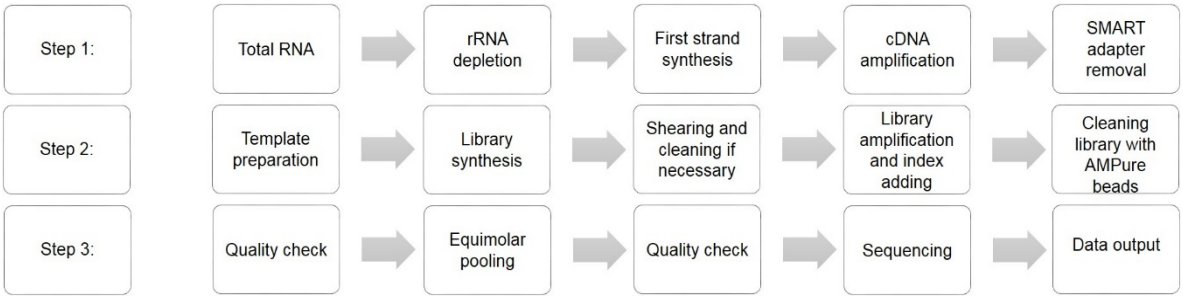


Figure 3-2: Library construction workflow

rRNAs are among the most dominant RNAs in the cell. To gain a better sequencing depth and to avoid repetitive reads, the rRNA was depleted using the Ribo-Zero™system (van Dijk *et al.*, 2014). The rRNA removal was performed using specific instructions provided by 'Clontech' for cDNA synthesis and library construction. The maximum possible RNA concentration available was used with as little as 4.5 ng for bovine sperm and testis, 2.4 ng for ovine sperm and testis, 16 ng for porcine sperm and testis and 10 ng for both human spermatozoa and testis. Clean up of the samples took place performing a column approach (RNeasy Mini Kit, QIAGEN) by using the protocol for purification of total RNA.

3.3.4.2 RNA amplification and library production for RNA sequencing using Clontech

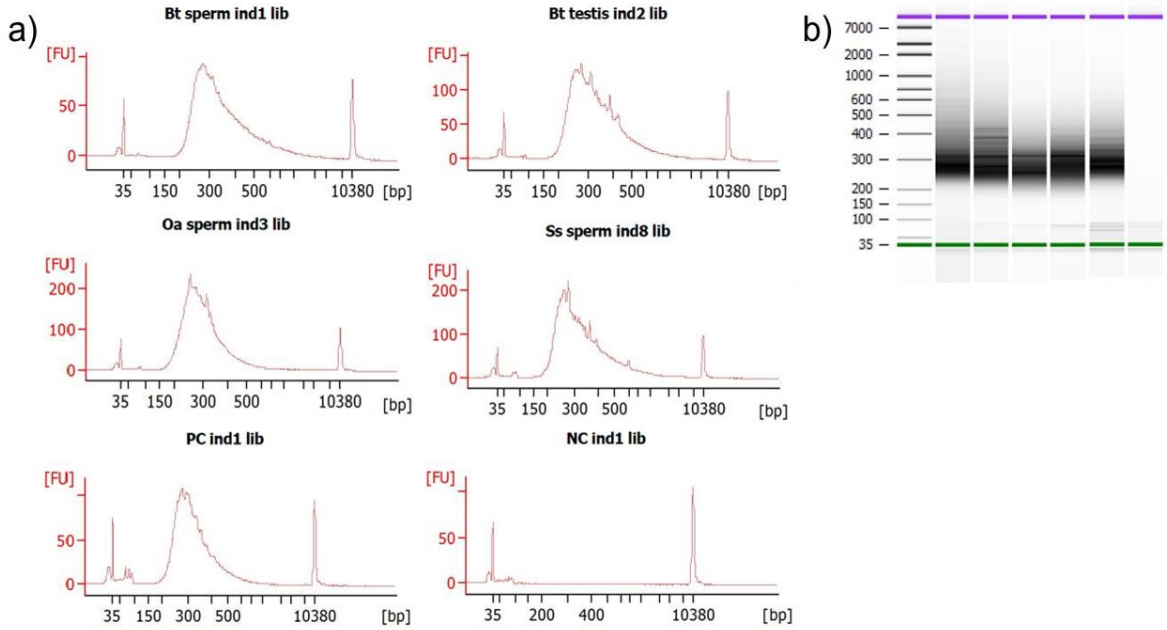


Figure 3-3: Amplified cDNA control. After sample amplification and indices attachment the quality of the libraries was controlled using an 'Agilent high sensitive DNA chip'. a) Bioanalyzer electropherogramms. Each electropherogramm is showing a single library with a peak around 200 bp - 300 bp; b) gel picture of each sample. Each line is showing the corresponding signal to the electropherogramms and is confirming the 300 bp peak. Bt =bovine (n=5), Oa = ovine (n=4); Ss = porcine (n=4); PC = positive control and NC = negative control. FU=Fluorescent Unit, bp=base pairs

SMARTer® Universal Low Input RNA Kit for Sequencing' (Clontech, version published: 17/04/2013) was used for step 1 (Figure 3-2). The input number per spermatozoal replicate was as follows: bovine: n=5, ovine: n=4; porcine: n=4 and human: n=5. Human was the only species where 5 different donors were used for both replicates. Therefore 10 donors in total were used (see demographics Table 3-1). The same animal individuals were available for ovine, ovine and porcine and were therefore used for both replicate sets. Only one testis samples was used for each species and a different batch was run as replicate. RNA transcription and cDNA amplification was performed according to the manufacturer's instructions. For library synthesis, a total input of 500 pg was used and the libraries prepared according to the manufacturer's instructions (Clontech, version published 30/4/2013). Library amplification ('Low Input Library Prep Kit') was performed using 7 PCR cycles (Step: 2, Figure 3-2). QC took place by running a 'High Sensitive DNA' chip on the Agilent

Bioanalyzer after every amplification step (Figure 3-3). If needed, the samples were sheared using covaris micro tubes. Shearing took place by using ultra sonication with an intensity set to 5 at 4°C for 140s, to achieve a 200 bp fragment peak. After shearing, a column clean up and concentration step was performed using 'MinElute®Reaction Clean Up Kit'. Illumina indices for later sample separation were chosen according to the low plexity guidelines (http://www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf): The Illumina HiSeq2500 uses a green laser to sequence G/T and a red laser to read A/C. During each reading cycle of the machine, each laser needs to be activated and each colour channel needs to be read to ensure correct registration It is important to maintain a colour balance, otherwise the indices may fail to be recognised and the samples cannot be separated after sequencing (Figure 3-4).

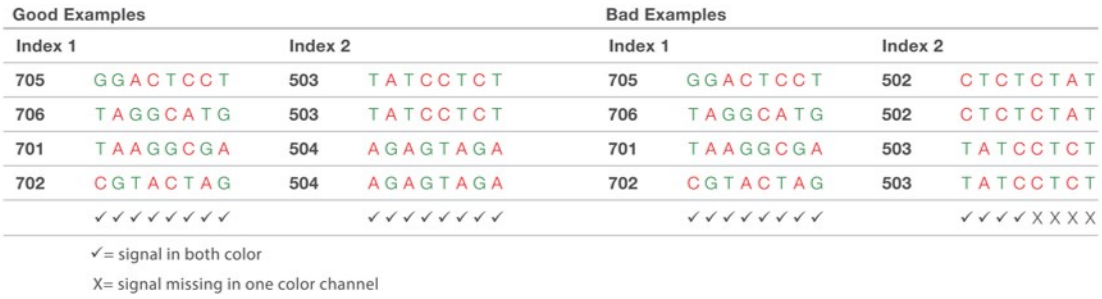


Figure 3-4: Pooling index guideline. HiSeq2500 uses a green laser to read G/T and a red laser for A/C. Both lasers need to be activated during each reading cycle, otherwise the attached indices cannot be separated after sequencing, which means the samples cannot be separated after pooling. The figure was adapted from Illumina pooing guidelines:

http://www.illumina.com/documents/products/technotes/technote_nextera_low_plex_pooling_guidelines.pdf

3.3.4.3 Equimolar Pooling

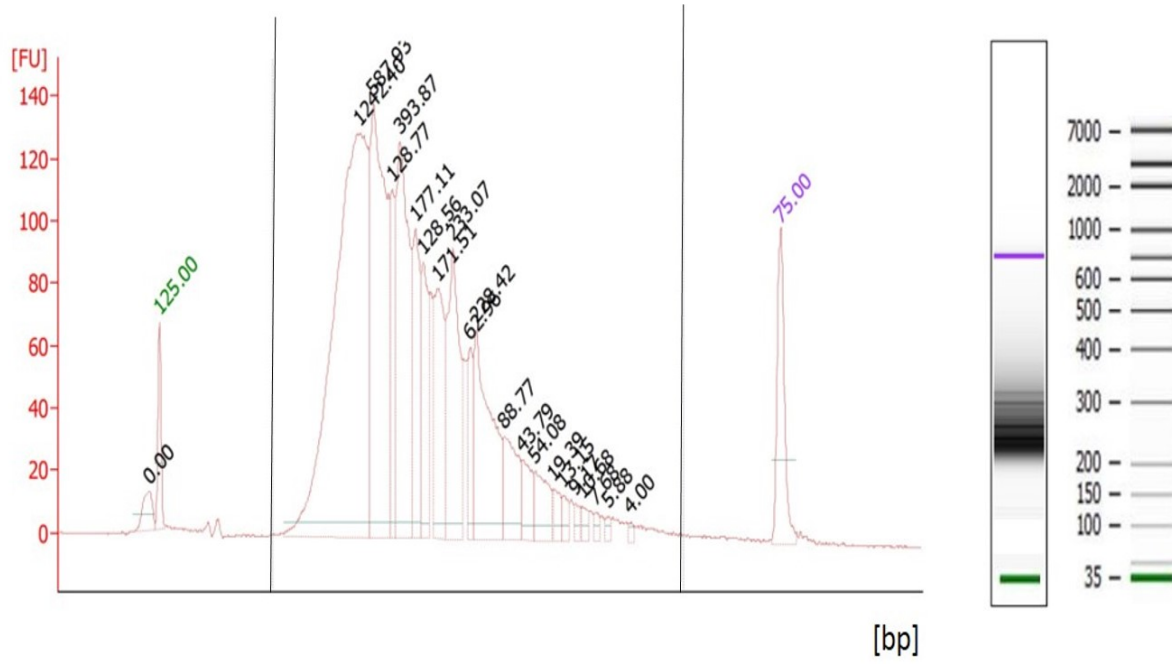


Figure 3-5: Bioanalyzer electropherogramm of amplified libraries used for equimolar pooling calculations. A border was drawn at the start and the end of each library peak and the molarity was calculated for each library by using the Agilent Bioanalyzer Software. Samples were pooled equimolar after calculation with a minimum of 30 pmol/l. FU=Fluorescent Unit, bp=base pairs

Samples to be pooled must be equimolar, a “correct pooling is significant for balanced distribution of reads across all samples. Furthermore, the number of pooled samples should be balanced for effective variant detection” (Harakalova *et al.*, 2011). For this, the Bioanalyzer electropherograms of each library QC were used. A border around each peak was drawn manually (Figure 3-5) and the molarity calculated for each trace by the Bioanalyzer Agilent software.

3.3.4.4 Sequencing

Libraries were sequenced on an Illumina HiSeq 2500 instrument (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseq2500/hiseq-2500-system-guide-15035786-01.pdf) using 100 bp paired end reads at the sequencing facility at St. James', Leeds, UK. Each library pool was clustered on the Illumina Cbot (used kit: PE-401-3001 TruSeq PE Cluster Kit v3, Illumina; protocol: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cbot/cbot-system-guide-15006165-02.pdf). The loading volume for the HiSeq2500 was either 10 pM or 11 pM and a 1% Phix Spike was loaded as an internal control for each lane (used kit: FC-401-3001 TruSeq SBS Kit v3, Illumina).

3.4 Results

3.4.1 Raw read coverage

The raw read coverage for each library is shown below (see Table 3-2 for bovine, Table 3-3 ovine, Table 3-4 porcine and Table 3-5 for human spermatozoal and testis raw reads). It can be seen throughout the species, that more testis reads could be assigned to features, which is due to the better quality RNA. The only exception is the first bovine testis replicate, which has less reads assigned.

Status	Bt_1_Sperm	Bt_1_Testis	Bt_2_sperm	Bt_2_testis
Assigned	654,717	505,697	428,485	2,102,780
Unassigned_NoFeatures	10,518,701	3,580,039	2,693,336	1,439,284
Sum	11,173,418	4,085,736	3,121,821	3,542,064

Table 3-2: Bovine read coverage

Status	Oa_1_sperm	Oa_1_testis	Oa_2_sperm	Oa_2_testis
Assigned	548,699	3,823,553	240,531	1,478,428
Unassigned_NoFeatures	9,577,995	5,168,509	307,411	1,797,034
Sum	10,126,694	8,992,062	547,942	3,275,462

Table 3-3: Ovine read coverage

Status	Ss_1_sperm	Ss_1_testis	Ss_2_sperm	Ss_2_testis
Assigned	282,276	2,818,809	621,749	3,790,132
Unassigned_Ambiguity	7,625	73,048	18,295	96,523
Unassigned_NoFeatures	1,621,231	2,908,033	626,217	4,522,047
Sum	1,911,132	5,799,890	1,266,261	8,408,702

Table 3-4: Porcine read coverage

Status	Hs_1_ fresh	Hs_1_ testis	Hs_2_ fresh	Hs_2_ testis
Assigned	339,530	1,993,306	119,650	2,177,533
Unassigned_NoFeatures	357,814	2,304,319	124,376	2,456,574
Sum	697,344	4,297,625	244,026	4,634,107

Table 3-5: Human read coverage

3.4.2 NGS Validation Spermatozoal RNA vs. Testis Transcript Profile

3.4.2.1 Non-coding RNAs-unassigned and ribosomal/mitochondrial Reads

The NGS- sequencing of spermatozoal and testis RNAs revealed ~42.7% of reads in bovine, ~48.7% of reads in ovine, ~34.2% of reads in porcine and ~7.2% of reads in human spermatozoa were unassigned whereas ~30.0% in bovine, ~41.1% ovine, ~26.7% in porcine and ~21% in human testis RNAs were unassigned using the currently released annotations. Although rRNA depletion was carried out, ~0.4% bovine, ~7% ovine, ~11% porcine and ~21.5% human spermatozoal RNAs mapped to ribosomal/mitochondrial transcripts. In contrast, ~1.2% bovine, ~0.3% ovine, ~1.6% porcine and ~3% human testis RNAs mapped to ribosomal/mitochondrial transcripts. All percentages are shown in the a) and b) in Figure 3-6 for bovine, Figure 3-7 for ovine, Figure 3-8 for porcine and Figure 3-9 for human. See Table 3-2 for a summary for bovine, Table 3-3 for ovine, Table 3-4 for porcine and Table 3-5 for human coding, unassigned and ribosomal/mitochondrial reads.

3.4.2.2 RNA Repeats

The majority of assigned sequences in spermatozoal RNAs mapped to the repeat mask (see e) and f) in Figure 3-6 to Figure 3-9 for bovine, ovine, porcine and human spermatozoa and testis repeats). Analysis revealed that ~49.7% of bovine, ~43.9% of ovine, ~37.4% of porcine and ~42.93% of human spermatozoal RNAs were repetitive compared to ~35.8% of bovine, ~15.7% of ovine, ~26.8% of porcine and ~31.8% of human testis RNAs.

Repeat RNAs were mainly derived from: Signal recognition particle RNA (7SLRNA), long and short subunits of rRNA (long: LSU-rRNA_Hsa; short: SSU-rRNA_Hsa), L1M long interspersed nuclear elements (LINES), mammalian-wide interspersed repeat (MWIR), mammalian long terminal repeats transposon (MLT), MEdium Reiterated repeats (MER), long terminal repeats (LTR) and other miscellaneous repeat types. Reads from bovine spermatozoa and testis indicated that the L1M LINE was the most abundant repetitive RNA in both, with the main difference between the two sources accounted for by 7SL RNA (spermatozoa ~9.5%, testis ~3.2%. See Figure 3-7 e) and f). Repeat RNAs in the ovine differed mainly in the large rRNA subunits between spermatozoa and testis. A clear trend in bovine, porcine and human spermatozoa was observed in having more abundant levels of 7SLRNA compared with testes. No difference, however, was seen between the two RNA sources in the ovine (Figure 3-7 e) and f). The main human “repeat RNAs” in spermatozoa were the large rRNA subunit (16.3%) and the small rRNA subunit (20.5%) rRNA subunits. These repetitive element RNAs were ten times less abundant in preparations from the

corresponding testis (~1.7% and ~2.2%). The third most abundant repeat RNA in human spermatozoa was the 7SLRNA (~8.2%), with human testis containing just over half that amount (~4.9%). L1M repeats were the most abundant in testis with ~13.5% of all assigned repeats. No reads were assigned to this family in RNA from human spermatozoa (see Figure 3-9 e) and f). Similar to the human repeats, porcine also showed an abundance of assigned reads for the two main rRNAs subunits (~22.4% and ~10.3%, respectively), followed by 7SLRNA with ~9.1% of all repeats in spermatozoal RNA. The most abundant repeat RNA from porcine testis was concordant to the human the L1M line at ~14.3% (Figure 3-8 c) and f). See Table 3-2 for a summary for bovine, Table 3-3 for ovine, Table 3-4 for porcine and Table 3-5 for human repeats.

3.4.2.3 Coding RNAs and Other RNA Types

Approximately ~7% of bovine, ~0.4% of ovine, ~14% of porcine and ~26% of human spermatozoal reads matched coding RNAs compared with 32.4% of bovine, ~43% of ovine, ~43% of porcine and ~43% human testis reads (see c) and d) in Figure 3-6 to Figure 3-9). The lowest percentage of RNAs are summarized as “Other RNA” types, mainly consisting of micro RNAs (MIRs), long intergenic non-coding RNAs (LNC) and small nucleolar *RNAs* (SNOR) and are shown in the c) and d) of Figure 3-6 for bovine, Figure 3-7 for ovine, Figure 3-8 for porcine and Figure 3-9 for human. Comparing the “other RNA” types in spermatozoa and testis of bovine, ovine and porcine showed that they each had similar ratios for both spermatozoa and testis (c) and d) panel of Figure 3-6 until Figure 3-9). However, differences in miRNA and snorRNA seen in both spermatozoal and testis reads were not significant. miRNA and long intergenic non-coding RNAs differed in more abundant reads of either ~10% in spermatozoa (miRNA) or testis (long intergenic non-coding RNAs). Approximately 0.94% more snorRNA were seen in bovine spermatozoa than in testis. Long intergenic non-coding RNAs were ~3% higher in preparations from spermatozoa whereas there was no difference in levels of snorRNA between sperm and testis. There were almost ~3% more MIR RNAs from porcine spermatozoa compared with corresponding RNA from testis. Long intergenic non-coding RNAss and snorRNAs were almost equal. Human spermatozoa and testis showed more differences than the other species with almost three times more MIRs from testis compared to spermatozoa, ~10% less abundant long intergenic non-coding RNAs in testis and almost seven times more snoRNAs in testis compared to spermatozoa. Since the raw counts were used, the comparison is more quantitative and the ranking of the transcript level between sources and species could be compared. See Table 3-2 for a summary for bovine, Table 3-3 for ovine, Table 3-4 for porcine and Table 3-5 for coding sequences and other mapped transcripts.

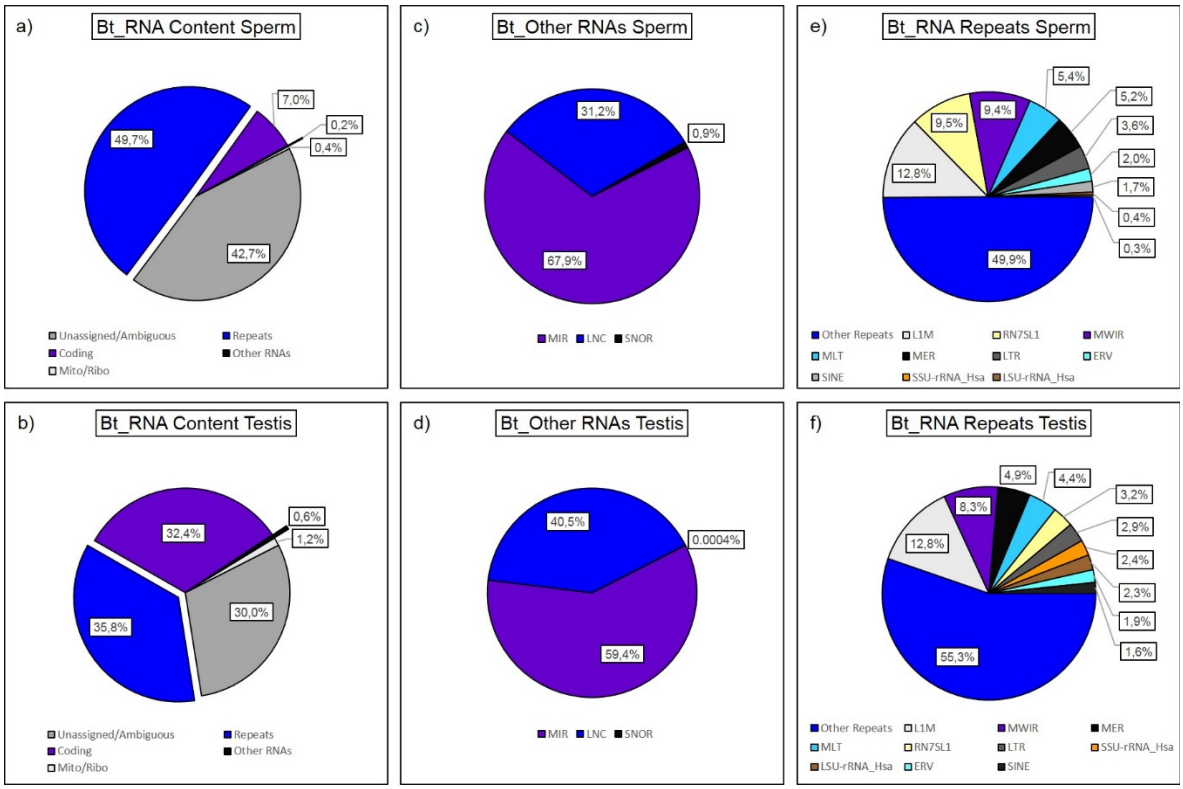


Figure 3-6: RNA proportioning bovine (Bt). RNA content in a) bovine sperm and b) in testis consisting of unassigned/ambiguous reads (sperm ~42.7%, testis: ~30%): mitochondrial and ribosomal (Mito/Ribo) reads (sperm: 0.4%, testis: 1.2%), coding RNA sequences (spermatozoa ~7%, testis 32.4%) and other RNAs of ~0.2% in spermatozoa and ~0.6% in testis. Other RNAs: c) spermatozoa and d) testis: divided into ~67.9% MIR in sperm, ~59.4% MIR in testis, ~31.2% LNC in spermatozoa, ~40.5% in testis and ~0.9% of SNORs in spermatozoa and 0.0004% in testis. The biggest subunit of the mapped RNAs were repeats: spermatozoal e) and testis f) with having mainly L1M long interspersed nuclear elements (LINES), mammalian-wide interspersed repeat (MWIR), mammalian long terminal repeats transposon (MLT), MEdium Reiterated repeats (MER), long terminal repeats LTR and other repeat types. Having the same percentage in LINEs (L1M) and similar percentages in both spermatozoa and testis and as the main difference appeared the RN7SL1 increased in spermatozoa and decreased rRNA subunits in spermatozoal. All graphs were calculated with the ratio of each panel. LNC= Long intergenic non-coding RNAs; MIR= miRNAs; SNOR= small nucleolar RNAs

Species	RNA content	Sperm	Testis
Bovine	Repeats	49.72%	35.75%
	Unassigned/Ambiguous	42.70%	30.00%
	Coding	7.02%	32.35%
	Mito/Ribo	0.37%	1.23%
	Other RNAs	0.18%	0.62%

Species	Other RNAs	Sperm	Testis
Bovine	MIR	67.86%	59.44%
	LNC	31.20%	40.52%
	SNOR	0.94%	0.03%

Species	RNA Repeats	Sperm	Testis
Bovine	Other Repeats	49.89%	55.29%
	L1M	12.80%	12.84%
	RN7SL1	9.48%	3.22%
	MWIR	9.40%	8.25%
	MLT	5.35%	4.38%
	MER	5.15%	4.86%
	LTR	3.58%	2.93%
	ERV	1.97%	1.92%
	SINE	1.68%	1.65%
	SSU-rRNA_Hsa	0.39%	2.36%
	LSU-rRNA_Hsa	0.31%	2.30%

Table 3-6: RNA proportioning bovine (Bt).

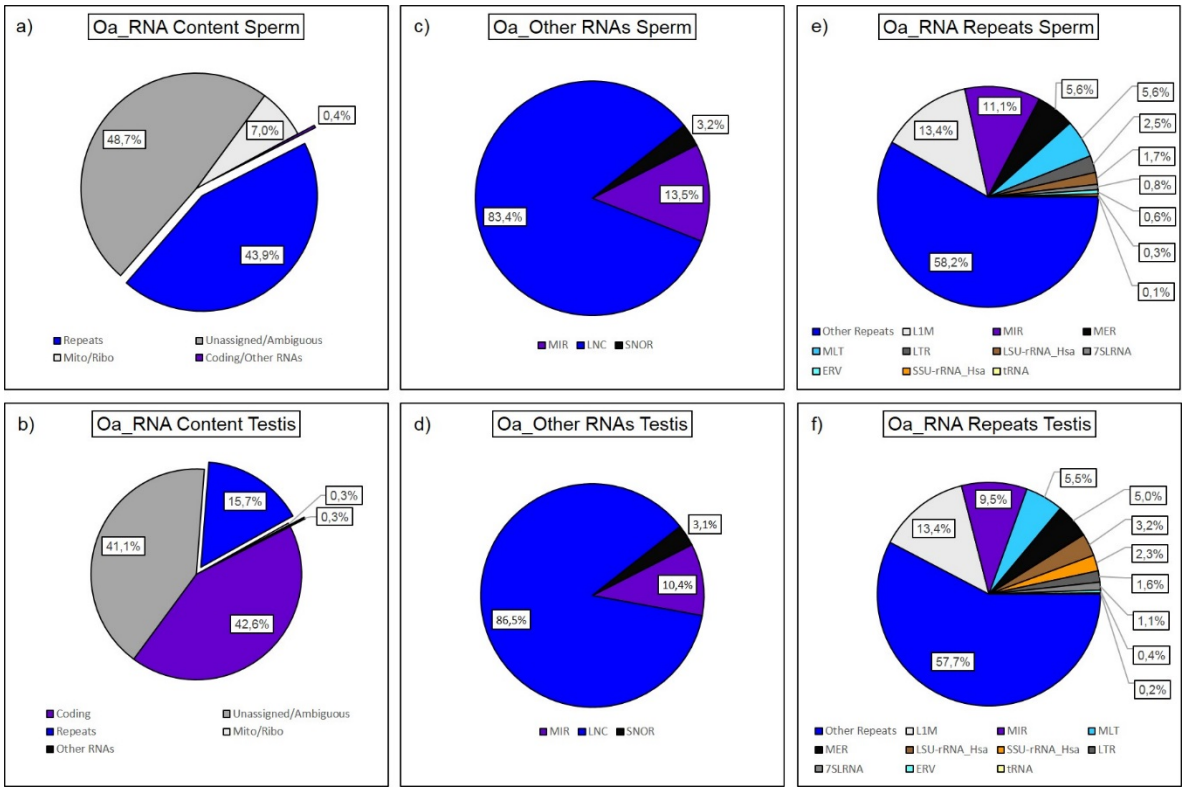


Figure 3-7: RNA proportioning in ovine (Oa). RNA content: a) spermatozoa and b) testis: ~48.7% of spermatozoal reads and ~41% reads in testis could not be assigned or were ambiguous. ~0.4% of the spermatozoal reads were coding or “other RNAs” in sperm and ~43% in testis. Ribosomal and mitochondrial reads were ~7% in sperm and ~0.3% in testis. Other RNAs: c) spermatozoa and d) testis and repeats e) spermatozoa and f) testis did not differ significantly, except for the MIR ratio, were more MIR was found in testis. LNC= Long intergenic non-coding RNAs; MIR= miRNAs; SNOR= small nucleolar RNAs

Species	RNA content	Sperm	Testis
Ovine	Repeats	43.90%	15.69%
	Unassigned/Ambiguous	48.71%	41.09%
	Coding/Other RNAs	0.36%	42.91%
	Mito/Ribo	7.04%	0.31%

Species	RNA Repeats	Sperm	Testis
Ovine	Other Repeats	58.23%	57.68%
	L1M	13.39%	13.37%
	MIR	11.09%	9.48%
	MER	5.64%	5.03%
	MLT	5.59%	5.55%
	LTR	2.49%	1.63%
	LSU-rRNA_Hsa	1.74%	3.25%
	7SLRNA	0.79%	1.14%
	ERV	0.59%	0.41%
	SSU-rRNA_Hsa	0.32%	2.30%
	tRNA	0.11%	0.17%

Species	Other RNAs	Sperm	Testis
Ovine	MIR	13.46%	10.38%
	LNC	83.37%	86.52%
	SNOR	3.17%	3.09%

Table 3-7: RNA proportioning in ovine (Oa).

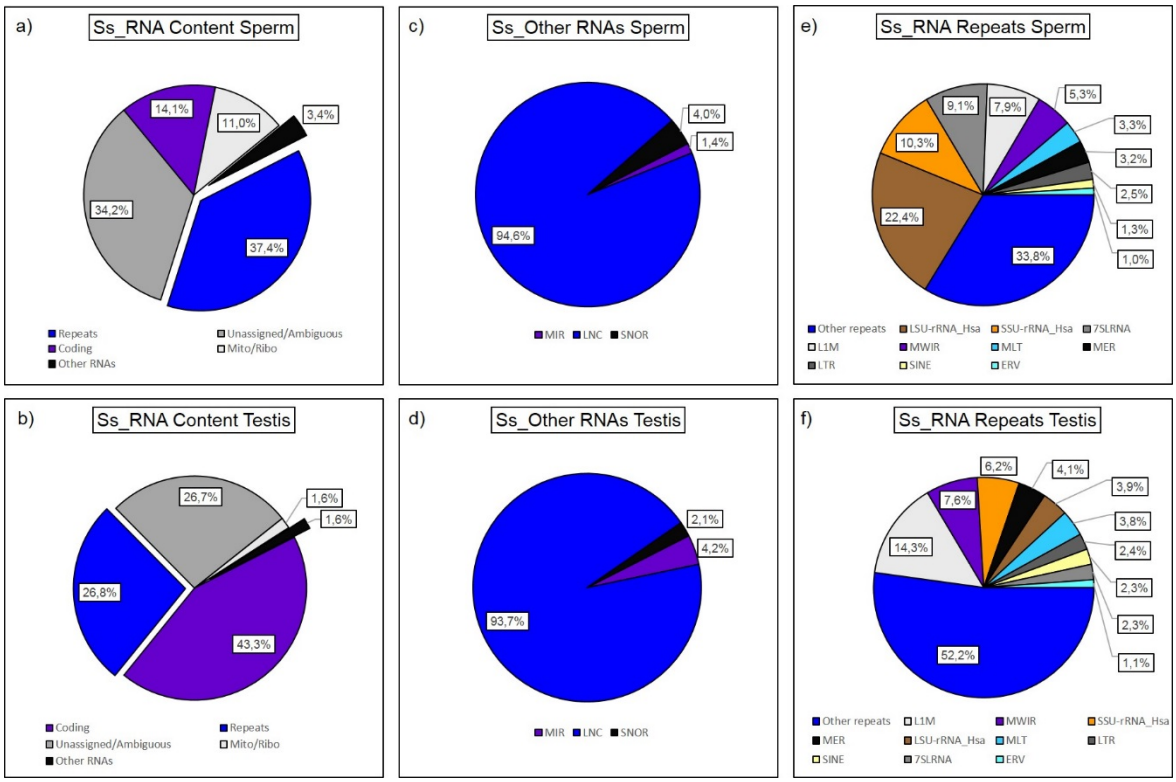


Figure 3-8: RNA proportioning in porcine (Ss). a) spermatozoa: ~34.2% of reads were not assigned in spermatozoa, whereas the percentage in b) testis lay at ~26.7%. The repeat mask revealed ~37% hits in spermatozoa and ~26.8% in testis. ~14% of the reads in spermatozoa and ~43.3% in testis were coding transcripts. More mitochondrial and ribosomal reads were found in porcine spermatozoa than in testis. Other RNAs in c) spermatozoa and d) testis is showing: double the amount of SNOR RNAs were found in spermatozoa, whereas the MIRs were found 4 times less in spermatozoa as in testis. Repeats: e) spermatozoa and f) testis: The main difference in the repeats was the two rRNA subunits and 7SLRNA with a higher percentage inspermatozoa, whereas more LINEs (L1M) were found in testis. LNC= Long intergenic non-coding RNAs; MIR= miRNAs; SNOR= small nucleolar RNAs

Species	RNA content	Sperm	Testis
Porcine	Repeats	37.37%	26.77%
	Unassigned/Ambiguous	34.18%	26.72%
	Coding	14.11%	43.34%
	Mito/Ribo	10.96%	1.60%
	Other RNAs	3.39%	1.58%

Species	Other RNAs	Sperm	Testis
Porcine	MIR	1.44%	4.22%
	LNC	94.58%	93.66%
	SNOR	3.98%	2.13%

Species	RNA Repeats	Sperm	Testis
Porcine	Other repeats	33.76%	52.18%
	LSU-rRNA_Hsa	22.41%	3.85%
	SSU-rRNA_Hsa	10.31%	6.16%
	7SLRNA	9.14%	2.25%
	L1M	7.86%	14.30%
	MWIR	5.30%	7.58%
	MLT	3.26%	3.82%
	MER	3.20%	4.09%
	LTR	2.52%	2.35%
	SINE	1.26%	2.27%
	ERV	0.98%	1.14%

Table 3-8: RNA proportioning in porcine (Ss).

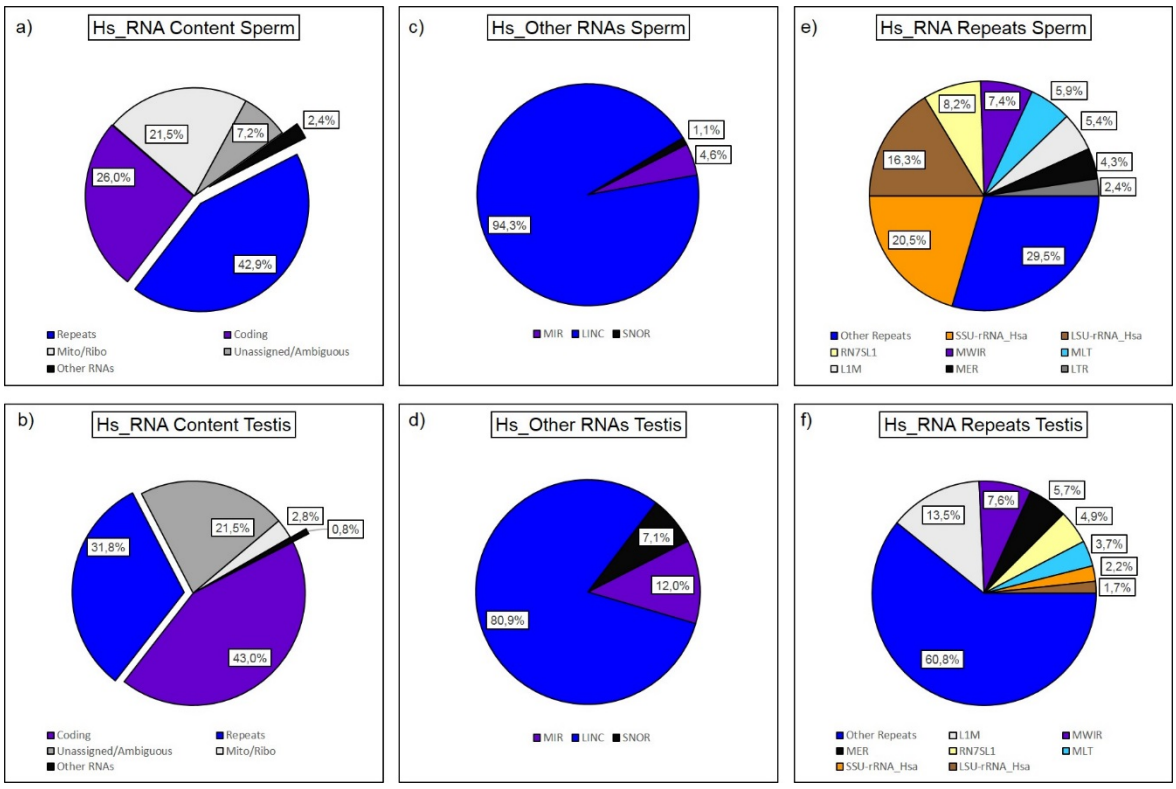


Figure 3-9: RNA proportioning human (Hs). RNA content: a) spermatozoa and b) testis: Coding sequences in spermatozoa were ~26% and ~43% in testis. While ~42.9% repeats mapped in spermatozoa, ~31.8% mapped in testis. ~7.2% of spermatozoa reads and ~21.5% in testis could not be assigned or were ambiguous. The mitochondrial and ribosomal percentage was higher in sperm than in testis with ~22% (sperm) vs. ~3% (testis). Other RNAs: c) spermatozoa and d) testis: More MIR and SNORs were seen in testis than in sperm. Repeats: e) spermatozoa and f) testis: As in porcine the repeats were similarly distributed in human. The two rRNA subunits and the 7SLRNA were higher in sperm and the LINEs in the testis. LNC= Long intergenic non-coding RNAs; MIR= miRNAs; SNOR= small nucleolar RNAs

Species	RNA content	Frozen	Testis
Human	Repeats	42.93%	31.83%
	Unassigned/Ambiguous	7.17%	21.47%
	Coding	25.97%	43.04%
	Mito/Ribo	21.51%	2.84%
	Other RNAs	2.42%	0.81%

Species	RNA Repeats	Frozen	Testis
Human	Other Repeats	29.52%	60.81%
	LSU-rRNA_Hsa	16.34%	1.69%
	RN7SL1	8.17%	4.87%
	SSU-rRNA_Hsa	20.49%	2.23%
	MWIR	7.36%	7.56%
	L1M	5.45%	13.47%
	MLT	5.95%	3.70%
	MER	4.31%	5.66%
	LTR	2.41%	-

Species	Other RNAs	Frozen	Testis
Human	MIR	4.64%	12.04%
	LNC	94.28%	80.86%
	SNOR	1.08%	7.10%

Table 3-9: RNA proportioning human (Hs).

3.4.2.4 DE Analysis Using the Bioconductor Package edgeR

3.4.2.4.1 Data Exploration and MDS Plots

The technical and biological variation between each spermatozoa and testis sample of each respective species was summarised using a MDS plot prior to the differential expression analysis using raw reads (see Figure 3-10). All testis samples of each species are visibly separated from the spermatozoal samples indicating variations between the different sources in the first dimension on the x-axis. The testis replicates of ovine (Figure 3-10b) and porcine (Figure 3-10c) cluster closer together and demonstrate little variation between the replicates, whereas all remaining replicates (Figure 3-10a) bovine and (Figure 3-10d) human throughout the species and source types show more variation between each other by separation in the second dimension (y-axis).

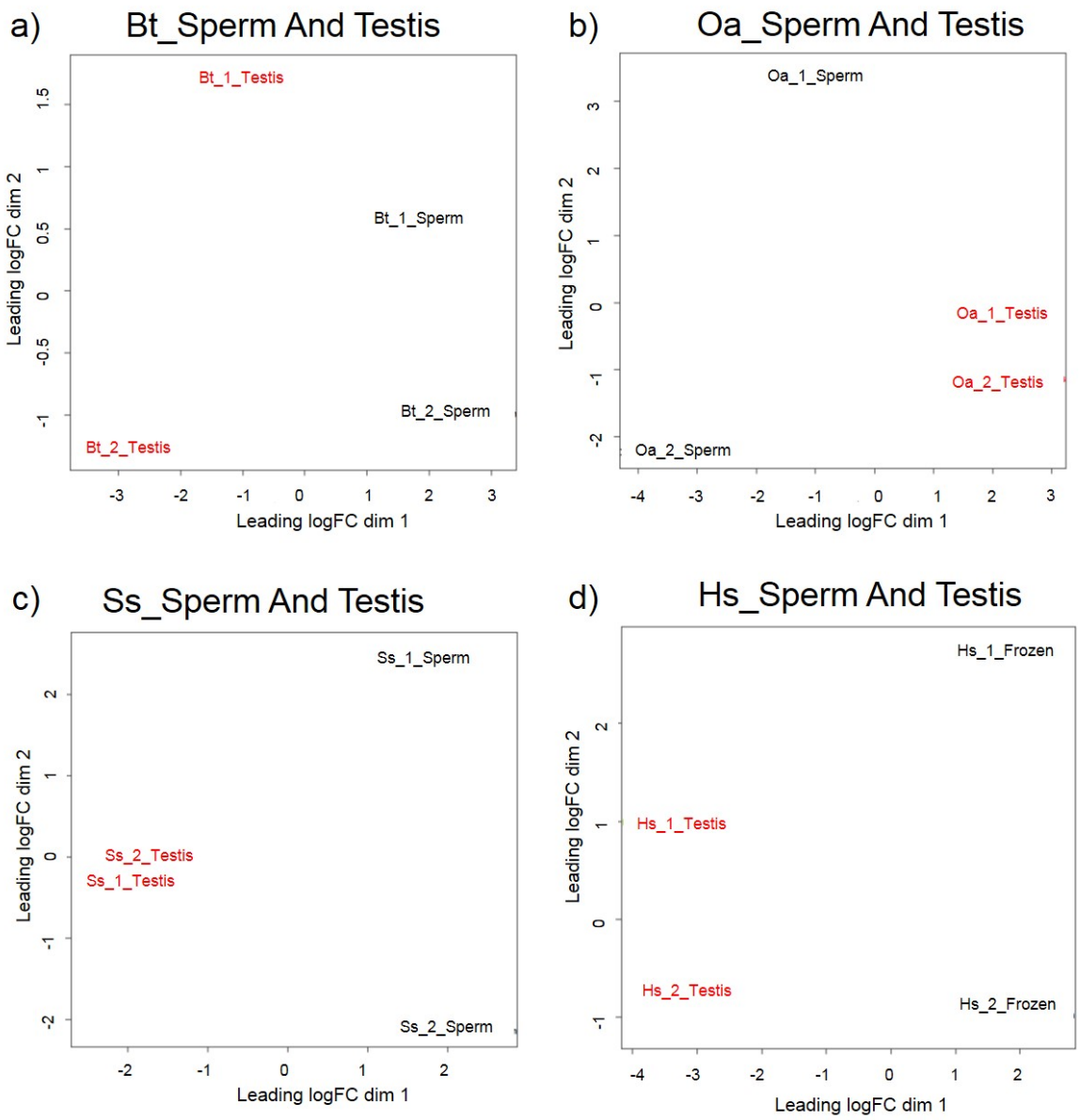
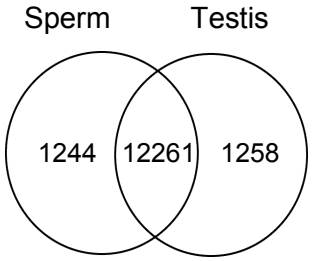


Figure 3-10: MDS plot of bovine, ovine, porcine and human sperm and testis replicates. a) bovine spermatozoa n=5 (black) and testis n=1 (red) per replicate, b) ovine spermatozoa n=4 (black) and testis n=1 (red) per replicate, c) porcine spermatozoa n=4 (black and testis n=1 (red) per replicate and d) human spermatozoa n=5 (black) and testis n=1 (red) per replicate. All replicates form a cluster in the first dimension, whereas they differ in the second. Testis and spermatozoa in all species significantly differ in the first dimension, showing less commonality. The closer the replicates and samples lie the less biological and technical difference can be seen.

3.4.2.4.2 Spearman's correlations analysis and DE transcripts

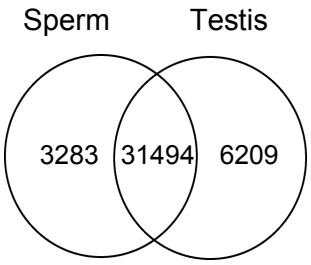
The statistical difference of the spermatozoa and testis of each species was calculated using Spearman's correlation to quantify the association. Analysis from porcine spermatozoa and testis shows r_s>0.5, which indicates a distinct correlation between them (Figure 3-11c). Bovine, ovine and human spermatozoa and their respective testis showed r_s<0.5, which indicates weak correlations (Figure 3-11 a) bovine, b) ovine and d) human). Differential expression analysis using edgeR (Bioconductor) in R was used to reveal shared and differentially expressed transcripts between spermatozoal and testis RNA. This comparison permits the distinction of transcripts and groups appearing significantly in spermatozoal RNA by noticeably higher representation relative to the testis and to earlier, intra-testicular stages in spermatogenesis using normalised libraries to 1,000,000 cpm. 1,244 mRNAs were more highly represented in bovine, 3,283 mRNAs were more highly represented in ovine, 2,187 mRNAs were more highly represented in porcine and 1,024 mRNAs were more highly represented in human spermatozoa compared to their corresponding testis. The majority of mRNAs were shared with testis (Figure 3-11a)-d)). To visualise the DE transcripts a smear plot was used, plotting the $\log_2$FC at the y-axis and the $\log_2$CPM onto the x-axis. The red dots indicate the higher representation of RNAs from each respective source with negative values representing the most highly abundant RNAs in spermatozoa and positive values representing the most highly abundant mRNAs in testis. Black dots are transcripts in common and not significantly represented in both spermatozoa and testis (Figure 3-12 a) bovine; b) ovine; c) porcine; d) human).

a) Bovine



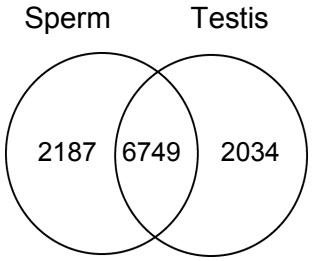
Bovine Sperm vs. Testis
r_s =0.48

b) Ovine



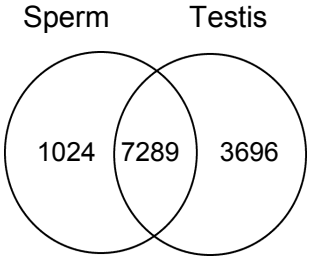
Ovine Sperm vs. Testis
r_s = 0.37

c) Porcine



Porcine Sperm vs. Testis
r_s = 0.71

d) Human



Human Sperm Frozen vs. Testis
r_s =0.41

Figure 3-11: Spearman's correlations and DE transcripts. a) Bovine, b) ovine and d) human showed a r_s<0.5 between spermatozoa and testis, which means that both have a minor to weak correlation, whereas the c) porcine correlation was distinct with r_s > 0.5. Venn diagrams show number of transcripts, significantly higher represented in either spermatozoa or testis at 5% FDR: showing higher represented genes in spermatozoa for bovine: 1244; ovine: 3282; porcine: 2187 and human: 1024.

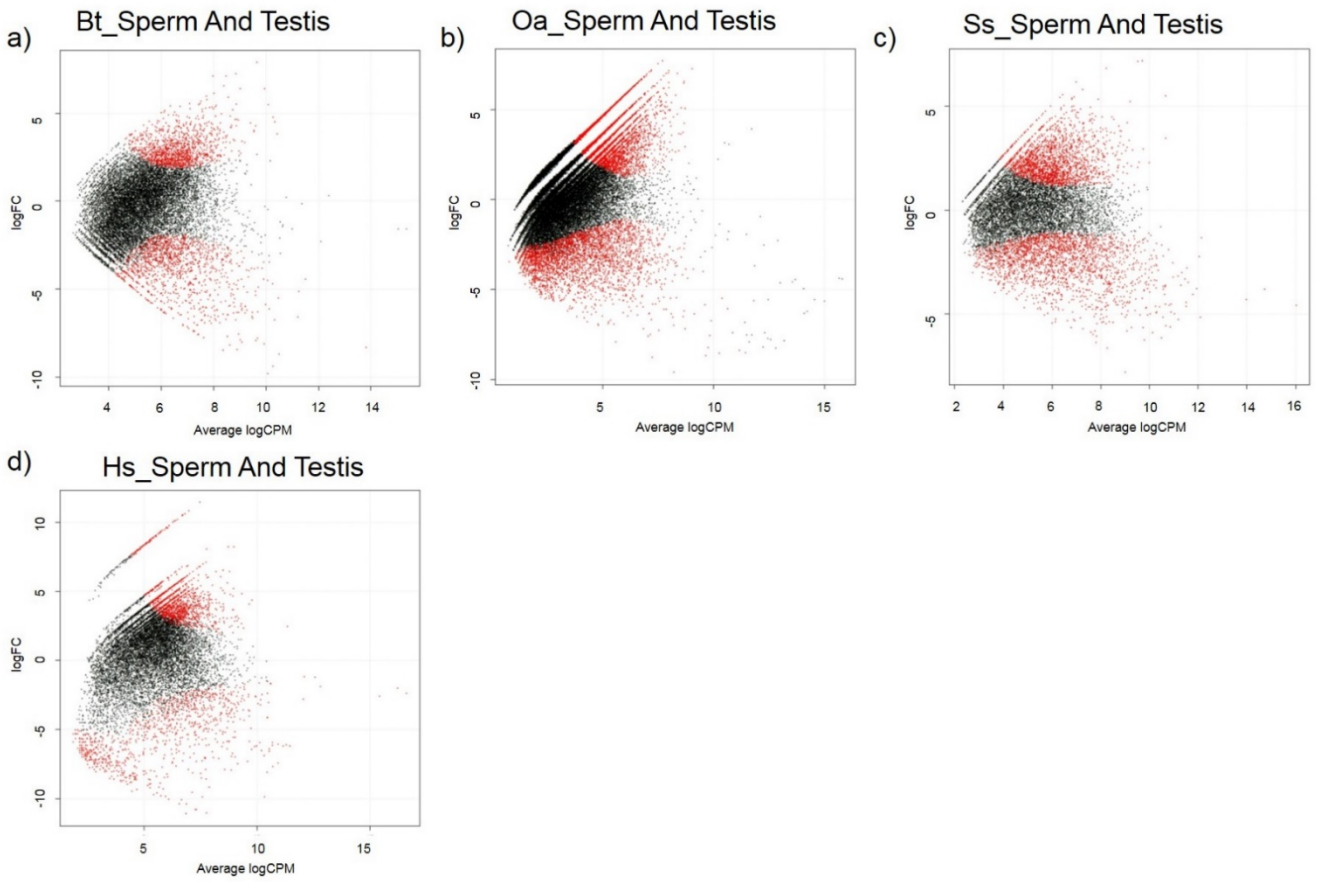


Figure 3-12: Visualisation of DE transcripts using a smear plot. a) bovine, b) ovine, c) porcine and d) human DE expresses transcripts in spermatozoa and testis. Significantly higher represented spermatozoal transcripts in the negative range of the y-axis are indicate in red, whereas higher represented testis transcripts are located in the positive range of the y-axis and are also indicated in red. Transcripts without any significant over-representation either in spermatozoal or in testis are indicated in black.

3.4.2.4.3 Functional annotation and clustered transcripts DE spermatozoa

Different functional annotations and clustering systems have been used during the development of the work described in this thesis. Pre-analysis and data comparison were carried out using the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID), PANTHER and *QIAGEN's Ingenuity Pathway Analysis (IPA)* (Huang *et al.*, 2009; Mi *et al.*, 2009). Since bioinformatics is a rapidly developing field, changes in our initial and follow up data analysis arose. Therefore, frequent re-analysis of functional annotation and gene ontology (GO) of the NGS data was performed. The most recent bioinformatically re-analysed data was processed with 'Cytoscape', using the add-on ClueGO (see also Chapter 2) as a tool, combining the newest releases of open source databases, functioning as a combination between DAVID and PANTHER as shown in this section of the thesis (Bindea *et al.*, 2009).

3.4.2.4.4 GO Functional Analysis of Bovine, Ovine, Porcine and Human Spermatozoa using the Cytoscape App ClueGO

Significantly higher represented transcripts with a $\log_2$FC of -4 - 0 (minus indicates upregulated transcripts in spermatozoa, compared to testis) were used to determine transcript clusters and relationships in bovine (Figure 3-13 and Table 3-10), ovine (Figure 3-14 and Table 3-11), porcine (Figure 3-15 and Table 3-12) and human (Figure 3-16 and Table 3-13) spermatozoa compared to their corresponding testis samples. A single transcript can be represented in multiple GO cluster categories. GO functional analysis of more highly represented RNAs in spermatozoa showed 11 clusters in bovine, 9 in ovine, 13 in porcine and 35 in human, revealing strong relevance to spermatozoal function. Annotation issues were improved by building our own annotations (see Chapter 2) and pathway analysis therefore showed expected functions throughout each species. These functions can be divided mainly into the following categories: 1. Spermatozoal motility (cytoskeleton organisation, mitotic cell cycle regulation); 2. Gamete generation and functions in fertilisation (including protein phosphorylation and dephosphorylation); 3. Spermatozoon-oocyte interaction (ion channels and transmembrane transport); 4. Embryo and tissue/organ development (including hormone regulation in the embryo) and 5. RNA and DNA (histone and chromatin) regulations/modifications; 6. Others (catabolic and metabolic processes).

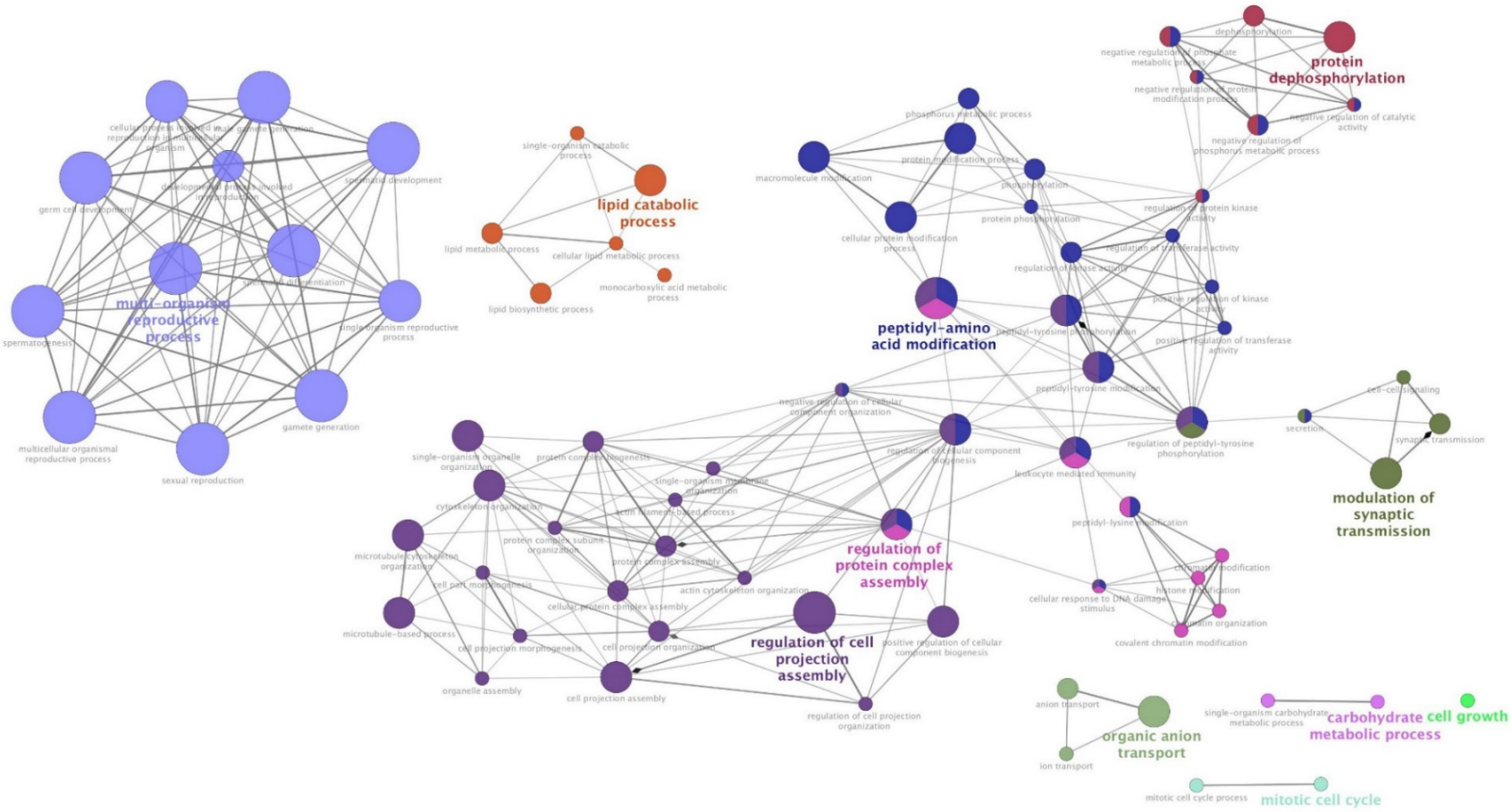


Figure 3-13: Functional pathway analysis of higher represented transcripts in bovine spermatozoa. ClueGO analysis revealed 11 clusters of functional biological processes of spermatozoa. Main functions revealed relevance in reproductive processes, including spermatozoal motility, gamete generation, fertilization, embryo development, RNA functions and catabolic processes.

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
organic anion transport	GO:0006811	ion transport	62,0E-3	100,0E-3	62,0E-3	130,0E-3	1,08	13,00
	GO:0006820	anion transport	26,0E-3	59,0E-3	62,0E-3	130,0E-3	1,59	7,00
	GO:0015711	organic anion transport	15,0E-3	36,0E-3	62,0E-3	130,0E-3	1,97	6,00
cell growth	GO:0016049	cell growth	63,0E-3	100,0E-3	63,0E-3	110,0E-3	1,53	5,00
peptidyl-amino acid modification**	GO:0044087	regulation of cellular component biogenesis	1,1E-3	6,1E-3	7,2E-3	40,0E-3	1,96	11,00
	GO:0002443	leukocyte mediated immunity	11,0E-3	31,0E-3	7,2E-3	40,0E-3	2,44	5,00
	GO:0006793	phosphorus metabolic process	28,0E-3	60,0E-3	7,2E-3	40,0E-3	1,01	25,00
	GO:0006974	cellular response to DNA damage stimulus	170,0E-3	210,0E-3	7,2E-3	40,0E-3	1,09	6,00
	GO:0043086	negative regulation of catalytic activity	290,0E-3	290,0E-3	7,2E-3	40,0E-3	1,00	6,00
	GO:0043412	macromolecule modification	11,0E-3	30,0E-3	7,2E-3	40,0E-3	1,00	31,00
	GO:0051129	negative regulation of cellular component organization	210,0E-3	230,0E-3	7,2E-3	40,0E-3	1,14	5,00
	GO:0036211	protein modification process	7,0E-3	25,0E-3	7,2E-3	40,0E-3	1,04	31,00
	GO:0043254	regulation of protein complex assembly	10,0E-3	29,0E-3	7,2E-3	40,0E-3	2,16	6,00
	GO:0046903	secretion	90,0E-3	120,0E-3	7,2E-3	40,0E-3	1,17	8,00
	GO:0051338	regulation of transferase activity	72,0E-3	110,0E-3	7,2E-3	40,0E-3	1,26	8,00
	GO:0010563	negative regulation of phosphorus metabolic process	50,0E-3	90,0E-3	7,2E-3	40,0E-3	1,48	6,00
	GO:0016310	phosphorylation	26,0E-3	58,0E-3	7,2E-3	40,0E-3	1,08	19,00
	GO:0051347	positive regulation of transferase activity	110,0E-3	140,0E-3	7,2E-3	40,0E-3	1,29	5,00
	GO:0006464	cellular protein modification process	7,0E-3	25,0E-3	7,2E-3	40,0E-3	1,04	31,00
	GO:0045936	negative regulation of phosphate metabolic process	50,0E-3	90,0E-3	7,2E-3	40,0E-3	1,48	6,00
	GO:0006468	protein phosphorylation	85,0E-3	120,0E-3	7,2E-3	40,0E-3	1,02	15,00
	GO:0018193	peptidyl-amino acid modification	230,0E-6	1,5E-3	7,2E-3	40,0E-3	1,78	16,00
	GO:0031400	negative regulation of protein modification process	71,0E-3	110,0E-3	7,2E-3	40,0E-3	1,36	6,00
	GO:0043549	regulation of kinase activity	33,0E-3	67,0E-3	7,2E-3	40,0E-3	1,41	8,00
	GO:0033674	positive regulation of kinase activity	83,0E-3	120,0E-3	7,2E-3	40,0E-3	1,42	5,00
	GO:0018205	peptidyl-lysine modification	23,0E-3	54,0E-3	7,2E-3	40,0E-3	2,02	5,00
	GO:0018212	peptidyl-tyrosine modification	8,0E-3	25,0E-3	7,2E-3	40,0E-3	2,26	6,00
	GO:0045859	regulation of protein kinase activity	150,0E-3	180,0E-3	7,2E-3	40,0E-3	1,16	6,00
	GO:0018108	peptidyl-tyrosine phosphorylation	7,7E-3	27,0E-3	7,2E-3	40,0E-3	2,28	6,00
	GO:0050730	regulation of peptidyl-tyrosine phosphorylation	4,5E-3	18,0E-3	7,2E-3	40,0E-3	3,05	5,00
regulation of protein complex assembly*	GO:0002443	leukocyte mediated immunity	11,0E-3	31,0E-3	10,0E-3	40,0E-3	2,44	5,00
	GO:0006974	cellular response to DNA damage stimulus	170,0E-3	210,0E-3	10,0E-3	40,0E-3	1,09	6,00
	GO:0006325	chromatin organization	180,0E-3	210,0E-3	10,0E-3	40,0E-3	1,08	6,00
	GO:0043254	regulation of protein complex assembly	10,0E-3	29,0E-3	10,0E-3	40,0E-3	2,16	6,00
	GO:0016568	chromatin modification	220,0E-3	230,0E-3	10,0E-3	40,0E-3	1,10	5,00
	GO:0016570	histone modification	80,0E-3	120,0E-3	10,0E-3	40,0E-3	1,43	5,00
	GO:0016569	covalent chromatin modification	87,0E-3	120,0E-3	10,0E-3	40,0E-3	1,40	5,00
	GO:0018193	peptidyl-amino acid modification	230,0E-6	1,5E-3	10,0E-3	40,0E-3	1,78	16,00
	GO:0018205	peptidyl-lysine modification	23,0E-3	54,0E-3	10,0E-3	40,0E-3	2,02	5,00
lipid catabolic process*	GO:0006629	lipid metabolic process	51,0E-3	90,0E-3	86,0E-3	130,0E-3	1,21	11,00
	GO:0044712	single-organism catabolic process	100,0E-3	130,0E-3	86,0E-3	130,0E-3	1,17	9,00
	GO:0044255	cellular lipid metabolic process	85,0E-3	120,0E-3	86,0E-3	130,0E-3	1,19	8,00
	GO:0008610	lipid biosynthetic process	42,0E-3	78,0E-3	86,0E-3	130,0E-3	1,55	6,00
	GO:0016042	lipid catabolic process	13,0E-3	33,0E-3	86,0E-3	130,0E-3	2,34	5,00
	GO:0032787	monocarboxylic acid metabolic process	190,0E-3	210,0E-3	86,0E-3	130,0E-3	1,27	5,00
carbohydrate metabolic process	GO:0005975	carbohydrate metabolic process	110,0E-3	150,0E-3	110,0E-3	140,0E-3	1,16	7,00
	GO:0044723	single-organism carbohydrate metabolic process	150,0E-3	180,0E-3	110,0E-3	140,0E-3	1,17	6,00
protein dephosphorylation*	GO:0043086	negative regulation of catalytic activity	290,0E-3	290,0E-3	280,0E-3	280,0E-3	1,00	6,00
	GO:0010563	negative regulation of phosphorus metabolic process	50,0E-3	90,0E-3	280,0E-3	280,0E-3	1,48	6,00
	GO:0016311	dephosphorylation	32,0E-3	67,0E-3	280,0E-3	280,0E-3	1,65	6,00
	GO:0045936	negative regulation of phosphate metabolic process	50,0E-3	90,0E-3	280,0E-3	280,0E-3	1,48	6,00
	GO:0006470	protein dephosphorylation	10,0E-3	31,0E-3	280,0E-3	280,0E-3	2,12	6,00
	GO:0031400	negative regulation of protein modification process	71,0E-3	110,0E-3	280,0E-3	280,0E-3	1,36	6,00
	GO:0045859	regulation of protein kinase activity	150,0E-3	180,0E-3	280,0E-3	280,0E-3	1,16	6,00
modulation of synaptic transmission*	GO:0007267	cell-cell signaling	200,0E-3	220,0E-3	86,0E-3	110,0E-3	1,08	7,00
	GO:0007268	synaptic transmission	36,0E-3	70,0E-3	86,0E-3	110,0E-3	1,61	6,00
	GO:0046903	secretion	90,0E-3	120,0E-3	86,0E-3	110,0E-3	1,17	8,00
	GO:0050804	modulation of synaptic transmission	4,7E-3	18,0E-3	86,0E-3	110,0E-3	3,01	5,00
	GO:0050730	regulation of peptidyl-tyrosine phosphorylation	4,5E-3	18,0E-3	86,0E-3	110,0E-3	3,05	5,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
multi-organism reproductive process**	GO:0003006	developmental process involved in reproduction	1,9E-3	10,0E-3	200,0E-6	2,2E-3	1,95	10,00
	GO:0044702	single organism reproductive process	94,0E-6	650,0E-6	200,0E-6	2,2E-3	1,93	16,00
	GO:0044703	multi-organism reproductive process	710,0E-9	9,1E-6	200,0E-6	2,2E-3	2,85	16,00
	GO:0048609	multicellular organismal reproductive process	11,0E-6	100,0E-6	200,0E-6	2,2E-3	2,78	13,00
	GO:0019953	sexual reproduction	610,0E-9	9,4E-6	200,0E-6	2,2E-3	3,09	15,00
	GO:0022412	cellular process involved in reproduction in multicellular organism	77,0E-6	590,0E-6	200,0E-6	2,2E-3	3,77	8,00
	GO:0007276	gamete generation	8,3E-6	92,0E-6	200,0E-6	2,2E-3	3,11	12,00
	GO:0007281	germ cell development	26,0E-6	220,0E-6	200,0E-6	2,2E-3	4,40	8,00
	GO:0048232	male gamete generation	550,0E-9	10,0E-6	200,0E-6	2,2E-3	4,04	12,00
	GO:0048515	spermatid differentiation	370,0E-9	14,0E-6	200,0E-6	2,2E-3	7,77	8,00
	GO:0007283	spermatogenesis	530,0E-9	13,0E-6	200,0E-6	2,2E-3	4,05	12,00
	GO:0007286	spermatid development	340,0E-9	26,0E-6	200,0E-6	2,2E-3	7,84	8,00
	GO:0044087	regulation of cellular component biogenesis	1,1E-3	6,1E-3	33,0E-3	92,0E-3	1,96	11,00
	GO:0002443	leukocyte mediated immunity	11,0E-3	31,0E-3	33,0E-3	92,0E-3	2,44	5,00
	GO:0007017	microtubule-based process	8,8E-3	27,0E-3	33,0E-3	92,0E-3	1,68	9,00
	GO:0030029	actin filament-based process	250,0E-3	260,0E-3	33,0E-3	92,0E-3	1,02	5,00
	GO:0030030	cell projection organization	34,0E-3	67,0E-3	33,0E-3	92,0E-3	1,25	12,00
regulation of cell projection assembly**	GO:0044089	positive regulation of cellular component biogenesis	4,3E-3	18,0E-3	33,0E-3	92,0E-3	2,26	7,00
	GO:0070271	protein complex biogenesis	53,0E-3	91,0E-3	33,0E-3	92,0E-3	1,19	11,00
	GO:0044802	single-organism membrane organization	270,0E-3	280,0E-3	33,0E-3	92,0E-3	1,06	6,00
	GO:1902589	single-organism organelle organization	2,3E-3	11,0E-3	33,0E-3	92,0E-3	1,22	24,00
	GO:0006974	cellular response to DNA damage stimulus	170,0E-3	210,0E-3	33,0E-3	92,0E-3	1,09	6,00
	GO:0007010	cytoskeleton organization	7,8E-3	26,0E-3	33,0E-3	92,0E-3	1,46	13,00
	GO:0030031	cell projection assembly	3,5E-3	15,0E-3	33,0E-3	92,0E-3	2,35	7,00
	GO:0051129	negative regulation of cellular component organization	210,0E-3	230,0E-3	33,0E-3	92,0E-3	1,14	5,00
	GO:0070925	organelle assembly	130,0E-3	160,0E-3	33,0E-3	92,0E-3	1,25	6,00
	GO:0071822	protein complex subunit organization	130,0E-3	160,0E-3	33,0E-3	92,0E-3	1,01	12,00
	GO:0000226	microtubule cytoskeleton organization	11,0E-3	30,0E-3	33,0E-3	92,0E-3	1,87	7,00
	GO:0006461	protein complex assembly	53,0E-3	91,0E-3	33,0E-3	92,0E-3	1,19	11,00
	GO:0030036	actin cytoskeleton organization	220,0E-3	230,0E-3	33,0E-3	92,0E-3	1,11	5,00
	GO:0031344	regulation of cell projection organization	86,0E-3	120,0E-3	33,0E-3	92,0E-3	1,40	5,00
	GO:0032990	cell part morphogenesis	210,0E-3	230,0E-3	33,0E-3	92,0E-3	1,05	7,00
	GO:0043254	regulation of protein complex assembly	10,0E-3	29,0E-3	33,0E-3	92,0E-3	2,16	6,00
	GO:0060491	regulation of cell projection assembly	400,0E-6	2,3E-3	33,0E-3	92,0E-3	5,26	5,00
	GO:0048858	cell projection morphogenesis	200,0E-3	230,0E-3	33,0E-3	92,0E-3	1,08	7,00
	GO:0043623	cellular protein complex assembly	40,0E-3	75,0E-3	33,0E-3	92,0E-3	1,57	6,00
	GO:0018193	peptidyl-amino acid modification	230,0E-6	1,5E-3	33,0E-3	92,0E-3	1,78	16,00
	GO:0018212	peptidyl-tyrosine modification	8,0E-3	25,0E-3	33,0E-3	92,0E-3	2,26	6,00
	GO:0018108	peptidyl-tyrosine phosphorylation	7,7E-3	27,0E-3	33,0E-3	92,0E-3	2,28	6,00
	GO:0050730	regulation of peptidyl-tyrosine phosphorylation	4,5E-3	18,0E-3	33,0E-3	92,0E-3	3,05	5,00
mitotic cell cycle	GO:0000278	mitotic cell cycle	120,0E-3	150,0E-3	120,0E-3	130,0E-3	1,15	7,00
	GO:1903047	mitotic cell cycle process	260,0E-3	270,0E-3	120,0E-3	130,0E-3	1,01	5,00

Table 3-10: Biological pathways for bovine spermatozoa. Showing represented and significantly enriched transcripts in bovine spermatozoa compared to testis including the subcategories and statistical considerations.

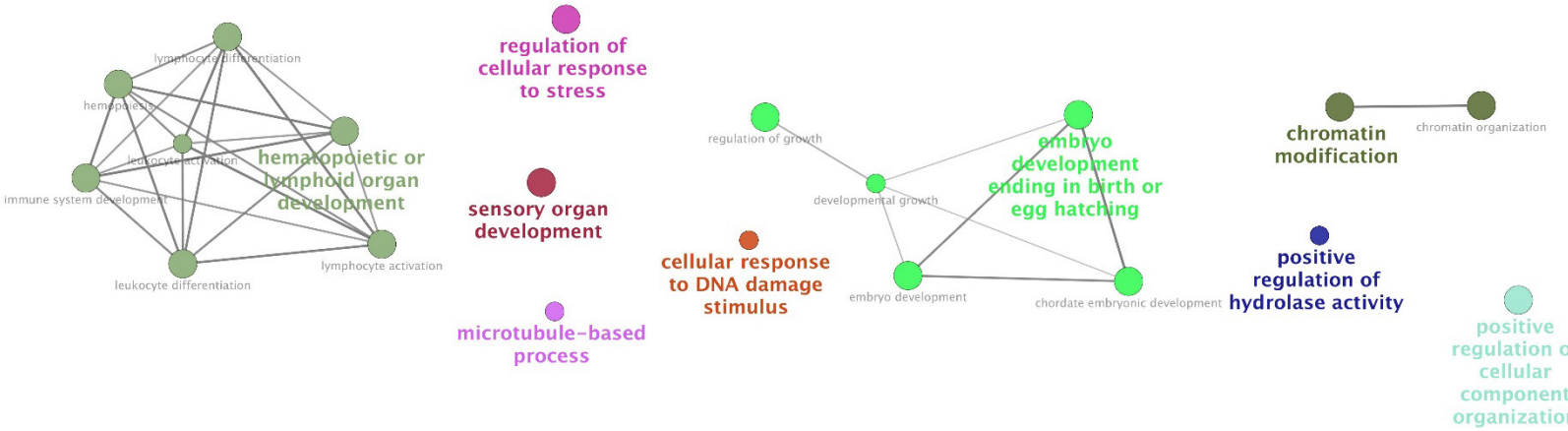


Figure 3-14: Ovine functional pathway analysis: biological processes in represented spermatozoal transcripts were examined and revealed 9 non-significant clusters, playing a role in gamete generation, fertilization and embryonic development.

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
hematopoietic or lymphoid organ development	GO:0045321	leukocyte activation	100,0E-3	100,0E-3	72,0E-3	100,0E-3	1,02	5,00
	GO:0002520	immune system development	15,0E-3	57,0E-3	72,0E-3	100,0E-3	1,24	8,00
	GO:0046649	lymphocyte activation	59,0E-3	87,0E-3	72,0E-3	100,0E-3	1,19	5,00
	GO:0048534	hematopoietic or lymphoid organ development	11,0E-3	52,0E-3	72,0E-3	100,0E-3	1,32	8,00
	GO:0030097	hemopoiesis	24,0E-3	67,0E-3	72,0E-3	100,0E-3	1,22	7,00
	GO:0002521	leukocyte differentiation	8,5E-3	53,0E-3	72,0E-3	100,0E-3	1,70	6,00
	GO:0030098	lymphocyte differentiation	6,7E-3	63,0E-3	72,0E-3	100,0E-3	2,12	5,00
embryo development ending in birth or egg hatching	GO:0040008	regulation of growth	46,0E-3	80,0E-3	32,0E-3	290,0E-3	1,28	5,00
	GO:0048589	developmental growth	5,8E-3	110,0E-3	32,0E-3	290,0E-3	1,62	7,00
	GO:0009790	embryo development	38,0E-3	81,0E-3	32,0E-3	290,0E-3	1,05	9,00
	GO:0009792	embryo development ending in birth or egg hatching	25,0E-3	59,0E-3	32,0E-3	290,0E-3	1,22	7,00
	GO:0043009	chordate embryonic development	24,0E-3	76,0E-3	32,0E-3	290,0E-3	1,23	7,00
positive regulation of cellular component organization	GO:0051130	positive regulation of cellular component organization	54,0E-3	85,0E-3	54,0E-3	240,0E-3	1,07	8,00
positive regulation of hydrolase activity	GO:0051345	positive regulation of hydrolase activity	100,0E-3	100,0E-3	100,0E-3	100,0E-3	1,01	5,00
regulation of cellular response to stress	GO:0080135	regulation of cellular response to stress	59,0E-3	87,0E-3	59,0E-3	170,0E-3	1,19	5,00
cellular response to DNA damage stimulus	GO:0006974	cellular response to DNA damage stimulus	93,0E-3	100,0E-3	93,0E-3	100,0E-3	1,04	5,00
microtubule-based process	GO:0007017	microtubule-based process	85,0E-3	100,0E-3	85,0E-3	110,0E-3	1,07	5,00
sensory organ development	GO:0007423	sensory organ development	67,0E-3	85,0E-3	67,0E-3	120,0E-3	1,15	5,00
chromatin modification	GO:0006325	chromatin organization	61,0E-3	83,0E-3	61,0E-3	130,0E-3	1,18	5,00
	GO:0016568	chromatin modification	42,0E-3	80,0E-3	61,0E-3	130,0E-3	1,32	5,00

Table 3-11: Biological pathways for ovine spermatozoa. Showing represented and significantly upregulated transcripts in ovine spermatozoa compared to testis including the subsections and statistical involvement

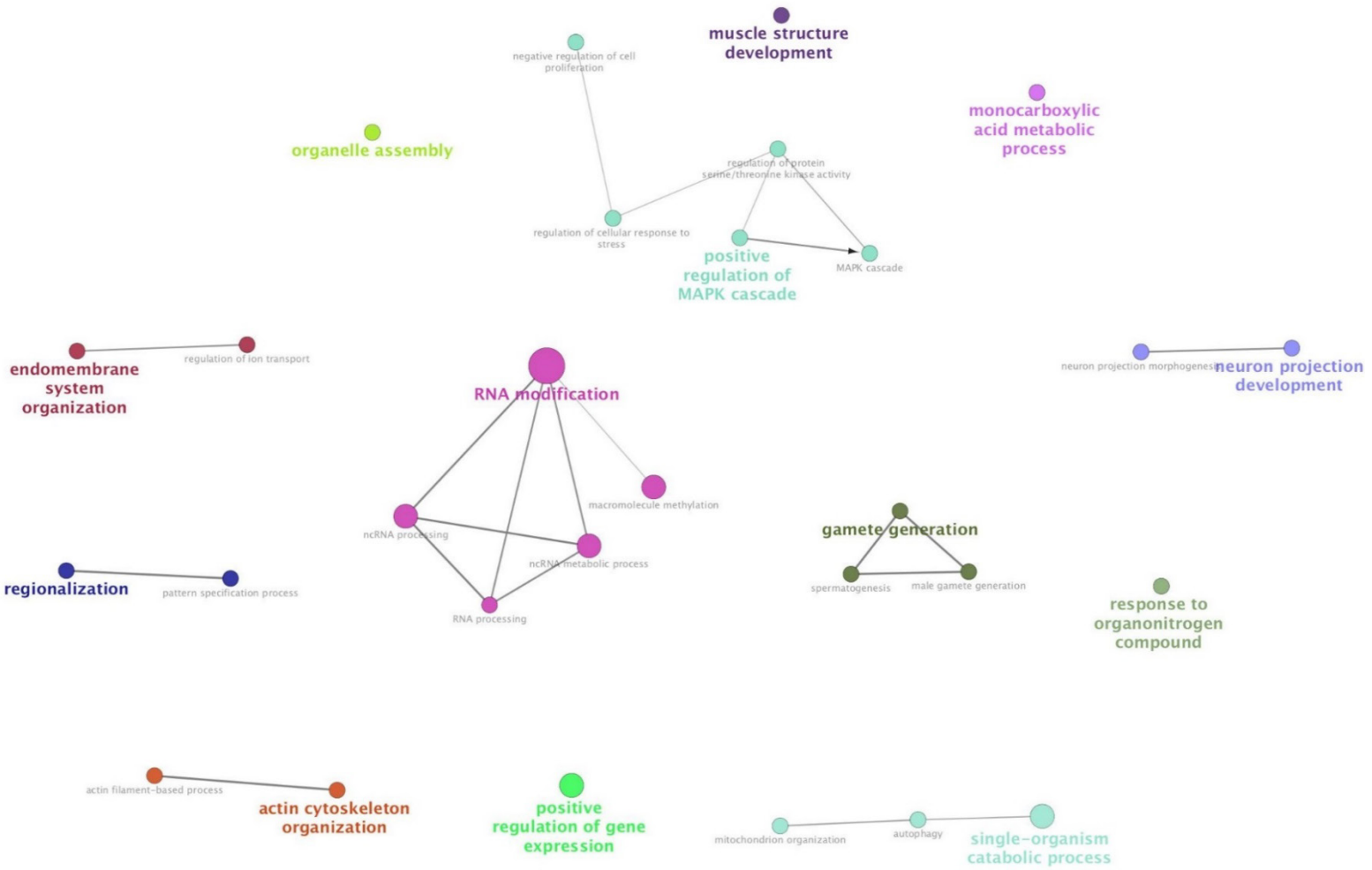


Figure 3-15: 13 clusters of biological processes using represented transcripts in porcine spermatozoa. 13 clustered biological processes were revealed using higher represented spermatozoal RNAs compared to testis in porcine.

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
response to organonitrogen compound	GO:0010243	response to organonitrogen compound	120.0E-3	130.0E-3	120.0E-3	130.0E-3	1.03	4.00
positive regulation of gene expression	GO:0010628	positive regulation of gene expression	11.0E-3	73.0E-3	11.0E-3	140.0E-3	1.02	13.00
actin cytoskeleton organization	GO:0030029	actin filament-based process	98.0E-3	130.0E-3	98.0E-3	140.0E-3	1.01	5.00
	GO:0030036	actin cytoskeleton organization	75.0E-3	130.0E-3	98.0E-3	140.0E-3	1.10	5.00
monocarboxylic acid metabolic process	GO:0032787	monocarboxylic acid metabolic process	110.0E-3	130.0E-3	110.0E-3	130.0E-3	1.06	4.00
endomembrane system organization	GO:0010256	endomembrane system organization	100.0E-3	120.0E-3	400.0E-3	400.0E-3	1.11	4.00
	GO:0043269	regulation of ion transport	120.0E-3	130.0E-3	400.0E-3	400.0E-3	1.03	4.00
gamete generation	GO:0007276	gamete generation	100.0E-3	120.0E-3	100.0E-3	130.0E-3	1.11	4.00
	GO:0048232	male gamete generation	40.0E-3	130.0E-3	100.0E-3	130.0E-3	1.54	4.00
	GO:0007283	spermatogenesis	39.0E-3	140.0E-3	100.0E-3	130.0E-3	1.54	4.00
neuron projection development	GO:0031175	neuron projection development	41.0E-3	110.0E-3	41.0E-3	100.0E-3	1.17	6.00
	GO:0048812	neuron projection morphogenesis	110.0E-3	130.0E-3	41.0E-3	100.0E-3	1.06	4.00
muscle structure development	GO:0061061	muscle structure development	53.0E-3	110.0E-3	53.0E-3	110.0E-3	1.22	5.00
organelle assembly	GO:0070925	organelle assembly	75.0E-3	130.0E-3	75.0E-3	130.0E-3	1.10	5.00
positive regulation of MAPK cascade	GO:0008285	negative regulation of cell proliferation	130.0E-3	130.0E-3	26.0E-3	87.0E-3	1.01	4.00
	GO:0080135	regulation of cellular response to stress	90.0E-3	130.0E-3	26.0E-3	87.0E-3	1.17	4.00
	GO:0000165	MAPK cascade	70.0E-3	130.0E-3	26.0E-3	87.0E-3	1.02	6.00
	GO:0043410	positive regulation of MAPK cascade	99.0E-3	120.0E-3	26.0E-3	87.0E-3	1.13	4.00
	GO:0071900	regulation of protein serine/threonine kinase activity	93.0E-3	130.0E-3	26.0E-3	87.0E-3	1.15	4.00
single-organism catabolic process	GO:0044712	single-organism catabolic process	11.0E-3	59.0E-3	19.0E-3	86.0E-3	1.21	9.00
	GO:0006914	autophagy	48.0E-3	120.0E-3	19.0E-3	86.0E-3	1.44	4.00
	GO:0007005	mitochondrion organization	53.0E-3	110.0E-3	19.0E-3	86.0E-3	1.22	5.00
regionalization	GO:0007389	pattern specification process	82.0E-3	130.0E-3	82.0E-3	130.0E-3	1.21	4.00
	GO:0003002	regionalization	50.0E-3	110.0E-3	82.0E-3	130.0E-3	1.43	4.00
RNA modification*	GO:0043414	macromolecule methylation	14.0E-3	64.0E-3	17.0E-3	110.0E-3	2.11	4.00
	GO:0006396	RNA processing	61.0E-3	120.0E-3	17.0E-3	110.0E-3	1.06	6.00
	GO:0009451	RNA modification	1.5E-3	40.0E-3	17.0E-3	110.0E-3	4.00	4.00
	GO:0034660	ncRNA metabolic process	5.9E-3	77.0E-3	17.0E-3	110.0E-3	1.82	6.00
	GO:0034470	ncRNA processing	6.7E-3	58.0E-3	17.0E-3	110.0E-3	2.09	5.00

Table 3-12: Biological pathways for porcine spermatozoa. Showing represented and significantly upregulated transcripts in porcine spermatozoa compared to testis including the subsections and statistical involvement.

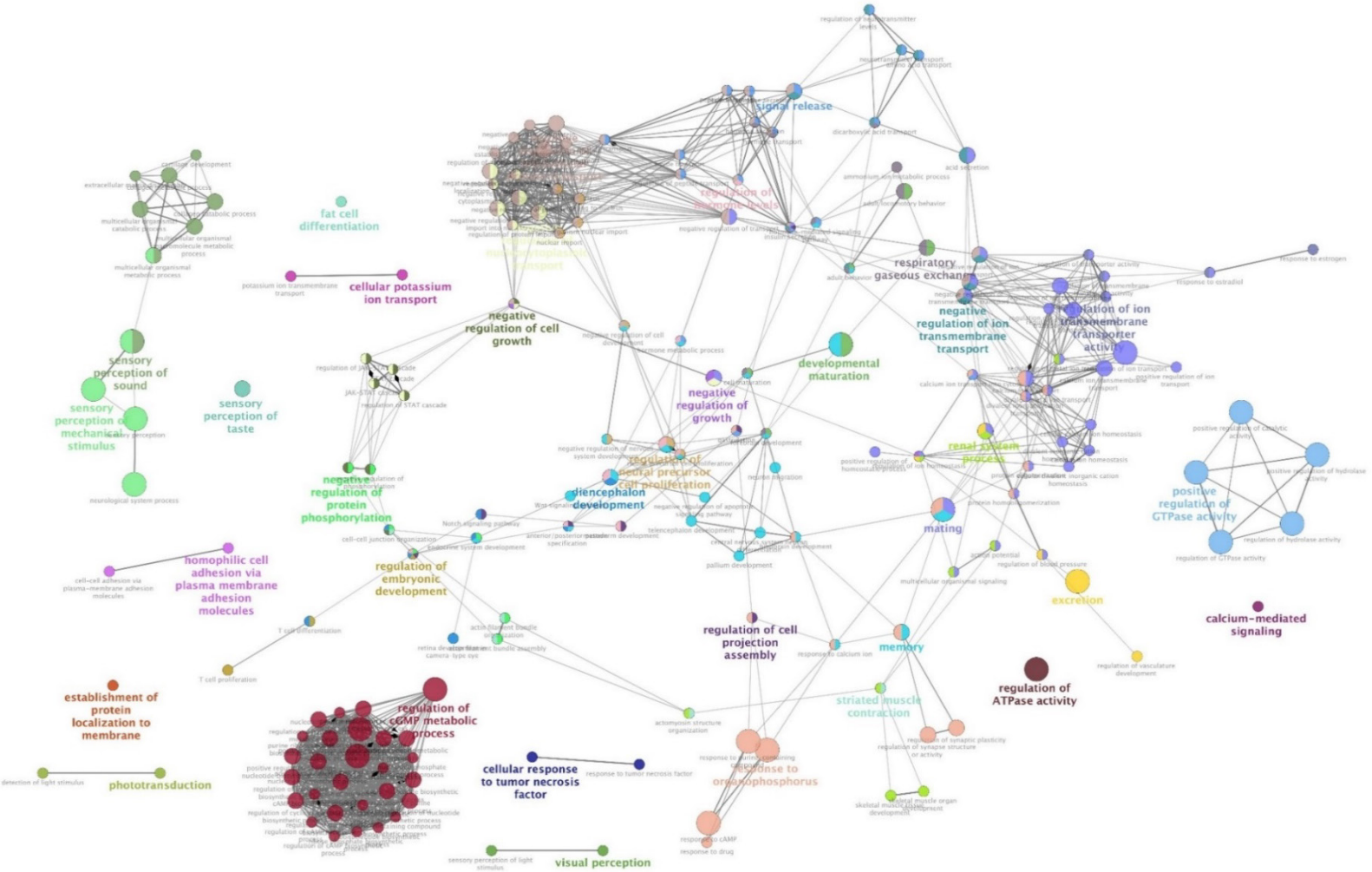


Figure 3-16: ClueGO network analysis of represented transcripts in human spermatozoa. Biological processes revealed 38 transcript clusters.

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
sensory perception of sound*	GO:0044236	multicellular organismal metabolic process	9,8E-3	60,0E-3	24,0E-3	76,0E-3	4,08	6,00
	GO:0044243	multicellular organismal catabolic process	22,0E-3	71,0E-3	24,0E-3	76,0E-3	4,60	4,00
	GO:0044259	multicellular organismal macromolecule metabolic process	23,0E-3	71,0E-3	24,0E-3	76,0E-3	3,79	5,00
	GO:0022617	extracellular matrix disassembly	100,0E-3	150,0E-3	24,0E-3	76,0E-3	2,76	4,00
	GO:0030574	collagen catabolic process	17,0E-3	71,0E-3	24,0E-3	76,0E-3	4,94	4,00
	GO:0032963	collagen metabolic process	20,0E-3	67,0E-3	24,0E-3	76,0E-3	3,94	5,00
	GO:0051216	cartilage development	310,0E-3	310,0E-3	24,0E-3	76,0E-3	2,03	4,00
	GO:0007605	sensory perception of sound	2,8E-3	39,0E-3	24,0E-3	76,0E-3	4,61	7,00
negative regulation of protein phosphorylation	GO:0045216	cell-cell junction organization	230,0E-3	250,0E-3	130,0E-3	170,0E-3	2,02	5,00
	GO:0035270	endocrine system development	140,0E-3	180,0E-3	130,0E-3	170,0E-3	2,44	4,00
	GO:0045995	regulation of embryonic development	64,0E-3	120,0E-3	130,0E-3	170,0E-3	3,28	4,00
	GO:0051017	actin filament bundle assembly	82,0E-3	140,0E-3	130,0E-3	170,0E-3	3,01	4,00
	GO:0061572	actin filament bundle organization	87,0E-3	140,0E-3	130,0E-3	170,0E-3	2,94	4,00
	GO:0042326	negative regulation of phosphorylation	120,0E-3	160,0E-3	130,0E-3	170,0E-3	2,05	9,00
	GO:0001933	negative regulation of protein phosphorylation	64,0E-3	120,0E-3	130,0E-3	170,0E-3	2,27	9,00
fat cell differentiation	GO:0045444	fat cell differentiation	56,0E-3	120,0E-3	56,0E-3	120,0E-3	2,71	6,00
negative regulation of growth	GO:0045926	negative regulation of growth	17,0E-3	71,0E-3	81,0E-3	130,0E-3	3,01	8,00
	GO:0030308	negative regulation of cell growth	77,0E-3	140,0E-3	81,0E-3	130,0E-3	2,72	5,00
	GO:2000021	regulation of ion homeostasis	190,0E-3	220,0E-3	81,0E-3	130,0E-3	2,35	5,00
negative regulation of ion transmembrane transport	GO:0001505	regulation of neurotransmitter levels	82,0E-3	140,0E-3	11,0E-3	64,0E-3	2,67	5,00
	GO:0030534	adult behavior	48,0E-3	110,0E-3	11,0E-3	64,0E-3	3,13	5,00
	GO:0006836	neurotransmitter transport	290,0E-3	310,0E-3	11,0E-3	64,0E-3	2,12	4,00
	GO:0034763	negative regulation of transmembrane transport	28,0E-3	80,0E-3	11,0E-3	64,0E-3	4,30	4,00
	GO:0043271	negative regulation of ion transport	27,0E-3	80,0E-3	11,0E-3	64,0E-3	3,65	5,00
	GO:0023061	signal release	18,0E-3	70,0E-3	11,0E-3	64,0E-3	2,50	12,00
	GO:0046717	acid secretion	23,0E-3	71,0E-3	11,0E-3	64,0E-3	3,79	5,00
	GO:0034766	negative regulation of ion transmembrane transport	20,0E-3	67,0E-3	11,0E-3	64,0E-3	4,76	4,00
	GO:0006865	amino acid transport	48,0E-3	110,0E-3	11,0E-3	64,0E-3	3,13	5,00
	GO:0006835	dicarboxylic acid transport	38,0E-3	100,0E-3	11,0E-3	64,0E-3	3,88	4,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	11,0E-3	64,0E-3	2,04	5,00
sensory perception of taste	GO:0050909	sensory perception of taste	12,0E-3	62,0E-3	12,0E-3	58,0E-3	5,56	4,00
visual perception	GO:0050953	sensory perception of light stimulus	74,0E-3	130,0E-3	74,0E-3	130,0E-3	2,52	6,00
	GO:0007601	visual perception	71,0E-3	130,0E-3	74,0E-3	130,0E-3	2,55	6,00
negative regulation of intracellular protein transport	GO:0010817	regulation of hormone levels	45,0E-3	110,0E-3	21,0E-3	79,0E-3	2,21	12,00
	GO:0051051	negative regulation of transport	22,0E-3	70,0E-3	21,0E-3	79,0E-3	2,41	12,00
	GO:1903828	negative regulation of cellular protein localization	33,0E-3	92,0E-3	21,0E-3	79,0E-3	3,45	5,00
	GO:1904950	negative regulation of establishment of protein localization	62,0E-3	120,0E-3	21,0E-3	79,0E-3	2,64	6,00
	GO:0009914	hormone transport	47,0E-3	110,0E-3	21,0E-3	79,0E-3	2,47	9,00
	GO:0030308	negative regulation of cell growth	77,0E-3	140,0E-3	21,0E-3	79,0E-3	2,72	5,00
	GO:0032387	negative regulation of intracellular transport	46,0E-3	110,0E-3	21,0E-3	79,0E-3	3,16	5,00
	GO:0042886	amide transport	150,0E-3	190,0E-3	21,0E-3	79,0E-3	2,05	8,00
	GO:0034763	negative regulation of transmembrane transport	28,0E-3	80,0E-3	21,0E-3	79,0E-3	4,30	4,00
	GO:0043271	negative regulation of ion transport	27,0E-3	80,0E-3	21,0E-3	79,0E-3	3,65	5,00
	GO:0051224	negative regulation of protein transport	52,0E-3	120,0E-3	21,0E-3	79,0E-3	2,76	6,00
	GO:0010721	negative regulation of cell development	130,0E-3	170,0E-3	21,0E-3	79,0E-3	2,11	7,00
	GO:0015833	peptide transport	91,0E-3	140,0E-3	21,0E-3	79,0E-3	2,20	8,00
	GO:0023061	signal release	18,0E-3	70,0E-3	21,0E-3	79,0E-3	2,50	12,00
	GO:0090087	regulation of peptide transport	180,0E-3	220,0E-3	21,0E-3	79,0E-3	2,02	6,00
	GO:0090317	negative regulation of intracellular protein transport	11,0E-3	64,0E-3	21,0E-3	79,0E-3	4,55	5,00
	GO:1900181	negative regulation of protein localization to nucleus	18,0E-3	68,0E-3	21,0E-3	79,0E-3	4,88	4,00
	GO:1903650	negative regulation of cytoplasmic transport	75,0E-3	130,0E-3	21,0E-3	79,0E-3	3,10	4,00
	GO:0002790	peptide secretion	190,0E-3	220,0E-3	21,0E-3	79,0E-3	2,00	6,00
	GO:0034766	negative regulation of ion transmembrane transport	20,0E-3	67,0E-3	21,0E-3	79,0E-3	4,76	4,00
	GO:0044744	protein targeting to nucleus	180,0E-3	220,0E-3	21,0E-3	79,0E-3	2,04	6,00
	GO:1904589	regulation of protein import	300,0E-3	310,0E-3	21,0E-3	79,0E-3	2,06	4,00
	GO:1904590	negative regulation of protein import	14,0E-3	67,0E-3	21,0E-3	79,0E-3	5,26	4,00
	GO:1902593	single-organism nuclear import	180,0E-3	220,0E-3	21,0E-3	79,0E-3	2,04	6,00
	GO:0006606	protein import into nucleus	180,0E-3	220,0E-3	21,0E-3	79,0E-3	2,04	6,00
	GO:0046823	negative regulation of nucleocytoplasmic transport	20,0E-3	67,0E-3	21,0E-3	79,0E-3	4,71	4,00
	GO:0046879	hormone secretion	43,0E-3	110,0E-3	21,0E-3	79,0E-3	2,54	9,00
	GO:0051170	nuclear import	180,0E-3	220,0E-3	21,0E-3	79,0E-3	2,01	6,00
	GO:1903533	regulation of protein targeting	66,0E-3	130,0E-3	21,0E-3	79,0E-3	2,45	8,00
	GO:0030072	peptide hormone secretion	170,0E-3	210,0E-3	21,0E-3	79,0E-3	2,05	6,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	21,0E-3	79,0E-3	2,04	5,00
	GO:0042306	regulation of protein import into nucleus	300,0E-3	310,0E-3	21,0E-3	79,0E-3	2,11	4,00
	GO:0042308	negative regulation of protein import into nucleus	14,0E-3	67,0E-3	21,0E-3	79,0E-3	5,26	4,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
positive regulation of GTPase activity*	GO:0043085	positive regulation of catalytic activity	1,2E-3	27,0E-3	4,9E-3	84,0E-3	2,15	35,00
	GO:0051336	regulation of hydrolase activity	2,7E-3	42,0E-3	4,9E-3	84,0E-3	2,17	29,00
	GO:0051345	positive regulation of hydrolase activity	100,0E-6	16,0E-3	4,9E-3	84,0E-3	2,83	25,00
	GO:0043087	regulation of GTPase activity	400,0E-6	20,0E-3	4,9E-3	84,0E-3	3,04	18,00
	GO:0043547	positive regulation of GTPase activity	130,0E-6	10,0E-3	4,9E-3	84,0E-3	3,33	18,00
response to organophosphorus	GO:0042493	response to drug	58,0E-3	120,0E-3	14,0E-3	59,0E-3	2,21	11,00
	GO:0050803	regulation of synapse structure or activity	16,0E-3	69,0E-3	14,0E-3	59,0E-3	3,05	8,00
	GO:0007618	mating	3,5E-3	39,0E-3	14,0E-3	59,0E-3	7,84	4,00
	GO:0046683	response to organophosphorus	520,0E-6	20,0E-3	14,0E-3	59,0E-3	5,37	8,00
	GO:0014074	response to purine-containing compound	1,2E-3	25,0E-3	14,0E-3	59,0E-3	4,68	8,00
	GO:0060491	regulation of cell projection assembly	59,0E-3	120,0E-3	14,0E-3	59,0E-3	2,94	5,00
	GO:0007369	gastrulation	300,0E-3	310,0E-3	14,0E-3	59,0E-3	2,05	4,00
	GO:0007613	memory	15,0E-3	68,0E-3	14,0E-3	59,0E-3	4,24	5,00
	GO:0051591	response to cAMP	2,5E-3	43,0E-3	14,0E-3	59,0E-3	5,41	6,00
	GO:0051592	response to calcium ion	69,0E-3	130,0E-3	14,0E-3	59,0E-3	3,20	4,00
	GO:0030902	hindbrain development	280,0E-3	300,0E-3	14,0E-3	59,0E-3	2,22	4,00
	GO:0051259	protein oligomerization	86,0E-3	140,0E-3	14,0E-3	59,0E-3	2,12	10,00
	GO:0048167	regulation of synaptic plasticity	18,0E-3	71,0E-3	14,0E-3	59,0E-3	3,55	6,00
	GO:0051260	protein homooligomerization	160,0E-3	200,0E-3	14,0E-3	59,0E-3	2,11	6,00
	GO:0072511	divalent inorganic cation transport	74,0E-3	130,0E-3	14,0E-3	59,0E-3	2,24	10,00
	GO:0070838	divalent metal ion transport	73,0E-3	130,0E-3	14,0E-3	59,0E-3	2,26	10,00
	GO:0006816	calcium ion transport	32,0E-3	88,0E-3	14,0E-3	59,0E-3	2,53	10,00
	GO:0060402	calcium ion transport into cytosol	95,0E-3	140,0E-3	14,0E-3	59,0E-3	2,86	4,00
striated muscle contraction	GO:0045216	cell-cell junction organization	230,0E-3	250,0E-3	41,0E-3	100,0E-3	2,02	5,00
	GO:0006941	striated muscle contraction	45,0E-3	110,0E-3	41,0E-3	100,0E-3	3,18	5,00
	GO:0031032	actomyosin structure organization	130,0E-3	170,0E-3	41,0E-3	100,0E-3	2,53	4,00
	GO:0051017	actin filament bundle assembly	82,0E-3	140,0E-3	41,0E-3	100,0E-3	3,01	4,00
	GO:0061572	actin filament bundle organization	87,0E-3	140,0E-3	41,0E-3	100,0E-3	2,94	4,00
cellular response to tumor necrosis factor	GO:0034612	response to tumor necrosis factor	180,0E-3	220,0E-3	180,0E-3	220,0E-3	2,03	6,00
	GO:0071356	cellular response to tumor necrosis factor	150,0E-3	190,0E-3	180,0E-3	220,0E-3	2,21	6,00
cellular potassium ion transport	GO:0071804	cellular potassium ion transport	280,0E-3	300,0E-3	280,0E-3	290,0E-3	2,23	4,00
	GO:0071805	potassium ion transmembrane transport	280,0E-3	300,0E-3	280,0E-3	290,0E-3	2,23	4,00
establishment of protein localization to membrane	GO:0090150	establishment of protein localization to membrane	150,0E-3	190,0E-3	150,0E-3	190,0E-3	2,05	8,00
homophilic cell adhesion via plasma membrane adhesion molecules	GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	210,0E-3	230,0E-3	210,0E-3	250,0E-3	2,16	5,00
	GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	140,0E-3	180,0E-3	210,0E-3	250,0E-3	2,45	4,00
regulation of cGMP metabolic process*	GO:0072522	purine-containing compound biosynthetic process	95,0E-3	140,0E-3	120,0E-3	180,0E-3	2,41	7,00
	GO:0046390	ribose phosphate biosynthetic process	94,0E-3	140,0E-3	120,0E-3	180,0E-3	2,42	7,00
	GO:1901293	nucleoside phosphate biosynthetic process	200,0E-3	230,0E-3	120,0E-3	180,0E-3	2,06	7,00
	GO:0006140	regulation of nucleotide metabolic process	8,4E-3	56,0E-3	120,0E-3	180,0E-3	3,43	8,00
	GO:0009165	nucleotide biosynthetic process	200,0E-3	230,0E-3	120,0E-3	180,0E-3	2,09	7,00
	GO:0009187	cyclic nucleotide metabolic process	17,0E-3	73,0E-3	120,0E-3	180,0E-3	3,26	7,00
	GO:0030808	regulation of nucleotide biosynthetic process	11,0E-3	63,0E-3	120,0E-3	180,0E-3	3,97	6,00
	GO:0045981	positive regulation of nucleotide metabolic process	8,3E-3	58,0E-3	120,0E-3	180,0E-3	4,23	6,00
	GO:0006164	purine nucleotide biosynthetic process	87,0E-3	140,0E-3	120,0E-3	180,0E-3	2,53	7,00
	GO:0009190	cyclic nucleotide biosynthetic process	19,0E-3	70,0E-3	120,0E-3	180,0E-3	3,49	6,00
	GO:0009260	ribonucleotide biosynthetic process	92,0E-3	140,0E-3	120,0E-3	180,0E-3	2,45	7,00
	GO:0030799	regulation of cyclic nucleotide metabolic process	5,0E-3	48,0E-3	120,0E-3	180,0E-3	4,14	7,00
	GO:0030810	positive regulation of nucleotide biosynthetic process	12,0E-3	60,0E-3	120,0E-3	180,0E-3	4,50	5,00
	GO:1900542	regulation of purine nucleotide metabolic process	6,8E-3	50,0E-3	120,0E-3	180,0E-3	3,56	8,00
	GO:0030801	positive regulation of cyclic nucleotide metabolic process	14,0E-3	65,0E-3	120,0E-3	180,0E-3	4,31	5,00
	GO:0030802	regulation of cyclic nucleotide biosynthetic process	9,2E-3	59,0E-3	120,0E-3	180,0E-3	4,14	6,00
	GO:1900371	regulation of purine nucleotide biosynthetic process	11,0E-3	63,0E-3	120,0E-3	180,0E-3	3,97	6,00
	GO:1900544	positive regulation of purine nucleotide metabolic process	8,3E-3	58,0E-3	120,0E-3	180,0E-3	4,23	6,00
	GO:0009152	purine ribonucleotide biosynthetic process	51,0E-3	120,0E-3	120,0E-3	180,0E-3	2,58	7,00
	GO:0030804	positive regulation of cyclic nucleotide biosynthetic process	10,0E-3	60,0E-3	120,0E-3	180,0E-3	4,72	5,00
	GO:0046058	cAMP metabolic process	140,0E-3	180,0E-3	120,0E-3	180,0E-3	2,42	4,00
	GO:0046068	cGMP metabolic process	4,3E-3	45,0E-3	120,0E-3	180,0E-3	7,41	4,00
	GO:0052652	cyclic purine nucleotide metabolic process	19,0E-3	70,0E-3	120,0E-3	180,0E-3	3,51	6,00
	GO:1900373	positive regulation of purine nucleotide biosynthetic process	12,0E-3	60,0E-3	120,0E-3	180,0E-3	4,50	5,00
	GO:0030814	regulation of cAMP metabolic process	95,0E-3	140,0E-3	120,0E-3	180,0E-3	2,86	4,00
	GO:0030823	regulation of cGMP metabolic process	690,0E-6	21,0E-3	120,0E-3	180,0E-3	12,12	4,00
	GO:0006171	cAMP biosynthetic process	84,0E-3	140,0E-3	120,0E-3	180,0E-3	2,99	4,00
	GO:0030817	regulation of cAMP biosynthetic process	64,0E-3	120,0E-3	120,0E-3	180,0E-3	3,28	4,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
negative regulation of cell growth	GO:0045216	cell-cell junction organization	230,0E-3	250,0E-3	280,0E-3	290,0E-3	2,02	5,00
	GO:0030308	negative regulation of cell growth	77,0E-3	140,0E-3	280,0E-3	290,0E-3	2,72	5,00
	GO:0097696	STAT cascade	300,0E-3	310,0E-3	280,0E-3	290,0E-3	2,11	4,00
	GO:1904892	regulation of STAT cascade	120,0E-3	160,0E-3	280,0E-3	290,0E-3	2,61	4,00
	GO:0007259	JAK-STAT cascade	300,0E-3	310,0E-3	280,0E-3	290,0E-3	2,11	4,00
	GO:0042326	negative regulation of phosphorylation	120,0E-3	160,0E-3	280,0E-3	290,0E-3	2,05	9,00
	GO:0046425	regulation of JAK-STAT cascade	120,0E-3	160,0E-3	280,0E-3	290,0E-3	2,61	4,00
	GO:0001933	negative regulation of protein phosphorylation	64,0E-3	120,0E-3	280,0E-3	290,0E-3	2,27	9,00
mating*	GO:0003014	renal system process	19,0E-3	68,0E-3	10,0E-3	74,0E-3	3,97	5,00
	GO:0035637	multicellular organismal signaling	310,0E-3	310,0E-3	10,0E-3	74,0E-3	2,03	4,00
	GO:0045926	negative regulation of growth	17,0E-3	71,0E-3	10,0E-3	74,0E-3	3,01	8,00
	GO:0032846	positive regulation of homeostatic process	62,0E-3	120,0E-3	10,0E-3	74,0E-3	2,64	6,00
	GO:0051051	negative regulation of transport	22,0E-3	70,0E-3	10,0E-3	74,0E-3	2,41	12,00
	GO:0001508	action potential	78,0E-3	140,0E-3	10,0E-3	74,0E-3	3,05	4,00
	GO:0007618	mating	3,5E-3	39,0E-3	10,0E-3	74,0E-3	7,84	4,00
	GO:0008217	regulation of blood pressure	310,0E-3	310,0E-3	10,0E-3	74,0E-3	2,02	4,00
	GO:0032409	regulation of transporter activity	110,0E-3	160,0E-3	10,0E-3	74,0E-3	2,39	5,00
	GO:0034762	regulation of transmembrane transport	13,0E-3	64,0E-3	10,0E-3	74,0E-3	2,68	12,00
	GO:2000021	regulation of ion homeostasis	190,0E-3	220,0E-3	10,0E-3	74,0E-3	2,35	5,00
	GO:0034763	negative regulation of transmembrane transport	28,0E-3	80,0E-3	10,0E-3	74,0E-3	4,30	4,00
	GO:0043269	regulation of ion transport	5,1E-3	46,0E-3	10,0E-3	74,0E-3	2,56	16,00
	GO:0043270	positive regulation of ion transport	130,0E-3	170,0E-3	10,0E-3	74,0E-3	2,39	6,00
	GO:0043271	negative regulation of ion transport	27,0E-3	80,0E-3	10,0E-3	74,0E-3	3,65	5,00
	GO:0022898	regulation of transmembrane transporter activity	92,0E-3	140,0E-3	10,0E-3	74,0E-3	2,58	5,00
	GO:0034765	regulation of ion transmembrane transport	11,0E-3	62,0E-3	10,0E-3	74,0E-3	2,79	12,00
	GO:0046717	acid secretion	23,0E-3	71,0E-3	10,0E-3	74,0E-3	3,79	5,00
	GO:0051259	protein oligomerization	86,0E-3	140,0E-3	10,0E-3	74,0E-3	2,12	10,00
	GO:0032355	response to estradiol	120,0E-3	170,0E-3	10,0E-3	74,0E-3	2,55	4,00
	GO:0034766	negative regulation of ion transmembrane transport	20,0E-3	67,0E-3	10,0E-3	74,0E-3	4,76	4,00
	GO:0010959	regulation of metal ion transport	79,0E-3	130,0E-3	10,0E-3	74,0E-3	2,31	8,00
	GO:0032412	regulation of ion transmembrane transporter activity	86,0E-3	140,0E-3	10,0E-3	74,0E-3	2,63	5,00
	GO:0051260	protein homooligomerization	160,0E-3	200,0E-3	10,0E-3	74,0E-3	2,11	6,00
	GO:0072511	divalent inorganic cation transport	74,0E-3	130,0E-3	10,0E-3	74,0E-3	2,24	10,00
	GO:0072507	divalent inorganic cation homeostasis	110,0E-3	160,0E-3	10,0E-3	74,0E-3	2,08	9,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	10,0E-3	74,0E-3	2,04	5,00
	GO:0070838	divalent metal ion transport	73,0E-3	130,0E-3	10,0E-3	74,0E-3	2,26	10,00
	GO:0072503	cellular divalent inorganic cation homeostasis	100,0E-3	150,0E-3	10,0E-3	74,0E-3	2,18	9,00
	GO:0051924	regulation of calcium ion transport	58,0E-3	120,0E-3	10,0E-3	74,0E-3	2,69	6,00
	GO:0055074	calcium ion homeostasis	100,0E-3	150,0E-3	10,0E-3	74,0E-3	2,22	9,00
	GO:0006816	calcium ion transport	32,0E-3	88,0E-3	10,0E-3	74,0E-3	2,53	10,00
	GO:0006874	cellular calcium ion homeostasis	62,0E-3	120,0E-3	10,0E-3	74,0E-3	2,29	9,00
	GO:0070588	calcium ion transmembrane transport	17,0E-3	71,0E-3	10,0E-3	74,0E-3	3,01	8,00
	GO:0060402	calcium ion transport into cytosol	95,0E-3	140,0E-3	10,0E-3	74,0E-3	2,86	4,00
regulation of cell projection assembly	GO:0007498	mesoderm development	80,0E-3	140,0E-3	100,0E-3	160,0E-3	3,03	4,00
	GO:0045995	regulation of embryonic development	64,0E-3	120,0E-3	100,0E-3	160,0E-3	3,28	4,00
	GO:0060491	regulation of cell projection assembly	59,0E-3	120,0E-3	100,0E-3	160,0E-3	2,94	5,00
	GO:0007369	gastrulation	300,0E-3	310,0E-3	100,0E-3	160,0E-3	2,05	4,00
	GO:0007219	Notch signaling pathway	190,0E-3	220,0E-3	100,0E-3	160,0E-3	2,33	5,00
	GO:0009952	anterior/posterior pattern specification	200,0E-3	230,0E-3	100,0E-3	160,0E-3	2,24	5,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	100,0E-3	160,0E-3	2,04	5,00
renal system process	GO:0003014	renal system process	19,0E-3	68,0E-3	23,0E-3	78,0E-3	3,97	5,00
	GO:0035637	multicellular organismal signaling	310,0E-3	310,0E-3	23,0E-3	78,0E-3	2,03	4,00
	GO:0001508	action potential	78,0E-3	140,0E-3	23,0E-3	78,0E-3	3,05	4,00
	GO:2000021	regulation of ion homeostasis	190,0E-3	220,0E-3	23,0E-3	78,0E-3	2,35	5,00
	GO:0006941	striated muscle contraction	45,0E-3	110,0E-3	23,0E-3	78,0E-3	3,18	5,00
	GO:0031032	actomyosin structure organization	130,0E-3	170,0E-3	23,0E-3	78,0E-3	2,53	4,00
	GO:0060538	skeletal muscle organ development	310,0E-3	310,0E-3	23,0E-3	78,0E-3	2,02	4,00
	GO:0010959	regulation of metal ion transport	79,0E-3	130,0E-3	23,0E-3	78,0E-3	2,31	8,00
	GO:0007519	skeletal muscle tissue development	290,0E-3	310,0E-3	23,0E-3	78,0E-3	2,15	4,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
developmental maturation*	GO:0007585	respiratory gaseous exchange	12,0E-3	62,0E-3	31,0E-3	89,0E-3	5,56	4,00
	GO:0021700	developmental maturation	5,1E-3	44,0E-3	31,0E-3	89,0E-3	3,45	9,00
	GO:0030534	adult behavior	48,0E-3	110,0E-3	31,0E-3	89,0E-3	3,13	5,00
	GO:0008344	adult locomotory behavior	30,0E-3	87,0E-3	31,0E-3	89,0E-3	4,17	4,00
	GO:0048469	cell maturation	62,0E-3	130,0E-3	31,0E-3	89,0E-3	2,91	5,00
	GO:0030900	forebrain development	100,0E-3	150,0E-3	31,0E-3	89,0E-3	2,14	9,00
sensory perception of mechanical stimulus*	GO:0044236	multicellular organismal metabolic process	9,8E-3	60,0E-3	3,2E-3	100,0E-3	4,08	6,00
	GO:0050877	neurological system process	5,1E-3	42,0E-3	3,2E-3	100,0E-3	2,07	29,00
	GO:0007600	sensory perception	2,9E-3	35,0E-3	3,2E-3	100,0E-3	2,34	24,00
	GO:0050954	sensory perception of mechanical stimulus	1,1E-3	30,0E-3	3,2E-3	100,0E-3	4,73	8,00
	GO:0007605	sensory perception of sound	2,8E-3	39,0E-3	3,2E-3	100,0E-3	4,61	7,00
excretion*	GO:0003014	renal system process	19,0E-3	68,0E-3	46,0E-3	110,0E-3	3,97	5,00
	GO:0007588	excretion	2,9E-3	37,0E-3	46,0E-3	110,0E-3	6,33	5,00
	GO:0008217	regulation of blood pressure	310,0E-3	310,0E-3	46,0E-3	110,0E-3	2,02	4,00
	GO:2000021	regulation of ion homeostasis	190,0E-3	220,0E-3	46,0E-3	110,0E-3	2,35	5,00
	GO:1901342	regulation of vasculature development	230,0E-3	250,0E-3	46,0E-3	110,0E-3	2,01	5,00
phototransduction	GO:0009583	detection of light stimulus	100,0E-3	150,0E-3	100,0E-3	160,0E-3	2,72	4,00
	GO:0007602	phototransduction	78,0E-3	140,0E-3	100,0E-3	160,0E-3	3,05	4,00
regulation of neural precursor cell proliferation	GO:0061351	neural precursor cell proliferation	130,0E-3	170,0E-3	290,0E-3	290,0E-3	2,52	4,00
	GO:2000177	regulation of neural precursor cell proliferation	23,0E-3	71,0E-3	290,0E-3	290,0E-3	4,55	4,00
	GO:0051961	negative regulation of nervous system development	170,0E-3	210,0E-3	290,0E-3	290,0E-3	2,06	6,00
	GO:0010721	negative regulation of cell development	130,0E-3	170,0E-3	290,0E-3	290,0E-3	2,11	7,00
	GO:0044744	protein targeting to nucleus	180,0E-3	220,0E-3	290,0E-3	290,0E-3	2,04	6,00
	GO:1902593	single-organism nuclear import	180,0E-3	220,0E-3	290,0E-3	290,0E-3	2,04	6,00
	GO:0006606	protein import into nucleus	180,0E-3	220,0E-3	290,0E-3	290,0E-3	2,04	6,00
	GO:0051170	nuclear import	180,0E-3	220,0E-3	290,0E-3	290,0E-3	2,01	6,00
signal release	GO:0001505	regulation of neurotransmitter levels	82,0E-3	140,0E-3	63,0E-3	120,0E-3	2,67	5,00
	GO:0010817	regulation of hormone levels	45,0E-3	110,0E-3	63,0E-3	120,0E-3	2,21	12,00
	GO:0042445	hormone metabolic process	110,0E-3	160,0E-3	63,0E-3	120,0E-3	2,38	5,00
	GO:0006836	neurotransmitter transport	290,0E-3	310,0E-3	63,0E-3	120,0E-3	2,12	4,00
	GO:0009914	hormone transport	47,0E-3	110,0E-3	63,0E-3	120,0E-3	2,47	9,00
	GO:0042886	amide transport	150,0E-3	190,0E-3	63,0E-3	120,0E-3	2,05	8,00
	GO:0007369	gastrulation	300,0E-3	310,0E-3	63,0E-3	120,0E-3	2,05	4,00
	GO:0009755	hormone-mediated signaling pathway	56,0E-3	120,0E-3	63,0E-3	120,0E-3	3,42	4,00
	GO:0015833	peptide transport	91,0E-3	140,0E-3	63,0E-3	120,0E-3	2,20	8,00
	GO:0023061	signal release	18,0E-3	70,0E-3	63,0E-3	120,0E-3	2,50	12,00
	GO:0090087	regulation of peptide transport	180,0E-3	220,0E-3	63,0E-3	120,0E-3	2,02	6,00
	GO:0002790	peptide secretion	190,0E-3	220,0E-3	63,0E-3	120,0E-3	2,00	6,00
	GO:0046879	hormone secretion	43,0E-3	110,0E-3	63,0E-3	120,0E-3	2,54	9,00
	GO:0006865	amino acid transport	48,0E-3	110,0E-3	63,0E-3	120,0E-3	3,13	5,00
	GO:0006835	dicarboxylic acid transport	38,0E-3	100,0E-3	63,0E-3	120,0E-3	3,88	4,00
	GO:0030072	peptide hormone secretion	170,0E-3	210,0E-3	63,0E-3	120,0E-3	2,05	6,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	63,0E-3	120,0E-3	2,04	5,00
	GO:0060402	calcium ion transport into cytosol	95,0E-3	140,0E-3	63,0E-3	120,0E-3	2,86	4,00
calcium-mediated signaling	GO:0019722	calcium-mediated signaling	100,0E-3	150,0E-3	100,0E-3	160,0E-3	2,72	4,00
memory	GO:0021700	developmental maturation	5,1E-3	44,0E-3	5,0E-3	57,0E-3	3,45	9,00
	GO:0030534	adult behavior	48,0E-3	110,0E-3	5,0E-3	57,0E-3	3,13	5,00
	GO:0042445	hormone metabolic process	110,0E-3	160,0E-3	5,0E-3	57,0E-3	2,38	5,00
	GO:0061351	neural precursor cell proliferation	130,0E-3	170,0E-3	5,0E-3	57,0E-3	2,52	4,00
	GO:0007618	mating	3,5E-3	39,0E-3	5,0E-3	57,0E-3	7,84	4,00
	GO:0035270	endocrine system development	140,0E-3	180,0E-3	5,0E-3	57,0E-3	2,44	4,00
	GO:2000177	regulation of neural precursor cell proliferation	23,0E-3	71,0E-3	5,0E-3	57,0E-3	4,55	4,00
	GO:0007613	memory	15,0E-3	68,0E-3	5,0E-3	57,0E-3	4,24	5,00
	GO:0009755	hormone-mediated signaling pathway	56,0E-3	120,0E-3	5,0E-3	57,0E-3	3,42	4,00
	GO:0048469	cell maturation	62,0E-3	130,0E-3	5,0E-3	57,0E-3	2,91	5,00
	GO:0051592	response to calcium ion	69,0E-3	130,0E-3	5,0E-3	57,0E-3	3,20	4,00
	GO:0051961	negative regulation of nervous system development	170,0E-3	210,0E-3	5,0E-3	57,0E-3	2,06	6,00
	GO:0009952	anterior/posterior pattern specification	200,0E-3	230,0E-3	5,0E-3	57,0E-3	2,24	5,00
	GO:0010721	negative regulation of cell development	130,0E-3	170,0E-3	5,0E-3	57,0E-3	2,11	7,00
	GO:0016055	Wnt signaling pathway	86,0E-3	140,0E-3	5,0E-3	57,0E-3	2,12	10,00
	GO:0030900	forebrain development	100,0E-3	150,0E-3	5,0E-3	57,0E-3	2,14	9,00
	GO:0030902	hindbrain development	280,0E-3	300,0E-3	5,0E-3	57,0E-3	2,22	4,00
	GO:0001764	neuron migration	120,0E-3	170,0E-3	5,0E-3	57,0E-3	2,55	4,00
	GO:0021536	diencephalon development	28,0E-3	80,0E-3	5,0E-3	57,0E-3	4,30	4,00
	GO:0021537	telencephalon development	49,0E-3	110,0E-3	5,0E-3	57,0E-3	2,60	7,00
	GO:2001234	negative regulation of apoptotic signaling pathway	210,0E-3	230,0E-3	5,0E-3	57,0E-3	2,16	5,00
	GO:0021543	pallium development	280,0E-3	300,0E-3	5,0E-3	57,0E-3	2,21	4,00
	GO:0021953	central nervous system neuron differentiation	95,0E-3	140,0E-3	5,0E-3	57,0E-3	2,55	5,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
regulation of embryonic development	GO:0045995	regulation of embryonic development	64,0E-3	120,0E-3	270,0E-3	300,0E-3	3,28	4,00
	GO:0042098	T cell proliferation	300,0E-3	310,0E-3	270,0E-3	300,0E-3	2,09	4,00
	GO:0030217	T cell differentiation	200,0E-3	230,0E-3	270,0E-3	300,0E-3	2,23	5,00
negative regulation of nucleocytoplasmic transport	GO:0045926	negative regulation of growth	17,0E-3	71,0E-3	140,0E-3	180,0E-3	3,01	8,00
	GO:0030308	negative regulation of cell growth	77,0E-3	140,0E-3	140,0E-3	180,0E-3	2,72	5,00
	GO:0097696	STAT cascade	300,0E-3	310,0E-3	140,0E-3	180,0E-3	2,11	4,00
	GO:1900181	negative regulation of protein localization to nucleus	18,0E-3	68,0E-3	140,0E-3	180,0E-3	4,88	4,00
	GO:1903650	negative regulation of cytoplasmic transport	75,0E-3	130,0E-3	140,0E-3	180,0E-3	3,10	4,00
	GO:1904589	regulation of protein import	300,0E-3	310,0E-3	140,0E-3	180,0E-3	2,06	4,00
	GO:1904590	negative regulation of protein import	14,0E-3	67,0E-3	140,0E-3	180,0E-3	5,26	4,00
	GO:1904892	regulation of STAT cascade	120,0E-3	160,0E-3	140,0E-3	180,0E-3	2,61	4,00
	GO:0007259	JAK-STAT cascade	300,0E-3	310,0E-3	140,0E-3	180,0E-3	2,11	4,00
	GO:0046823	negative regulation of nucleocytoplasmic transport	20,0E-3	67,0E-3	140,0E-3	180,0E-3	4,71	4,00
	GO:0046425	regulation of JAK-STAT cascade	120,0E-3	160,0E-3	140,0E-3	180,0E-3	2,61	4,00
	GO:0042306	regulation of protein import into nucleus	300,0E-3	310,0E-3	140,0E-3	180,0E-3	2,11	4,00
	GO:0042308	negative regulation of protein import into nucleus	14,0E-3	67,0E-3	140,0E-3	180,0E-3	5,26	4,00
respiratory gaseous exchange	GO:0007585	respiratory gaseous exchange	12,0E-3	62,0E-3	120,0E-3	170,0E-3	5,56	4,00
	GO:0030534	adult behavior	48,0E-3	110,0E-3	120,0E-3	170,0E-3	3,13	5,00
	GO:0097164	ammonium ion metabolic process	310,0E-3	310,0E-3	120,0E-3	170,0E-3	2,01	4,00
	GO:0008344	adult locomotory behavior	30,0E-3	87,0E-3	120,0E-3	170,0E-3	4,17	4,00
	GO:0009914	hormone transport	47,0E-3	110,0E-3	120,0E-3	170,0E-3	2,47	9,00
	GO:0009755	hormone-mediated signaling pathway	56,0E-3	120,0E-3	120,0E-3	170,0E-3	3,42	4,00
	GO:0048469	cell maturation	62,0E-3	130,0E-3	120,0E-3	170,0E-3	2,91	5,00
	GO:0030900	forebrain development	100,0E-3	150,0E-3	120,0E-3	170,0E-3	2,14	9,00
	GO:0046879	hormone secretion	43,0E-3	110,0E-3	120,0E-3	170,0E-3	2,54	9,00
	GO:0006835	dicarboxylic acid transport	38,0E-3	100,0E-3	120,0E-3	170,0E-3	3,88	4,00
	GO:0030073	insulin secretion	230,0E-3	250,0E-3	120,0E-3	170,0E-3	2,04	5,00
regulation of ion transmembrane transporter activity	GO:0032409	regulation of transporter activity	110,0E-3	160,0E-3	260,0E-3	290,0E-3	2,39	5,00
	GO:0022898	regulation of transmembrane transporter activity	92,0E-3	140,0E-3	260,0E-3	290,0E-3	2,58	5,00
	GO:0043627	response to estrogen	220,0E-3	250,0E-3	260,0E-3	290,0E-3	2,06	5,00
	GO:0032355	response to estradiol	120,0E-3	170,0E-3	260,0E-3	290,0E-3	2,55	4,00
	GO:0032412	regulation of ion transmembrane transporter activity	86,0E-3	140,0E-3	260,0E-3	290,0E-3	2,63	5,00
diencephalon development	GO:0045216	cell-cell junction organization	230,0E-3	250,0E-3	70,0E-3	130,0E-3	2,02	5,00
	GO:0035270	endocrine system development	140,0E-3	180,0E-3	70,0E-3	130,0E-3	2,44	4,00
	GO:0045995	regulation of embryonic development	64,0E-3	120,0E-3	70,0E-3	130,0E-3	3,28	4,00
	GO:0007219	Notch signaling pathway	190,0E-3	220,0E-3	70,0E-3	130,0E-3	2,33	5,00
	GO:0009952	anterior/posterior pattern specification	200,0E-3	230,0E-3	70,0E-3	130,0E-3	2,24	5,00
	GO:0021536	diencephalon development	28,0E-3	80,0E-3	70,0E-3	130,0E-3	4,30	4,00
	GO:0060041	retina development in camera-type eye	53,0E-3	120,0E-3	70,0E-3	130,0E-3	3,03	5,00
	GO:0030217	T cell differentiation	200,0E-3	230,0E-3	70,0E-3	130,0E-3	2,23	5,00
regulation of hormone levels	GO:0010817	regulation of hormone levels	45,0E-3	110,0E-3	57,0E-3	120,0E-3	2,21	12,00
	GO:0042445	hormone metabolic process	110,0E-3	160,0E-3	57,0E-3	120,0E-3	2,38	5,00
	GO:0061351	neural precursor cell proliferation	130,0E-3	170,0E-3	57,0E-3	120,0E-3	2,52	4,00
	GO:0007498	mesoderm development	80,0E-3	140,0E-3	57,0E-3	120,0E-3	3,03	4,00
	GO:0045995	regulation of embryonic development	64,0E-3	120,0E-3	57,0E-3	120,0E-3	3,28	4,00
	GO:2000177	regulation of neural precursor cell proliferation	23,0E-3	71,0E-3	57,0E-3	120,0E-3	4,55	4,00
	GO:0009952	anterior/posterior pattern specification	200,0E-3	230,0E-3	57,0E-3	120,0E-3	2,24	5,00
	GO:0021536	diencephalon development	28,0E-3	80,0E-3	57,0E-3	120,0E-3	4,30	4,00
regulation of ATPase activity*	GO:0043462	regulation of ATPase activity	6,3E-3	49,0E-3	6,3E-3	54,0E-3	6,67	4,00

Table 3-13: Biological pathways for human spermatozoa. Showing represented and significantly upregulated transcripts in human spermatozoa compared to testis including the subsections and statistical involvement. Represented spermatozoal transcripts were involved in 38 clusters of biological processes.

3.4.2.5 Transcriptomic inter-species NGS validation

3.4.2.5.1 Transcriptome commonality between bovine, ovine, porcine and human spermatozoa

The following step was an inter-species comparison of the bovine, ovine, porcine and human spermatozoal RNA. Analysing commonality in-between the species the first 1000cpm most abundant reads of each species were used for the comparison. The highest level of commonality was between porcine and ovine (34.80%), followed by porcine and bovine (34.60%) (see Table 3-14). Human spermatozoal samples showed a balanced level of commonality among the other three species (22%-28%) where ovine was the most comparable to human. Surprisingly, the ovine and bovine revealed less commonality compared to the other species, sharing just 16% of mutual RNAs. All species together shared 122 mutual RNAs (see Table 3-15) involved in 19 biological processes (see Figure 3-17 and Table 3-16). The main biological processes include gamete generation, motility, histone and chromosome modifications/organisation, embryo/organ development, protein phosphorylation and catabolic processes (see Figure 3-17 and Table 3-16).

Percentages/ Species	human	bovine	ovine	porcine
human	100.00	26.10	28.20	21.70
bovine	26.10	100.00	16.00	34.60
ovine	28.20	16.00	100.00	34.80
porcine	21.70	34.60	34.80	100.00

Table 3-14: Interspecies comparison. Porcine showed most transcripts in common with ovine, followed by bovine. Human spermatozoa was equal in having between 28% - 21% in common compared to the other species. Ovine and bovine had less transcripts in common, with 16%.

Symbol	Gene Name	Symbol	Gene Name
CEP112	Centrosomal protein of 112 kDa;CEP112;ortholog	*BAZ2B*	Bromodomain adjacent to zinc finger domain protein 2B;BAZ2B;ortholog
HIP1	Huntingtin-interacting protein 1;HIP1;ortholog	*NSD1*	Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific;NSD1;ortholog
CLIP1	CAP-Gly domain-containing linker protein 1;CLIP1;ortholog	*KDM1A*	Lysine-specific histone demethylase 1A;KDM1A;ortholog
MROH7	Maestro heat-like repeat-containing protein family member 7;MROH7;ortholog	*AFF4*	AF4/FMR2 family member 4;AFF4;ortholog
ODF2	Outer dense fiber protein 2;ODF2;ortholog	*USP2*	Ubiquitin carboxyl-terminal hydrolase 2;USP2;ortholog
GIGYF2	PERQ amino acid-rich with GYF domain-containing protein 2;GIGYF2;ortholog	*CEP128*	Centrosomal protein of 128 kDa;CEP128;ortholog
TTC7A	Tetratricopeptide repeat protein 7A;TTC7A;ortholog	*DENND1A*	DENN domain-containing protein 1A;DENND1A;ortholog
MED13	Mediator of RNA polymerase II transcription subunit 13;MED13;ortholog	*FMN1*	Formin-1;FMN1;ortholog

Symbol	Gene Name	Symbol	Gene Name
CLMN	Calmin;CLMN;ortholog	*LPIN1*	Phosphatidate phosphatase LPIN1;LPIN1;ortholog
MYO9A	Unconventional myosin-IXa;MYO9A;ortholog	*BAG1*	BAG family molecular chaperone regulator 1;BAG1;ortholog
ACE	Angiotensin-converting enzyme;ACE;ortholog	*CHD2*	Down syndrome cell adhesion molecule;DSCAM;ortholog
KIF21B	Kinesin-like protein KIF21B;KIF21B;ortholog	*VRK3*	Inactive serine/threonine-protein kinase VRK3;VRK3;ortholog
NBR1	Next to BRCA1 gene 1 protein;NBR1;ortholog	*NRD1*	Nardilysin;NRD1;ortholog
MYCBP2	E3 ubiquitin-protein ligase MYCBP2;MYCBP2;ortholog	*WHSC1*	Histone-lysine N-methyltransferase NSD2;WHSC1;ortholog
POLB	DNA polymerase beta;POLB;ortholog	*TEX2*	Testis-expressed sequence 2 protein;TEX2;ortholog
PROCA1	Protein PROCA1;PROCA1;ortholog	*PCNX*	Pecanex-like protein 1;PCNX;ortholog

Symbol	Gene Name	Symbol	Gene Name
SGSM2	Small G protein signaling modulator 2;SGSM2;ortholog	*KDM5B*	Lysine-specific demethylase 5B;KDM5B;ortholog
FAM104A	Protein FAM104A;FAM104A;ortholog	*OAZ3*	Ornithine decarboxylase antizyme 3;OAZ3;ortholog
CCHCR1	Coiled-coil alpha-helical rod protein 1;CCHCR1;ortholog	*CSPP1*	Centrosome and spindle pole-associated protein 1;CSPP1;ortholog
MAP4	Microtubule-associated protein 4;MAP4;ortholog	*TSSK6*	Testis-specific serine/threonine-protein kinase 6;TSSK6;ortholog
RIMS1	Regulating synaptic membrane exocytosis protein 1;RIMS1;ortholog	*UBXN6*	UBX domain-containing protein 6;UBXN6;ortholog
CCDC7	Coiled-coil domain-containing protein 7;CCDC7;ortholog	*PXDNL*	Peroxidasin-like protein;PXDNL;ortholog
NCOR1	Nuclear receptor corepressor 1;NCOR1;ortholog	*SMARCC1*	SWI/SNF complex subunit SMARCC1;SMARCC1;ortholog
DHX36	ATP-dependent RNA helicase DHX36;DHX36;ortholog	*FAM71D*	Protein FAM71D;FAM71D;ortholog
SETX	Probable helicase senataxin;SETX;ortholog	*EHBP1*	EH domain-binding protein 1;EHBP1;ortholog

Symbol	Gene Name	Symbol	Gene Name
PHC2	Polyhomeotic-like protein 2;PHC2;ortholog	CRISP2	Cysteine-rich secretory protein 2;CRISP2;ortholog
PDE4A	cAMP-specific 3',5'-cyclic phosphodiesterase 4A;PDE4A;ortholog	*CRIP2*	Cysteine-rich protein 2;CRIP2;ortholog
C16orf82	Protein TNT;C16orf82;ortholog	*CALCOCO2*	Calcium-binding and coiled-coil domain-containing protein 2;CALCOCO2;ortholog
MALAT1	Metastasis-associated lung adenocarcinoma transcript 1;MALAT1;ortholog	*EPN1*	Epsin-1;EPN1;ortholog
CSNK1G2	Casein kinase I isoform gamma-2;CSNK1G2;ortholog	*HDAC11*	Histone deacetylase 11;HDAC11;ortholog
PKM	Pyruvate kinase PKM;PKM;ortholog	*AZIN2*	Antizyme inhibitor 2;AZIN2;ortholog
C2CD3	C2 domain-containing protein 3;C2CD3;ortholog	*ALMS1*	Alstrom syndrome protein 1;ALMS1;ortholog
CCDC136	Coiled-coil domain-containing protein 136;CCDC136;ortholog	*RGS22*	Regulator of G-protein signaling 22;RGS22;ortholog
HSPA4L	Heat shock 70 kDa protein 4L;HSPA4L;ortholog	*BRWD1*	Bromodomain and WD repeat-containing protein 1;BRWD1;ortholog
TCP11	T-complex protein 11 homolog;TCP11;ortholog	*PRR30*	Proline-rich protein 30;PRR30;ortholog

Symbol	Gene Name	Symbol	Gene Name
CHD5	Chromodomain-helicase-DNA-binding protein 5;CHD5;ortholog	*CCNY*	Cyclin-Y;CCNY;ortholog
CABYR	Calcium-binding tyrosine phosphorylation-regulated protein;CABYR;ortholog	*TMCO5B*	Transmembrane and coiled-coil domain-containing protein 5B;TMCO5B;ortholog
CLEC16A	Protein CLEC16A;CLEC16A;ortholog	*EIF4G3*	Eukaryotic translation initiation factor 4 gamma 3;EIF4G3;ortholog
UBAP2	Ubiquitin-associated protein 2;UBAP2;ortholog	*SPATA18*	Mitochondria-eating protein;SPATA18;ortholog
C22orf46	Uncharacterized protein C22orf46;C22orf46;ortholog	*TSKS*	Testis-specific serine kinase substrate;TSKS;ortholog
GPX4	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial;GPX4;ortholog	*PRM1*	Sperm protamine P1;PRM1;ortholog
USP25	Ubiquitin carboxyl-terminal hydrolase 25;USP25;ortholog	*HIPK1*	Homeodomain-interacting protein kinase 1;HIPK1;ortholog
ZMIZ2	Zinc finger MIZ domain-containing protein 2;ZMIZ2;ortholog	*SEC14L1*	SEC14-like protein 1;SEC14L1;ortholog

Symbol	Gene Name	Symbol	Gene Name
UBC	Polyubiquitin-C;UBC;ortholog	*ST6GALNAC2*	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2;ST6GALNAC2;ortholog
STARD9	StAR-related lipid transfer protein 9;STARD9;ortholog	*RERE*	Arginine-glutamic acid dipeptide repeats protein;RERE;ortholog
CEP350	Centrosome-associated protein 350;CEP350;ortholog	*TRIP12*	E3 ubiquitin-protein ligase TRIP12;TRIP12;ortholog
RNF44	RING finger protein 44;RNF44;ortholog	*ISG20L2*	Interferon-stimulated 20 kDa exonuclease-like 2;ISG20L2;ortholog
SPAG9	C-Jun-amino-terminal kinase-interacting protein 4;SPAG9;ortholog	*HDLBP*	Vigilin;HDLBP;ortholog
DYRK1B	Dual specificity tyrosine-phosphorylation-regulated kinase 1B;DYRK1B;ortholog	*CHD5*	Tail-anchored protein insertion receptor WRB;WRB;ortholog
SRPK2	SRSF protein kinase 2;SRPK2;ortholog	*GLUL*	Glutamine synthetase;GLUL;ortholog
SPATA20	Spermatogenesis-associated protein 20;SPATA20;ortholog	*CDK5RAP2*	CDK5 regulatory subunit-associated protein 2;CDK5RAP2;ortholog
ERC1	ELKS/Rab6-interacting/CAST family member 1;ERC1;ortholog	*CHD1*	Chromodomain-helicase-DNA-binding protein 1;CHD1;ortholog

Symbol	Gene Name	Symbol	Gene Name
EIF4G1	Eukaryotic translation initiation factor 4 gamma 1;EIF4G1;ortholog	*HGS*	Hepatocyte growth factor-regulated tyrosine kinase substrate;HGS;ortholog
FAM71A	Protein FAM71A;FAM71A;ortholog	*CHD2*	Chromodomain-helicase-DNA-binding protein 2;CHD2;ortholog
CLPB	Caseinolytic peptidase B protein homolog;CLPB;ortholog	*AKAP13*	A-kinase anchor protein 13;AKAP13;ortholog
ANKRD28	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A;ANKRD28;ortholog	*KLHDC3*	Kelch domain-containing protein 3;KLHDC3;ortholog
ULK4	Serine/threonine-protein kinase ULK4;ULK4;ortholog	*SPATA6*	Spermatogenesis-associated protein 6;SPATA6;ortholog
RANGAP1	Ran GTPase-activating protein 1;RANGAP1;ortholog	*KHDRBS3*	KH domain-containing, RNA-binding, signal transduction-associated protein 3;KHDRBS3;ortholog
VPRBP	Protein VPRBP;VPRBP;ortholog	*EEF1D*	Elongation factor 1-delta;EEF1D;ortholog

Table 3-15: Common transcripts between bovine, porcine, ovine and human spermatozoa. The first 1000 transcripts (normalised) of each species were used to identify inter-species commonality and revealed 122 transcripts in common.

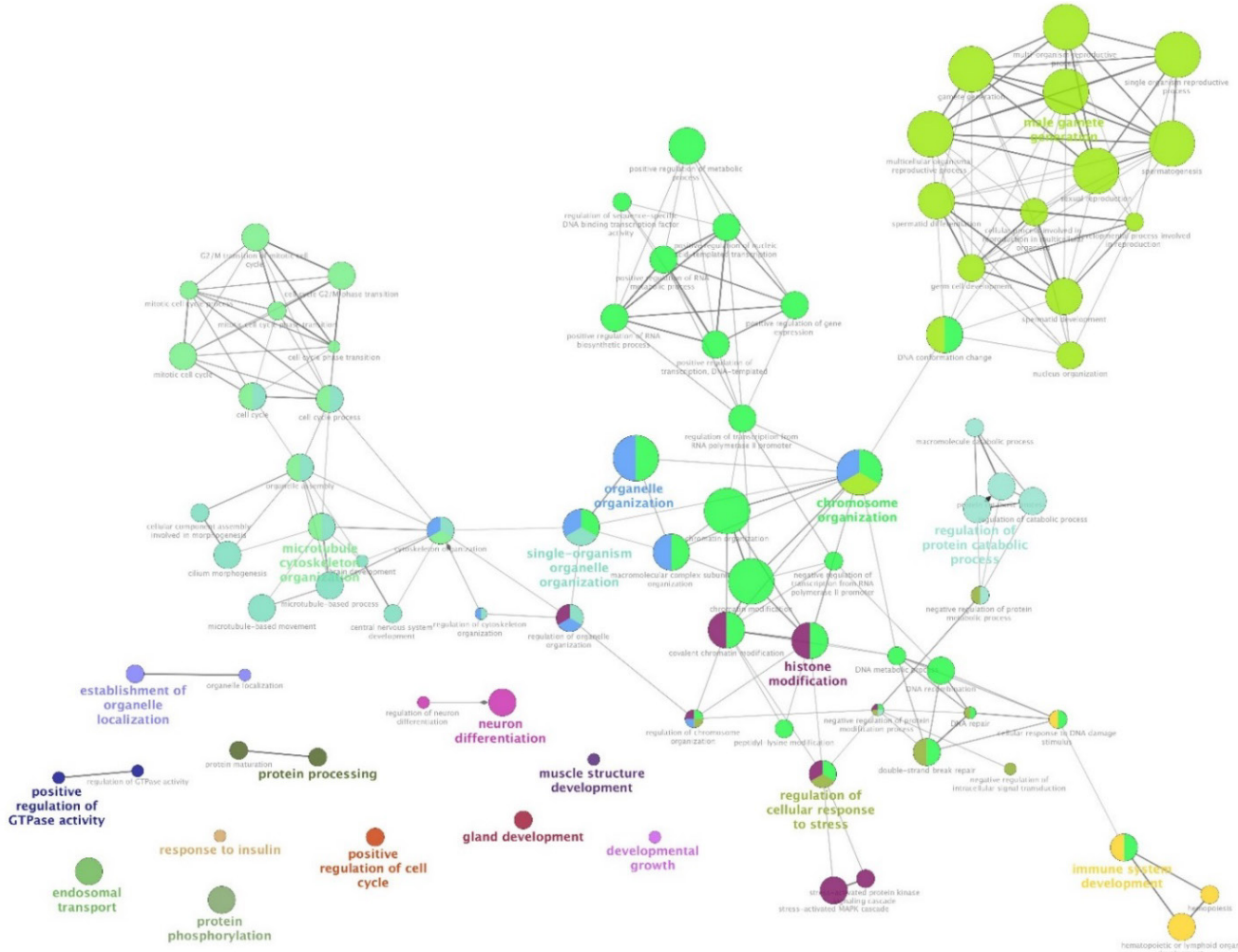


Figure 3-17: Inter-species functional annotation and pathway analysis of biological processes and clustering of the common RNAs among species. 122 mutual transcripts showed 19 clusters. Main functions contain gamete generation and motility, histone and chromosome modifications/organisation, embryo/organ development and protein phosphorylation and catabolic processes.

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
Protein phosphorylation*	GO:0006468	protein phosphorylation	6,10E-03	1,70E-02	6,10E-03	1,70E-02	1,09	21,00
Chromosome organisation **	GO:0006996	organelle organization	1,80E-05	1,50E-04	1,60E-06	1,60E-05	1,10	42,00
	GO:0009893	positive regulation of metabolic process	3,90E-04	1,80E-03	1,60E-06	1,60E-05	1,01	38,00
	GO:0043933	macromolecular complex subunit organization	1,70E-04	1,10E-03	1,60E-06	1,60E-05	1,14	31,00
	GO:1902589	single-organism organelle organization	1,10E-04	8,00E-04	1,60E-06	1,60E-05	1,15	32,00
	GO:0002520	immune system development	1,90E-02	3,60E-02	1,60E-06	1,60E-05	1,25	11,00
	GO:0006974	cellular response to DNA damage stimulus	3,10E-02	5,00E-02	1,60E-06	1,60E-05	1,20	10,00
	GO:0051276	chromosome organization	2,80E-06	3,30E-05	1,60E-06	1,60E-05	1,83	21,00
	GO:0080135	regulation of cellular response to stress	7,70E-03	1,90E-02	1,60E-06	1,60E-05	1,44	10,00
	GO:0006325	chromatin organization	9,90E-06	9,30E-05	1,60E-06	1,60E-05	2,09	16,00
	GO:0006259	DNA metabolic process	5,00E-02	6,80E-02	1,60E-06	1,60E-05	1,09	11,00
	GO:0010628	positive regulation of gene expression	3,20E-03	1,10E-02	1,60E-06	1,60E-05	1,13	21,00
	GO:0071103	DNA conformation change	7,80E-04	3,40E-03	1,60E-06	1,60E-05	2,75	7,00
	GO:0006281	DNA repair	1,30E-01	1,40E-01	1,60E-06	1,60E-05	1,13	6,00
	GO:0016568	chromatin modification	1,90E-05	1,40E-04	1,60E-06	1,60E-05	2,22	14,00
	GO:0016570	histone modification	2,60E-04	1,30E-03	1,60E-06	1,60E-05	2,28	10,00
	GO:0033044	regulation of chromosome organization	3,50E-02	5,30E-02	1,60E-06	1,60E-05	1,63	5,00
	GO:0006310	DNA recombination	2,30E-03	9,50E-03	1,60E-06	1,60E-05	2,26	7,00
	GO:0051254	positive regulation of RNA metabolic process	1,40E-02	3,00E-02	1,60E-06	1,60E-05	1,07	17,00
	GO:0006302	double-strand break repair	1,10E-02	2,70E-02	1,60E-06	1,60E-05	2,16	5,00
	GO:0016569	covalent chromatin modification	3,40E-04	1,60E-03	1,60E-06	1,60E-05	2,21	10,00
	GO:0031400	negative regulation of protein modification process	1,10E-01	1,20E-01	1,60E-06	1,60E-05	1,07	7,00
	GO:1902680	positive regulation of RNA biosynthetic process	1,30E-02	2,80E-02	1,60E-06	1,60E-05	1,09	17,00
	GO:0018205	peptidyl-lysine modification	3,90E-02	5,80E-02	1,60E-06	1,60E-05	1,42	6,00
	GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	7,90E-02	9,40E-02	1,60E-06	1,60E-05	1,29	5,00
	GO:0045893	positive regulation of transcription, DNA-templated	2,10E-02	4,00E-02	1,60E-06	1,60E-05	1,05	16,00
	GO:1903508	positive regulation of nucleic acid-templated transcription	2,10E-02	4,00E-02	1,60E-06	1,60E-05	1,05	16,00
	GO:0006357	regulation of transcription from RNA polymerase II promoter	9,80E-03	2,30E-02	1,60E-06	1,60E-05	1,05	20,00
	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	5,00E-02	6,90E-02	1,60E-06	1,60E-05	1,15	9,00
Single-organism organell organisation **	GO:0007017	microtubule-based process	4,20E-03	1,30E-02	1,70E-04	8,80E-04	1,58	10,00
	GO:0007049	cell cycle	1,50E-02	3,10E-02	1,70E-04	8,80E-04	1,01	19,00
	GO:0022402	cell cycle process	2,50E-02	4,30E-02	1,70E-04	8,80E-04	1,08	15,00
	GO:1902589	single-organism organelle organization	1,10E-04	8,00E-04	1,70E-04	8,80E-04	1,15	32,00
	GO:0007010	cytoskeleton organization	2,40E-02	4,40E-02	1,70E-04	8,80E-04	1,13	13,00
	GO:0007018	microtubule-based movement	9,00E-03	2,20E-02	1,70E-04	8,80E-04	2,31	5,00
	GO:0010927	cellular component assembly involved in morphogenesis	4,40E-02	6,40E-02	1,70E-04	8,80E-04	1,52	5,00
	GO:0070925	organelle assembly	1,30E-02	2,70E-02	1,70E-04	8,80E-04	1,41	9,00
	GO:0000226	microtubule cytoskeleton organization	5,40E-03	1,60E-02	1,70E-04	8,80E-04	1,76	8,00
	GO:0033043	regulation of organelle organization	2,60E-02	4,30E-02	1,70E-04	8,80E-04	1,11	13,00
	GO:0007417	central nervous system development	5,20E-02	7,00E-02	1,70E-04	8,80E-04	1,08	11,00
	GO:0007420	brain development	1,40E-01	1,50E-01	1,70E-04	8,80E-04	1,03	8,00
	GO:0051493	regulation of cytoskeleton organization	1,90E-01	2,00E-01	1,70E-04	8,80E-04	1,11	5,00
	GO:0060271	cilium morphogenesis	2,20E-02	4,00E-02	1,70E-04	8,80E-04	1,85	5,00
Endosomal transport *	GO:0016197	endosomal transport	4,50E-03	1,40E-02	4,50E-03	1,50E-02	2,29	6,00
Microtuble Cytoskeleton Organization "	GO:0007049	cell cycle	1,50E-02	3,10E-02	3,20E-02	6,50E-02	1,01	19,00
	GO:0022402	cell cycle process	2,50E-02	4,30E-02	3,20E-02	6,50E-02	1,08	15,00
	GO:0000278	mitotic cell cycle	1,20E-02	2,70E-02	3,20E-02	6,50E-02	1,20	13,00
	GO:0007010	cytoskeleton organization	2,40E-02	4,40E-02	3,20E-02	6,50E-02	1,13	13,00
	GO:0070925	organelle assembly	1,30E-02	2,70E-02	3,20E-02	6,50E-02	1,41	9,00
	GO:0000226	microtubule cytoskeleton organization	5,40E-03	1,60E-02	3,20E-02	6,50E-02	1,76	8,00
	GO:0044770	cell cycle phase transition	1,30E-01	1,50E-01	3,20E-02	6,50E-02	1,13	6,00
	GO:1903047	mitotic cell cycle process	6,80E-02	8,60E-02	3,20E-02	6,50E-02	1,10	10,00
	GO:0044772	mitotic cell cycle phase transition	7,70E-02	9,30E-02	3,20E-02	6,50E-02	1,19	6,00
	GO:0044839	cell cycle G2/M phase transition	7,10E-03	1,90E-02	3,20E-02	6,50E-02	2,45	5,00
	GO:0000086	G2/M transition of mitotic cell cycle	6,60E-03	1,80E-02	3,20E-02	6,50E-02	2,50	5,00
Immune system development**	GO:0002520	immune system development	1,90E-02	3,60E-02	3,60E-02	6,10E-02	1,25	11,00
	GO:0006974	cellular response to DNA damage stimulus	3,10E-02	5,00E-02	3,60E-02	6,10E-02	1,20	10,00
	GO:0048534	hematopoietic or lymphoid organ development	2,90E-02	4,90E-02	3,60E-02	6,10E-02	1,22	10,00
	GO:0030097	hemopoiesis	4,80E-02	6,70E-02	3,60E-02	6,10E-02	1,16	9,00

Group and Significance	GOID	GOTerm	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Group PValue	Group PValue Corrected with Benjamini-Hochberg	%	Nr. Genes
Regulation of cellular response to stress **	GO:0080135	regulation of cellular response to stress	7,70E-03	1,90E-02	1,10E-02	2,90E-02	1,44	10,00
	GO:0051248	negative regulation of protein metabolic process	4,70E-02	6,70E-02	1,10E-02	2,90E-02	1,03	12,00
	GO:0006281	DNA repair	1,30E-01	1,40E-01	1,10E-02	2,90E-02	1,13	6,00
	GO:0033044	regulation of chromosome organization	3,50E-02	5,30E-02	1,10E-02	2,90E-02	1,63	5,00
	GO:1902532	negative regulation of intracellular signal transduction	2,10E-01	2,10E-01	1,10E-02	2,90E-02	1,04	5,00
	GO:0006302	double-strand break repair	1,10E-02	2,70E-02	1,10E-02	2,90E-02	2,16	5,00
	GO:0031400	negative regulation of protein modification process	1,10E-01	1,20E-01	1,10E-02	2,90E-02	1,07	7,00
Response to insulin	GO:0032868	response to insulin	2,00E-01	2,10E-01	2,00E-01	2,10E-01	1,05	5,00
Organelle organization **	GO:0006996	organelle organization	1,80E-05	1,50E-04	6,20E-05	4,10E-04	1,10	42,00
	GO:0043933	macromolecular complex subunit organization	1,70E-04	1,10E-03	6,20E-05	4,10E-04	1,14	31,00
	GO:1902589	single-organism organelle organization	1,10E-04	8,00E-04	6,20E-05	4,10E-04	1,15	32,00
	GO:0007010	cytoskeleton organization	2,40E-02	4,40E-02	6,20E-05	4,10E-04	1,13	13,00
	GO:0051276	chromosome organization	2,80E-06	3,30E-05	6,20E-05	4,10E-04	1,83	21,00
	GO:0033043	regulation of organelle organization	2,60E-02	4,30E-02	6,20E-05	4,10E-04	1,11	13,00
	GO:0033044	regulation of chromosome organization	3,50E-02	5,30E-02	6,20E-05	4,10E-04	1,63	5,00
	GO:0051493	regulation of cytoskeleton organization	1,90E-01	2,00E-01	6,20E-05	4,10E-04	1,11	5,00
Histone modification **	GO:0080135	regulation of cellular response to stress	7,70E-03	1,90E-02	1,00E-03	4,30E-03	1,44	10,00
	GO:0033043	regulation of organelle organization	2,60E-02	4,30E-02	1,00E-03	4,30E-03	1,11	13,00
	GO:0031098	stress-activated protein kinase signaling cascade	3,40E-02	5,30E-02	1,00E-03	4,30E-03	1,64	5,00
	GO:0016570	histone modification	2,60E-04	1,30E-03	1,00E-03	4,30E-03	2,28	10,00
	GO:0033044	regulation of chromosome organization	3,50E-02	5,30E-02	1,00E-03	4,30E-03	1,63	5,00
	GO:0016569	covalent chromatin modification	3,40E-04	1,60E-03	1,00E-03	4,30E-03	2,21	10,00
	GO:0031400	negative regulation of protein modification process	1,10E-01	1,20E-01	1,00E-03	4,30E-03	1,07	7,00
	GO:0051403	stress-activated MAPK cascade	2,50E-02	4,30E-02	1,00E-03	4,30E-03	1,78	5,00
Regulation of protein catabolic process	GO:0009894	regulation of catabolic process	9,70E-03	2,30E-02	5,20E-02	8,10E-02	1,39	10,00
	GO:0009057	macromolecule catabolic process	3,10E-02	5,00E-02	5,20E-02	8,10E-02	1,07	13,00
	GO:0030163	protein catabolic process	3,60E-03	1,20E-02	5,20E-02	8,10E-02	1,44	12,00
	GO:0042176	regulation of protein catabolic process	2,90E-03	1,10E-02	5,20E-02	8,10E-02	1,95	8,00
	GO:0051248	negative regulation of protein metabolic process	4,70E-02	6,70E-02	5,20E-02	8,10E-02	1,03	12,00
	GO:0031400	negative regulation of protein modification process	1,10E-01	1,20E-01	5,20E-02	8,10E-02	1,07	7,00
Positive Regulation of GTPase activity	GO:0043087	regulation of GTPase activity	1,60E-01	1,70E-01	1,60E-01	1,80E-01	1,01	6,00
	GO:0043547	positive regulation of GTPase activity	1,30E-01	1,50E-01	1,60E-01	1,80E-01	1,11	6,00
Neuron differentiation *	GO:0030182	neuron differentiation	1,60E-02	3,20E-02	1,60E-02	3,70E-02	1,05	17,00
	GO:0045664	regulation of neuron differentiation	1,60E-01	1,70E-01	1,60E-02	3,70E-02	1,01	6,00
Positive regulation of cell cycle	GO:0045787	positive regulation of cell cycle	6,10E-02	8,00E-02	6,10E-02	8,80E-02	1,39	5,00
Developmental growth	GO:0048589	developmental growth	1,00E-01	1,20E-01	1,00E-01	1,20E-01	1,08	7,00
Gland development	GO:0048732	gland development	3,60E-02	5,40E-02	3,60E-02	6,50E-02	1,33	7,00
Protein processing	GO:0051604	protein maturation	7,40E-02	9,10E-02	7,40E-02	9,90E-02	1,31	5,00
	GO:0016485	protein processing	5,90E-02	7,80E-02	7,40E-02	9,90E-02	1,40	5,00
Establishment of organelle localization	GO:0051640	organelle localization	2,10E-01	2,10E-01	2,10E-01	2,10E-01	1,01	5,00
	GO:0051656	establishment of organelle localization	6,70E-02	8,60E-02	2,10E-01	2,10E-01	1,35	5,00
Muscle structure development	GO:0061061	muscle structure development	1,00E-01	1,10E-01	1,00E-01	1,20E-01	1,11	7,00
Male gamete generation **	GO:0003006	developmental process involved in reproduction	7,10E-02	8,80E-02	8,90E-10	1,70E-08	1,13	8,00
	GO:0044702	single organism reproductive process	7,80E-06	8,20E-05	8,90E-10	1,70E-08	1,65	22,00
	GO:0044703	multi-organism reproductive process	4,20E-07	5,90E-06	8,90E-10	1,70E-08	2,05	21,00
	GO:0048609	multicellular organismal reproductive process	4,00E-07	6,70E-06	8,90E-10	1,70E-08	2,26	19,00
	GO:0019953	sexual reproduction	2,90E-07	6,20E-06	8,90E-10	1,70E-08	2,30	19,00
	GO:0022412	cellular process involved in reproduction in multicellular organism	7,50E-03	1,90E-02	8,90E-10	1,70E-08	2,06	6,00
	GO:0006997	nucleus organization	1,30E-03	5,60E-03	8,90E-10	1,70E-08	3,62	5,00
	GO:0007276	gamete generation	4,90E-08	1,30E-06	8,90E-10	1,70E-08	2,74	18,00
	GO:0051276	chromosome organization	2,80E-06	3,30E-05	8,90E-10	1,70E-08	1,83	21,00
	GO:0007281	germ cell development	2,90E-03	1,00E-02	8,90E-10	1,70E-08	2,50	6,00
	GO:0048232	male gamete generation	1,50E-09	6,50E-08	8,90E-10	1,70E-08	3,42	18,00
	GO:0071103	DNA conformation change	7,80E-04	3,40E-03	8,90E-10	1,70E-08	2,75	7,00
	GO:0048515	spermatid differentiation	2,10E-04	1,20E-03	8,90E-10	1,70E-08	4,14	6,00
	GO:0007283	spermatogenesis	1,50E-09	1,20E-07	8,90E-10	1,70E-08	3,43	18,00
	GO:0007286	spermatid development	1,70E-04	1,00E-03	8,90E-10	1,70E-08	4,32	6,00

Table 3-16: ClueGO transcript analysis for inter-species commonality. Comparing the first 1000 transcripts of each species, 122 mutual transcripts were found between the four species. 19 clusters showed main processes in gamete generation and motility, histone and chromosome modifications/organisation, embryo/organ development and protein phosphorylation and catabolic processes.

3.4.2.5.2 Transcript validation

Validation of the transcripts was performed for 16 selected genes for the initial bioinformatical analysis in mixed bovine spermatozoal cDNA (see Chapter 4 for candidate selection and validation), involved in spermatogenesis, fertilisation, pregnancy outcome, placental development, DNA repair or binding or involved in RNA transport and nucleotide binding: *CRISP2*, *GTSF1L*, *ODF1*, *PRM1*, *SPATA3*, *SPEM1*, *TEX26*, *REEP6, ADAMTS6*, *FBXW5*, *KIF5C*, *MACF1, DDB1, HMGB4, KIF5C*, *SLEBP2* (see Table 4-1, PCR pictures in Chapter 4 and Appendix V). All selected candidate transcripts, except: *FBXW5, DDB1, CRISP2* could be successfully verified.

3.5 Discussion

Since the first spermaotzoal RNA was reported in 1963, scientific progression and usage of new techniques, starting with PCR to micro-arrays to NGS, has led to the idea of the spermatozoon carrying a complex repertoire of RNAs, playing a possible role in fertilisation, in the developing embryo and in the inheritance of acquired traits to progeny (Abraham and Bhargava, 1963; Miller *et al.*, 1999; Ostermeier *et al.*, 2002; Moldenhauer *et al.*, 2003; Boerke *et al.*, 2007; Garrido *et al.*, 2009; Krawetz, 2005; Sendler *et al.*, 2013; Dias and Ressler, 2014). This is one of the first studies using spermatozoal and testis RNA of four different species for comparison.

The NGS reads derived from the RNA isolated from motile spermatozoa of each species can be divided into 5 major categories: unassigned/ambiguous reads (Bt: ~42.7%; Oa: ~45.7% Ss: ~34.2%; Hs: ~14%), repeats (Bt: ~49.7%; Oa: ~43.9%; Ss: ~37.4%; Hs: ~47.2%), coding sequences (Bt: ~7%; Oa: ~0.036% incl. other RNAs; Ss: ~14.%; Hs: ~25.9%), other RNA types (Bt: ~0.2%; Oa: ~0.036% incl. coding sequences; Ss: ~3.4%; Hs: ~2.4%) and mitochondrial/ribosomal reads (Bt: ~0.4%; Oa: ~7%; Ss: ~11%; Hs: ~21.5%). Considering the fact that spermatozoa contain degraded RNA, fewer reads could be mapped compared with the corresponding testis, suggesting more fragmented RNA was only partially transcribed into cDNA library synthesis. Due to the fact that many of the smaller RNAs are lost in column treatments and library creation, they only represent a minor proportion of reads in these libraries. There are several reports, however of miRNAs in spermatozoa from human and other species (Krawetz *et al.*, 2011; McIver *et al.*, 2012; Peng *et al.*, 2012; Govindaraju *et al.*, 2012; Liu *et al.*, 2012; Fagerlind *et al.*, 2015; Curry *et al.*, 2011; Kawano *et al.*, 2012; Dias and Ressler, 2014; Gapp *et al.*, 2014; Chen *et al.*, 2016). Human was the only species with more miRNA reads in sperm than in testis (Krawetz *et al.*, 2011; Liu *et al.*, 2012). Liu *et al.* (2012) reported the requirement of miR-34c in the first cleavage of the embryo and Du *et al.* (2014) found that miRNAs in bovine may be involved in later fertilization events. This study is in line with recent studies, confirming that spermatozoal RNA is conserved across mammalian and non-mammalian species (Yang *et al.*, 2010; Das *et al.*, 2013; Fang *et al.*, 2014; Gapp *et al.*, 2014; Fischer *et al.*, 2012).

The most abundant RNA in bovine, porcine and human spermatozoa and the one showing the widest difference in expression between spermatozoa and testis was the 7SLRNA. Ovine did not show any difference in expression between spermatozoa and testis of either repetitive RNAs or any other RNA types. In total, most reads in both spermatozoa and testis aligned to the repeat mask, which contained short and long rRNA

subunits. Ullu and Tschudi (1984) mentioned that ALU repetitive elements are derived from 7SLRNAs and that the LINE and SINE frequency may be linked to the activity of a RNA dependent DNA polymerase in the testis and spermatozoa (Miller, 2000). This was also discussed in the research of Sciamanna *et al.* (2003), where the activity of an endogenous spermatozoal reverse transcriptase was found. The reverse transcriptase is not only transcribing RNA into cDNA, these cDNA is also transmitted from the founder generation to the progeny in mice. These results indicate that the transferred cDNA is fully active and additionally seen expressed as a fully protein product in the F1 generation. It was seen earlier that reproductive tracks and tissues are an abundant source of repetitive elements (Miller, 2015). Some repeats are found to fulfil a role in the early embryo e.g. LTR. LTR were seen to provide alternative transcription-activating promoters and could be involved into epigenetic processes (Peaston *et al.*, 2007). Repetitive elements are also able to "reshuffle the genome, to create new genes and alter existing genes" to mediate genomic plasticity (Brosius and Tiedge, 1996; Moran *et al.*, 1999; Ade *et al.*, 2013; Alzohairy *et al.*, 2013; Belan, 2013).

Although rRNA reads were significantly reduced using Ribo™Zero as rRNA depletion kit, rRNA showed high abundance of reads mapping to the repeat mask. rRNA depletion techniques rely on secondary structures to recognise and remove large subunits, and degraded rRNA, which are highly abundant in spermatozoa may escape this structure-dependent clearance (He *et al.*, 2010). It is interesting that while available commercial kits aimed at removing the rRNA from human RNA, rRNA was more abundant in human sources than sources from other species. The identification of different RNA types including mitochondrial RNA and the rRNA representation and variation of repeats are in agreement with previous studies (Lalancette *et al.*, 2008b; Kempisty *et al.*, 2008; Curry *et al.*, 2008; Lalancette *et al.*, 2009; Zhao *et al.*, 2009; Feugang *et al.*, 2010; Kawano *et al.*, 2012; Card *et al.*, 2013; Das *et al.*, 2013).

Gur and Breitbart (2006) reported the *de novo* translation of spermatozoal RNA into proteins through mitochondrial-type ribosomes. If mitochondrial-based translational events do occur, the translated proteins may fulfil a role in capacitation, spermatozoal-oocyte interactions and fertilisation (Zhao *et al.*, 2009; Miller and Ostermeier, 2006a; Das *et al.*, 2013). Miller and Ostermeier (2006a) discussed the role of the RNA in stabilising chromatin, which escaped repackaging into protamines and additionally may help prevent repackaging of nucleosomal DNA which is required for early developmental processes into protamines (Hammoud *et al.*, 2009). Furthermore it is possible that non-coding RNAs may play a role in regulating epigenetic changes in early embryonic development (Dadoune, 2009; Cuzin and Rassoulzadegan, 2010; Daxinger and Whitelaw, 2012; Puri *et al.*, 2010). Recent studies have shown that paternal RNA,

especially miRNAs and small tRNAs, could play a more significant role in transferring epigenetical changes to the offspring, potentially giving rise to long term health consequences in offspring (Miller and Ostermeier, 2006a; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Bohacek *et al.*, 2015; Chen *et al.*, 2016; Schagdarsurengin and Steger, 2016; Sharma *et al.*, 2016). Hence, a closer look at pre-miRNAs, miRNAs and other sncRNA types is necessary to understand their currently unknown functions and roles.

Further analysis of the raw data was performed and a MDS plot for data visualization was generated. It revealed differences in the spermatozoa and testis samples and can be seen in the MDS plot (Figure 3-10), where both RNA sources form a cluster in the first but not in the second dimension. Additionally, the raw data showed that the replicates of testis RNA expression, clusters together in ovine and porcine whereas the other replicates of both testis and spermatozoal RNAs do not, suggesting that protocol specific variations and artifacts may have been introduced through initial library preparations (Robinson *et al.*, 2010). Normalisation of the libraries was performed at a later step for statistical DE analysis to provide a more robust comparison between both spermatozoa and testis. DE analysis revealed that the majority of spermatozoal transcripts showed concordance with the corresponding testis (Bt: 14,763 transcripts; Oa: 40,986 transcripts; Ss: 10,967 transcripts; Hs: 12,009 transcripts), confirming the hypothesis of spermatozoal transcripts being historical records of spermatogenesis as reported in previous studies (Krawetz, 2005; Ostermeier *et al.*, 2002; Ostermeier *et al.*, 2005b; Moldenhauer *et al.*, 2003). However, the inter-species and spermatozoal-testis comparisons allows to distinguish between transcripts and enables grouping of those RNAs that are more abundant in spermatozoa relative to the testis. In this regard, 1244, 3282, 2187 and 1024 transcripts were more highly represented in spermatozoa of bovine, ovine, porcine and human spermatozoa, respectively, which is in agreement with former studies finding certain transcripts more highly represented in mature spermatozoa than in testis (Li *et al.*, 2002; Das *et al.*, 2010).

Higher represented spermatozoal transcripts were analysed by using ClueGO for pathway analysis showing a trend in six major pathway classes: 1. spermatozoal motility (cytoskeleton organisation, mitotic cell cycle regulation) 2. gamete generation and functions in fertilization (including protein phosphorylation and dephosphorization), 3. Spermatozoon - oocyte interaction (ion channels and transmembrane transport), 4. embryo and tissue/organ development (including hormone regulation in the embryo) and 5. RNA and DNA (histone and chromatin modifications) regulations/modifications 6. others (catabolic and metabolic processes). Protein phosphorylation and dephosphorylation processes occur during capacitation where the spermatozoa undergo

metabolic and cellular changes before and during oocyte binding (Urner and Sakkas, 2003). These changes are controlled by signalling pathways that are located in the midpiece and flagellum (less so in the spermatozoal head). Phosphorylation events are involved in capacitation, hyper activated motility, binding to the zona pellucida, acrosome reaction and spermatozoon-oocyte binding and fusion (Urner and Sakkas, 2003). Ion transport, ion channel activation and the endomembrane system play a role in enabling exchanges of the spermatozoa with its environment during capacitation, spermatozoon - oocyte recognition, acrosome reaction, chemotaxis and spermatozoa guidance (Florman *et al.*, 2007; Florman *et al.*, 2004; Teves *et al.*, 2009).

The goal of spermatozoal transcriptomic analysis was to investigate the likely existence of mutual gene expression networks and pathways that confer a fertile phenotype across species, disruption of which may lead to infertility. Having a better knowledge of the networks and pathways involved should help to improve diagnostic tools for fertility assessment (Anton and Krawetz, 2012; Garrido *et al.*, 2013; Kovac and Lamb, 2014; Malcher *et al.*, 2013). Standard examination of semen in clinics takes place using WHO guidelines that do not normally involve or consider deeper genetic causes of infertility (WHO, 2010). Some mutations have already been found in genes (*AR*; *AZF* gene families; *CFTR*, *DM-1*, *DNAH* gene family, *FGFR1*, *FSHR*, *INSL3*, *KAL-1*, *LGR8-GREAT*, *LHR*, *POLG* and dysplasia of the fibrous sheath), known to be involved in fertility pathways (Platts *et al.*, 2007; Vogt, 2004; Meschede and Horst, 1997; Lima-Souza *et al.*, 2012). However, a deeper understanding is needed of how these pathways work and the genes involved (Platts *et al.*, 2007; Vogt, 2004; Meschede and Horst, 1997; Lima-Souza *et al.*, 2012). Speculations about imprinting effects in the developing embryo associated with ART also needs a better knowledge of what roles transcripts play during fertilisation and embryogenesis (Vogt, 2004). Developmental pathways found in this study for embryo and organ/tissue (brain, muscle, neurons, etc.), especially in egg hatching indicate a potential functional role of the paternal RNA in the developing embryo.

Follow up analysis of all types of RNAs, especially non-coding and small RNAs using a NGS approach and additional RT-qPCR validation level are needed (Miller, 2014).

To find common networks and pathways playing a role that may be involved in post-fertilisation events, comparisons in RNA content among bovine, ovine, porcine and human spermatozoa were carried out. Interestingly, porcine spermatozoa were found to have more RNAs in common with ovine than ovine had in common with bovine and less between the bovine and ovine and to the human. The challenge in this data set was the well annotated human genome compared to the less well annotated genomes of the other three species. Datasets will need to be reanalysed as soon as better annotations

are available, which may in turn reveal more inter-species commonality. To overcome the interspecies issues self-built annotations were used for further analysis after re-visiting this data. This self-built annotations included the XenoRef annotations (UCSC) and ovine sequencing data of the group (with special thanks to Dr. David Iles). This approach enabled more extensive analysis of the data, without fully overcoming all limitations in the available annotations. Analysing the first 1000 most highly abundant spermatozoal RNAs of each species using the online available annotations, we found 23 transcripts (used in Chapter 4) compared to 122 mutual transcripts using the self-built annotations in this chapter.

The initial bioinformatical analysis revealed 23 mutual genes between bovine, ovine, porcine and human. 16 out of 23 transcripts were validated (see Chapter 4 and Appendix V for transcript selection and qPCR validation). All transcripts, except *FBXW5, DDB1, CRISP2* showed a signal and confirmed the NGS data. Further primers for these three transcripts need to be designed using mapped NGS sections to repeat the RT-qPCR analysis. Furthermore, the performed analysis in this chapter, revealing 122 mutual transcripts need to be analysed for the involvement in spermatogenesis, fertilisation, placental development and pregnancy outcome and embryo development. Transcripts selected in the before mentioned have to be taken further for RT-qPCR validation and the fate of the RNA in the developing embryo has to be examined to find a functional role of the paternal RNA given to the embryo.

Pathway analysis has to be considered carefully, since there is a bias for available pathways to be associated with pathologies (Khatri *et al.*, 2012). It needs to be taken into account, that the human species is the most carefully examined species associated with pathologies, however, existing human pathways contain more information about processes, genes, nodes, interactions compared to other species. To achieve a comparison in all species, the network analysis reported above had to be set to a relatively general setting to manage the smaller amounts of data available for the bovine, ovine and porcine and not to be rejected. The GO functionality analysis of both approaches showed trends about potential post-fertilisation roles and functionality in embryogenesis of spermatozoal RNA, indicating that spermatozoal RNA is not simply a random and untranslated remnant of spermatogenesis, which confirms recent research (Mizushima *et al.*, 2009; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Schagdarsurengin and Steger, 2016).

3.6 Conclusion

Although isolating spermatozoal RNA was found challenging and different methods needed to be tested to improve the RNA yield for sequencing, this study is in agreement with former studies confirming the existence of a complex repertoire of RNAs in mature spermatozoa. Spermatozoa of all species differ and they needed to be processed as similar as possible to obtain comparable data. This requirement brought the work close to the current limitations of the technology. In general, a closer look into RNA integrity and functionality in the developing embryo of mature vs. immature spermatozoa and into the small RNAs (only briefly investigated in this study) will be the next steps of this analysis. Certainly, a deeper analysis will be required to help develop biomarkers for male gamete quality and fertility analysis. The growth of NGS technologies and lower starting material inputs required to generate sequencing data will certainly aid this task (Sendler *et al.*, 2013; Lima-Souza *et al.*, 2012). Comparing spermatozoal RNA in four species, showed common functional pathways, indicating a possible post-fertilisation role in the developing embryo, and additionally strengthen other studies, which has to be further validated using qPCR.

Chapter 4: The Embryonic Fate of Spermatozoal RNA

4.1 Introduction

The long prevailing opinion that the paternal genome is the only important male contribution delivered to the oocyte by the spermatozoon has been revised in view of the presence of a wide variety of different RNA species that must also be carried by the spermatozoon to the oocyte and that could carry additional information (Krawetz, 2005; Miller and Ostermeier, 2006a; Lalancette *et al.*, 2008a). To date, the evidence for spermatozoal RNA translation has proved controversial and spermatozoa seem both transcriptionally and translationally inactive (Grunewald et al., 2005; Díez-Sánchez et al., 2003a; Gur and Breitbart, 2006). Several groups speculate, however, that once transferred to the oocyte, paternal RNA may serve some function in zygotic and embryonic development (Ostermeier *et al.*, 2004; Miller and Ostermeier, 2006a; Sendler *et al.*, 2013; Boerke *et al.*, 2007).

4.1.1 Contribution of maternal and paternal RNA to the developing embryo

Whereas the role of paternal transcripts have been controversial to date, it is well known that maternal transcripts play an important role in zygotic and embryonic development (De Loos *et al.*, 1992; Schultz and Heyner, 1992; Prather, 1992; Memili and First, 1999a; Memili and First, 1999b). In cattle, the transition from embryonic reliance on maternal transcripts to zygotic genome activation mainly occurs during the fourth or between the fourth to the fifth cell cycle transition, whereas it can be observed from the third to the fourth cell cycle in human (Memili and First, 1999b). The major transcriptional burst in cattle occurs at the eight plus-cell stage and is detected up to the 16+ cell stage (Vigneault *et al.*, 2009; Telford *et al.*, 1990). This equates to approximately 62 hours after fertilisation, giving ample time for a paternal RNA effect (such as translation into protein) before zygotic control takes over (Memili and First, 1999b). Spermatozoal RNA may be capable of affecting gene expression and epigenetic regulation in the embryo before transition from the maternal to zygotic control over gene expression occurs. In mouse, the paternal pronucleus may support transcription ~four to ~fivefold higher level than from the maternal pronucleus (Schultz, 2002). To avoid interference of maternal RNA at a later time point of embryogenesis, two major events in the embryo lead to maternal RNA clearance: maternal-zygotic transition (MZT) and zygotic genome activation (ZGA) (Schier, 2007; Walser and Lipshitz, 2011). MZT is encoded maternally and RNA binding proteins are activated to degrad maternal mRNAs, which would

leave sncRNA preserved as there is no evidence of sncRNA clearance to date (Walser and Lipshitz, 2011). After zygotic genome activation, the embryo relies on its own transcribed RNAs and it is possible that paternal RNAs escape the clearance machinery (Memili and First, 1999b; Meirelles *et al.*, 2004; De Sousa *et al.*, 1998; Walser and Lipshitz, 2011). Although many transcripts in the spermatozoon are degraded, functional roles in fertilisation, post-fertilisation and in the progeny are suggested (Rassoulzadegan *et al.*, 2006; Gilbert *et al.*, 2007; Avendaño *et al.*, 2009; Sone *et al.*, 2005; Krawetz *et al.*, 2011; Rodgers *et al.*, 2013; Gapp *et al.*, 2014; Dias and Ressler, 2014; Bohacek *et al.*, 2015). These roles include structural functions in paternal chromatin packaging, gene imprinting, spermatozoa maturation and moreover possible functions in post-fertilisation development in the oocyte, that could impact on the health of progeny during adulthood (Miller and Ostermeier, 2006a; Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Grandjean *et al.*, 2015; Chen *et al.*, 2016; Bohacek *et al.*, 2015). Ostermeier *et al.* (2004) and Kempisty *et al.* (2008) found evidence for spermatozoal RNA being delivered to the oocyte in human, porcine and bovine embryos, where the function remains unknown. These transcripts included *AKAP4*, *CLGN, CLU*, *DDX3Y*, *PLC-ζ*, *PRM1*, *PRM2*, *SPAG9* and *SRY* (Suri, 2004; Boerke *et al.*, 2007; Anderson, 2013; Nasr-Esfahani *et al.*, 2004). *PLC-ζ*, *PRM1*/*PRM2* and *microRNA-34c* are the only transcripts with a known function during fertilisation and later in the embryo to date. The first transcript triggers the Ca^{2+} oscillations during oocyte activation and suggestively has a role in embryo signalling (Swann *et al.*, 2006; Boerke *et al.*, 2007). Altered *PRM1*/*PRM2* ratios in the spermatozoon have an effect on fertilisation rates and therefore in embryo quality. The third transcript, *microRNA-34c* seems to be important in the first cleavage of mouse embryos and leads to a termination of cell division once silenced according to Liu *et al.* (2012) and was seen additionally involved in spermatogenesis (Bouhallier *et al.*, 2010; Liang *et al.*, 2012). In contrast, a *miR-34c* knockout mice strain showed normal development, which is a conflicting result compared to Liu *et al.* (2012) and the role to date remains therefore unclear (Concepcion *et al.*, 2012). As the functions of the paternally derived transcripts given to the developing embryo are unclear, this chapter focuses on the detection and stability of selected transcripts in the developing embryo.

4.2 Aims

Early bioinformatic analysis revealed 23 RNAs shared between bovine, ovine, porcine and human spermatozoa. To further characterise the potential transfer, function and stability of these transcripts, the fate of 16 of these paternal RNAs was followed in developing bovine embryos. Transcripts were selected on the basis that they were present at high levels in spermatozoal RNA and could be involved in a range of functions including fertilisation, embryo or placental development, DNA repair or binding and RNA transport.

4.3 Material and methods

4.3.1 Media, stock and culture preparations

See Appendix III for stock, media preparation and expiry. All cell culture experiments were performed using sterile or autoclaved materials.

The balance was calibrated before first usage and cleaned with 70% ethanol before and afterwards. The chemicals were fully dissolved in water and if necessary at 4°C overnight. All media and drops for *in vitro* maturation (IVM), *in vitro* fertilisation (IVF) and embryo culture were prepared in sterile and autoclaved glassware the day before usage to give the media enough time to warm up at 39°C in a non-gassed incubator (SI30H hybridisation oven/shaker, Scientific Laboratory Supplies Ltd.) or to equilibrate to 37°C and 5% CO_2 in a humidified incubator (Galaxy 170S, New Brunswick-Eppendorf, UL, Ltd., Cambridge, UK). All stocks and media were sterile filtered using 0.22 µm PS Millipore syringe-top filter. The first 5 ml were discarded, since the filter transmits embryo-toxic compounds (Harrison *et al.*, 1990). Osmolality was measured after calibration of the instrument (Model 3320 Osmometer; Advanced Instruments, Inc.) and should be between 280-290 mOsm/kg. Between each measurement the probe was cleaned with cotton sticks. To decrease osmolality sterile embryo tested cell culture water was used to dilute the solution and PBS to concentrate the solution and to increase the osmolality value. The media was adjusted up to 5% of the original amount of solution and if this was exceeded, new media was prepared. To minimise osmatic shock, all media were prepared having a similar osmolality not exceeding the range of 5 mOsm/kg. Culture dishes containing bovine germinal vesicle oocytes (GVs), mature metaphase II arrested oocytes (MII) or developing embryos were always held on a heating plate set to a temperature of 39°C if any preparations outside the incubator needed to be done or if the culture needed to be observed using a microscope. Aseptic techniques were used during all processes and media / stock preparations and all culture dishes were processed using a laminar flow hood, class 1.

4.3.2 Tissue collection and *in* vitro maturation

The investigations of this chapter were performed using bovine ovarian tissue provided by a local abattoir (JC Penny and Sons, Rawdon, Leeds, UK).

4.3.2.1 Preparation of mineral oil for IVM, IVF and embryo culture

Used sterile embryo tested mineral oil for IVM (*in vitro* maturation), IVF (*in vitro* fertilisation) and embryo culture was allowed to equilibrate to 39°C and 5% CO_2 for a minimum of 24 hours after the addition of 1 ml of the corresponding medium to the mineral oil.

4.3.2.2 Ovarian follicle collection and wash

After tissue reception from the abattoir, ovarian follicles were removed from the oviduct and placed into a pre-heated sterile ovary wash until all the tissue was processed. All ovaries were washed thrice with ovary wash and placed into pre-warmed follicle isolation media (FIM) at 39°C until further processing.

4.3.2.3 Aspiration of ovarian follicles, oocyte-cumulus complex (OCC) search and in vitro maturation

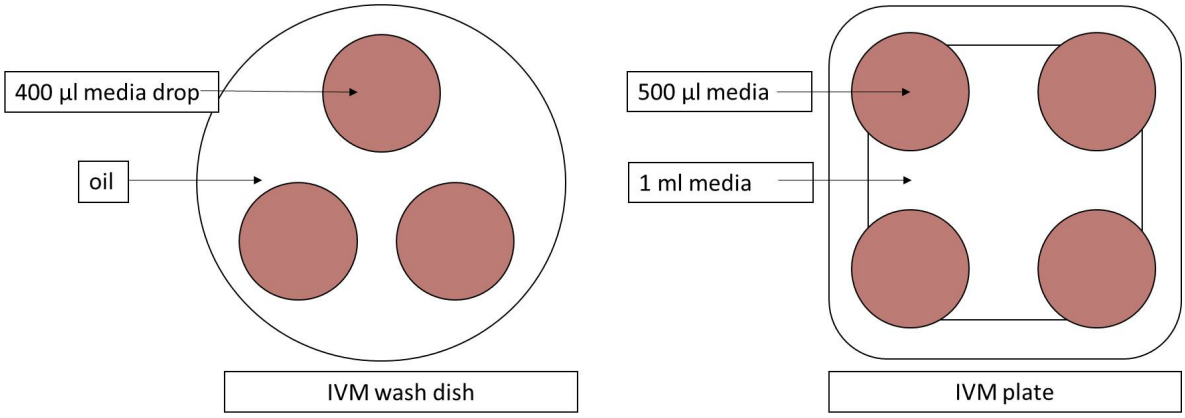


Figure 4-1: Wash and media dishes for IVM. Figure adapted from McKeegan (2015)

A wash dish and one four well nunc® dish were heated up to 39°C and equilibrated with IVM media in an 5% CO_2 incubator for a minimum of 2 hours before aspiration of the ovaries took place as follows: 400 µl drops of IVM medium were carefully placed onto a coated culture well, through reverse pipetting air drops were minimised. The bubbles were overlaid carefully with equilibrated embryo culture oil, washed with IVM media. Maturation dishes were set up by reverse pipetting 500 µl into each of the four culture wells and a 1 ml drop of media was place into the middle of the culture dish to prevent media evaporation (see Figure 4-1).

Pre-warmed holding media was pulled into a 10 ml syringe and a 19-gauge needle for aspiration was attached. Follicular fluid was aspirated from all visible follicles and after emptying the syringe the holding media containing follicles were pooled and kept at 39°C in

a universal tube. The tube was held in the oven for a minimum of 5 minutes to let the oocyte-cumulus complex (OCC) settle. A grid was drawn onto the bottom of a sterile petri-dish to ease the search for oocytes and it was filled with pre-warmed holding media. Additionally, two 60 mm petri-dishes containing pre-warmed holding media were prepared. A stereomicroscope was used to select oocytes surrounded by cumulus oophorous cells using aseptic techniques. The dishes were kept at 39°C the whole time, using a heating plate set to 39°C. Intact and none mucus-like OCCs were collected using a P100 set to 50 µl and collected in one of the 60 mm petri-dishes until the collection was completed. All visible OOCs were counted and transferred into the second prepared petri-dish containing 39°C holding media. Up to 50 OOCs were moved to the previously prepared washing dish containing IVM media and slowly released and collected into each drop until reaching the last drop for washing purposes. OOCs were transferred into the IVM plate, taking as little media as possible, and incubated at 38.5°C in 5% CO_2 in air for 18-24h.

4.3.3 Spermatozoa preparation and *in vitro* fertilisation

4.3.3.1 Fertilisation media preparation

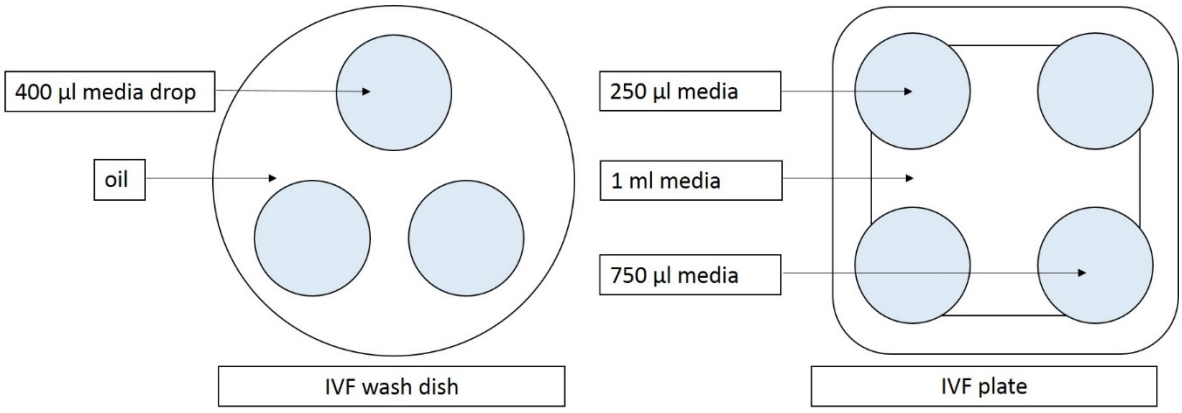


Figure 4-2: Wash and media dishes for IVF. Figure adapted from McKeegan (2015)

Fertilisation Tyrode's Albumin Lactate Pyruvate medium (Fert-Talp) was prepared a day prior to usage. IVF wash drops and fertilisation dishes were prepared at least 2 hours before usage to equilibrate the drops and the dish. 400 µl wash drops of Fert-Talp were carefully placed onto a coated IVF dish (see Figure 4-2) and covered with embryo culture tested oil. The 4-well culture dish was prepared pipetting 250 µl into three wells and 750 µl of Fert-Talp into the last well. To avoid media evaporation 1 ml of media was added into the middle of the dish and spread equally (see Figure 4-2). The reverse pipetting technique was used for preparation to avoid air bubbles.

4.3.3.2 Gradient centrifugation for spermatozoa

Frozen bovine spermatozoa were thawed and gradient centrifugation was performed to separate motile and immotile spermatozoa. See Chapter 2 for the full method of density gradient centrifugation and Appendix III for the 90% Percoll and Fert-Talp (washing and dilution media) preparation used for *in vitro* fertilisation. Motile spermatozoa of five different bulls (Genus, Cheshire, UK) were tested and spermatozoa of the bull "Classic" yielded the best *in vitro* fertilisation rates and was therefore used for all experiments. An aliquot of the motile spermatozoal fraction was counted using a Neubauer chamber to determine the spermatozoal concentration. The spermatozoal count was multiplied by the dilution factor and x 10^4 to calculate the concentration in 10^6/ml.

4.3.3.3 In vitro fertilisation

The mucoid-like matured OCCs were washed in each prepared washing drop and transferred carefully into the IVF plate transferring as little media as possible. $1x10^6$ spermatozoa (calculation: (500 µl (final volume) / spermatozoal concentration) = (volume to add in µl)) were added to each well containing the OCCs and each well was filled up to 500 µl with Fert-Talp. The OCCs were kept outside the incubator on a 39°C heating plate during the whole washing and IVF processes as short as possible. The IVF dish was returned to the incubator at 39°C and 5% CO_2 in air for 18-24h.

4.3.4 *In vitro* embryo culture

4.3.4.1 OCCs denudation

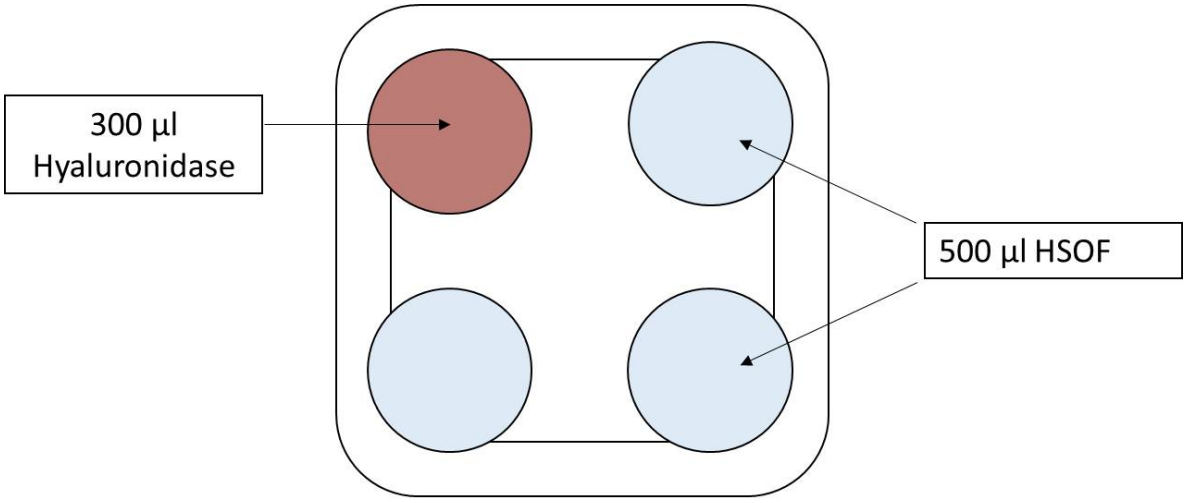


Figure 4-3: OCC denudation dish. Figure adapted from McKeegan (2015)

OCCs denudation was necessary to minimise spermatozoal RNA arising from spermatozoa attached to the cumulus cells. The denudation took place in two steps. Firstly, all OCCs were transferred into a 5 ml snap-cap centrifugation tube containing 500 µl of pre-warmed Hepes Synthetic Oviduct Fluid media (H-SOF) and vortexed at high speed for 2 minutes. The content off the tube was transferred into a small dish containing pre-heated H-SOF and all putative zygotes were visually checked for remaining cumulus cells. Secondly, all putative zygotes found surrounded by cumulus cells after vortexing were treated using pre-heated (39°C) hyaluronidase. 300 µl of hyaluronidase was transferred into one well of a four well dish (see Figure 4-3), whereas the other wells were filled with 500 µl H-SOF. Denudation took place using a flexi-pet (Cook Medical, Brisbane, Australia) starting with a 170 µm tip and after removal of the main complex, single cumulus cells were removed and the tip was changed to a 130 µm tip. The denudation process took place in a well containing hyaluronidase and the putative zygotes were pipetted up and down for several times until the cumulus cells were separated from the zygote, this process must be completed in no longer than 1 minute. If the time was exceeded, the cells were transferred back into H-SOF for regeneration and the process was repeated until no cumulus cells were visible. After denudation, all cumulus cells were washed off the embryos by washing the embryos in all three H-SOF containing wells to minimise the amount of cumulus cells transferred.

4.3.4.2 In vitro culture media preparation

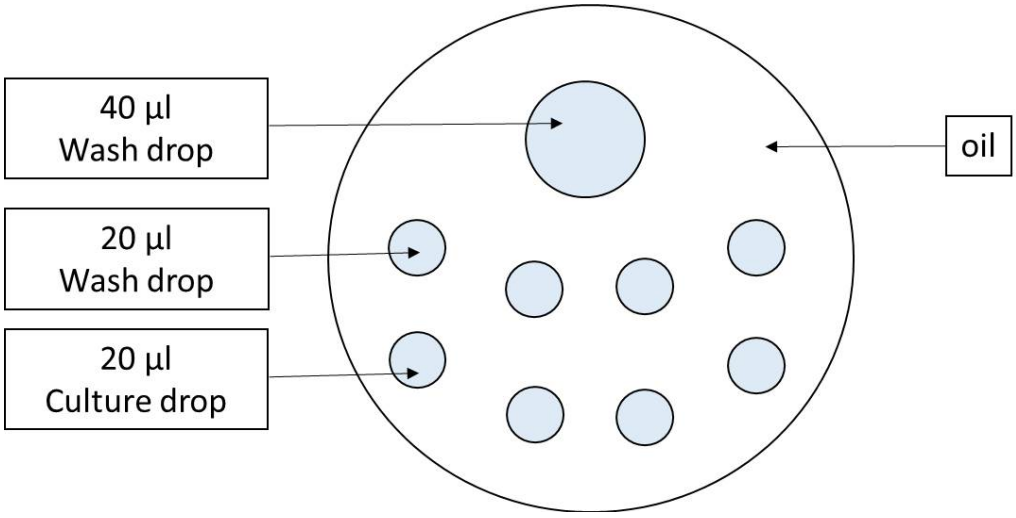


Figure 4-4: Embryo culture dish. Figure adapted from McKeegan (2015)

Embryo culture dishes were prepared the day before usage (see Figure 4-4). 40 µl wash drops and eight 20 µl culture drops were prepared and Synthetic Oviduct Fluid media supplemented with amino acids (SOFaaBSA media) was used to overlay the drops with embryo tested mineral oil, pre-heated and equilibrated for at least 2 hours. Putative embryos were carefully transferred in groups of 20 using a 170 µm glass pipette attached to a flexi-pet into both wash drops before placing the embryos together into the culture drop as close to each other as possible. Cell culture dishes containing bovine embryos were gently placed into a MINC™-1000 Benchtop Incubator (Cook®Medical Technology, Limerick, Ireland) at 39°C and air was replaced with 5% CO_2, 5% O_2, 90% N_2 through a Dreschel bottle containing 100 ml sterile water and 100 µl Ab/Am.

4.3.5 Embryo check, collection and storage of samples for molecular analysis

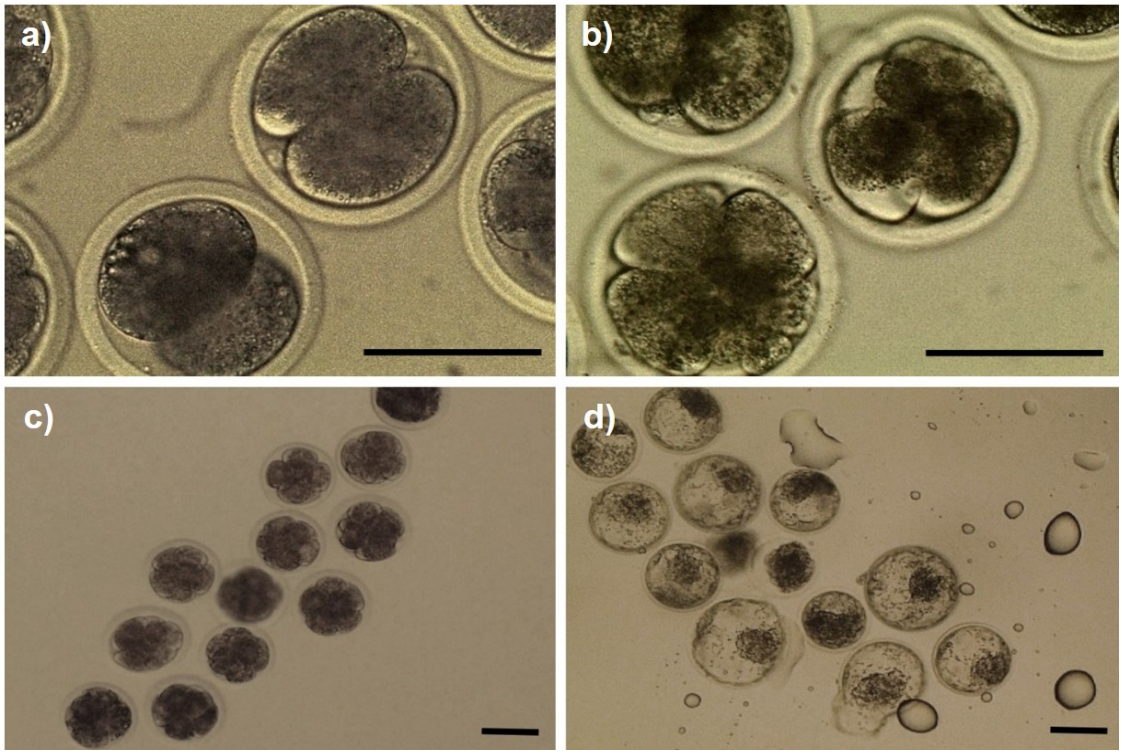


Figure 4-5: Collected bovine embryos stages. a) two plus cell embryos; b) four plus cell embryos; c) eight plus cell embryos and d) early and late blastocysts. A pool of n=20 each was collected and frozen in RLT buffer. Scale bar: 100 µm

Embryos were checked according to the required stage and collected in RLT buffer using a 170 µm glass pipette attached to a flexi-pet and an inverted microscope or a stereomicroscope. 24 h after placing the putative zygotes into the embryo culture, fertilised cells should have divided and a two-cell cleavage stage should be visible. After another 24 h a three to four-cell stage embryo should develop and after another 24 h, five to eight-cell embryos could be collected (see Figure 4-5 a-c). Morula and early blastocysts developed between the fifth and sixth day, late and hatching blastocysts between day seven and eight (see Figure 4-5 d). Embryos were cultured and collected until day eight. The cleavage rate of the embryos averaged ~80% and the blastocyst rate ~40%, depending on collection at earlier stages. Pools of 20 cells of each stage (GV, MII, two plus cell embryos, four plus cell embryos, eight plus cell embryos, early and late blastocysts) were collected in a sterile RNase/DNAse free Eppendorf tube containing 50 µl of RLT buffer and frozen at -80°C until usage.

4.3.6 RNA extraction

To increase the yield of the RNA and to use the same technique for bovine spermatozoal and embryo RNA, several isolation techniques were optimised and adapted.

A modified column based Trizol® RNA extraction approach was used, which is described further in Chapter 2. This was followed by 'on the column' DNA digestion and RNA elution using 25 µl of RNAse free water.

4.3.7 cDNA synthesis

An improved cDNA synthesis method was used, using random hexamers in combination with oligo$(dT)_{27}$ to maximise the RNA transcription. See Chapter 2.

4.3.8 Candidate transcripts, primer design and optimisation

Candidate transcripts for investigation were chosen from the initial bioinformatic analysis (Table 4-2). The top 1000 raw reads generated from bovine, ovine, porcine and human spermatozoal RNA were compared and revealed 23 mutually-expressed transcripts of which, 16 were chosen for further analysis (see Table 4-2 for all 23 mutually expressed transcripts). Selected transcripts were 1) spermatozoal specific or if not, 2) considered to play a role in fertilisation and pregnancy outcome (*CRISP2*, *GTSF1L*, *ODF1*, *PRM1*, *SPATA3*, *SPEM1*, *TEX26*, *REEP6*) or 3) considered to play a role in embryo or placental development (*ADAMTS6*, *FBXW5*, *KIF5C*, *MACF1*) or 4) involved in DNA repair (*DDB1*) or binding (*HMGB4*) or 5) involved in RNA transport and nucleotide binding (*KIF5C*, *MAFC1*, *SLEBP2*) (see Figure 4-6 and Table 4-2 for the selection process and characteristics of the chosen transcripts) (Argasinska *et al.*, 2003; Busso *et al.*, 2007; Gilbert *et al.*, 2007; Malcher *et al.*, 2013; Takemoto *et al.*, 2016; Yang *et al.*, 2012; Nasr-Esfahani *et al.*, 2004; Fu *et al.*, 2008; Zheng *et al.*, 2007; Shiyanov *et al.*, 1999; Yang *et al.*, 2013; Tsunematsu *et al.*, 2006; Dathe *et al.*, 2004; Chen *et al.*, 2006; Arangasamy *et al.*, 2011; Whitfield *et al.*, 2000). Expression of the transcripts in the different embryo stages and spermatozoa were checked in NCBI, GEO profiles (https://www.ncbi.nlm.nih. gov/geoprofiles/) and in existing literature. If no data or no expression in the GV and MII oocyte was found, the transcript was chosen and primers designed using Primer 3, version 4.0 (http://bioinfo.ut.ee/primer3-0.4.0/) (Rozen and Skaletsky, 2000). All primers (see Table 4-3) were chosen to be complementary to bovine RNA sequences from 'NCBI' and confirmed by reference to the 'Ensemble' database. The melting temperature (T_m) was set between 58°C and 61°C (optimal: 60°C) and %G-C was selected between 40% and 60%. Primers with a maximum of 3 G or 3 C in the last bases at their 3' end were accepted, in order to reduce the risk of forming a G C clamp. If

possible, repeats were avoided and primers spanning or flanking introns were preferred to monitor potential gDNA contamination. Product sizes were between 100 bp and 150 bp and all designed primers were checked for aligning to unique sequences using the bovine nucleotide 'BLAST' tool from the 'NCBI' database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE= Blast Search&LINK_LOC=blasthome) and ordered from Life Technologies Ltd. Primers were tested and optimised using a mixed bovine cDNA pool (taken from heart, lung, muscle, brain, ovary and testis). The first step was to test the primers on a standard curve with different cDNA concentrations. The second step was to confirm that the RNA was present in pelleted spermatozoa. Additionally, the primers should produce one peak in the melting curve since shoulders and more than one peak indicate primer dimers (see Figure 4-13 b).

Symbol	Gene Name	Symbol	Gene Name
CRISP2	cysteine-rich secretory protein 2	*DDB1*	damage-specific DNA binding protein 1
FAM71D	family with sequence similarity 71 member D	*FBXW5*	F-box and WD repeat domain containing 5
FNDC8	fibronectin type III domain containing 8	*GTSF1L*	gametocyte specific factor 1-like
HABP4	hyaluronan binding protein 4	*HMGB4*	high-mobility group box 4
HSF2BP	heat shock transcription factor 2 binding protein	*KIF5C*	kinesin family member 5C
MACF1	microtubule-actin crosslinking factor 1	*MYCBP2*	MYC binding protein 2, E3 ubiquitin protein ligase
NUPR1L	nuclear protein transcriptional regulator 1	*ODF1*	outer dense fiber of sperm tails 1
POLB	DNA polymerase beta	*PRM1*	Protamine 1
REEP6	receptor accessory protein 6	*RNF44*	ring finger protein 44
SPATA3	spermatogenesis associated 3	*SPEM1*	spermatid maturation 1
TEKT2	tektin 2	*TMEM8A*	transmembrane protein 8A
TSKS	testis specific serine kinase substrate		

Table 4-1: 23 mutual transcripts between bovine, ovine, porcine and human.

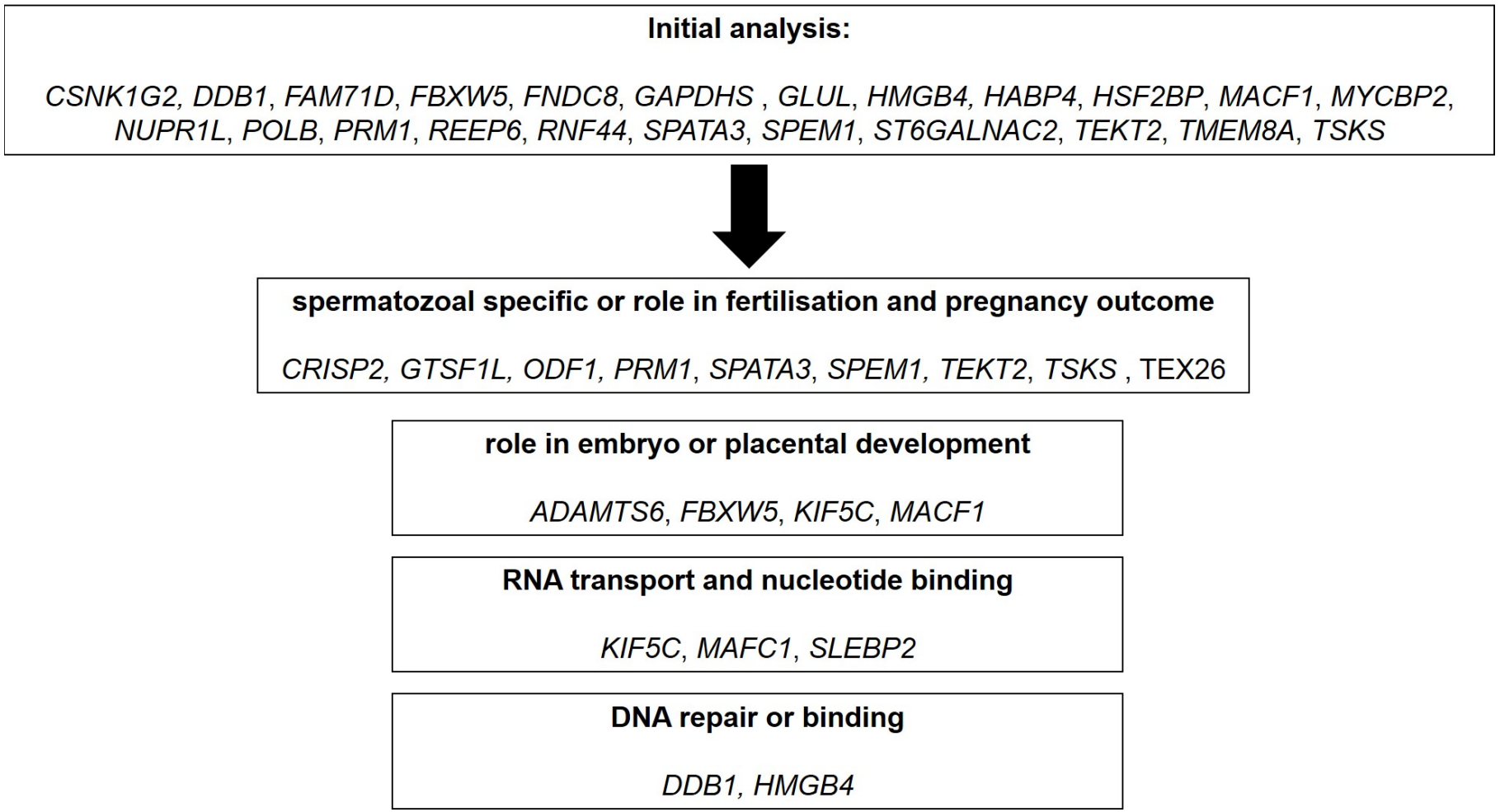


Figure 4-6: Transcript selection process. 16 transcripts were chosen to examine their involvement in either spermatogenesis, fertilisation, pregnancy outcome, placental or embryo development, RNA transport and nucleotide binding, DNA repair or binding.

Symbol	Accession Number	Name	Function/Role
GAPDH	NM_001034034.2	glyceraldehyde-3-phosphate dehydrogenase	Oxidoreductase in glycolysis and gluconeogenesis
S28	NM_001025316.2	ribosomal protein S28	Poly(A) RNA binding, ribosome biogenesis, translation
CDH1	NM_001002763.1	cadherin 1	Calcium ion binding
CRISP2	NM_001038089.1	cysteine-rich secretory protein 2	Egg-sperm fusion, regulate ryanodine receptors Ca^{2+} gating, role in male fertility[1]
GTSF1L	NM_001079601.2	gametocyte specific factor 1-like	Metal ion binding; retrotransposon suppression, function in spermatogenesis unclear[2]
ODF1	NM_174131.3	outer dense fiber of sperm tails 1	Protein binding, molecular chaperone, linkage between sperm head and tail, male fertility[3]
PRM1	NM_174156.2	protamine 1	DNA binding[4]
SPATA3	NM_001076480.2	spermatogenesis associated 3	May be involved in spermatogenesis or spermatogenesis, cell apoptosis[5]
SPEM1	NM_001079586.2	spermatid maturation 1	Cell differentiation, multicellular organism development, fertility, sperm formation and motility[6]
TEX26	NM_001098913.2	testis expressed 26	Function unknown

Symbol	Accession Number	Name	Function/Role
ADAMTS6	NM_001193016.1	ADAM metallopeptidase with thrombospondin type 1 motif 6	Cardiac and vasculature development
DDB1	NM_001080262.1	damage-specific DNA binding protein 1	DNA repair[7]
FAM71D	NM_001046625.1	family with sequence similarity 71 member D	Role unknown
HMGB4	NM_001040562.3	high mobility group box 4	DNA binding; chromatin remodelling, regulation of transcription[8]
FBXW5	NM_001192469.1	F-box and WD repeat domain containing 5	Protein binding, mitotic nuclear division, protein ubiquitination, essential in placenta development[9]
KIF5C	NM_001076127.2	kinesin family member 5C	mRNA transport, nucleotide binding. Protein trafficking, involved in formation of left-right body axis in embryos[10]
MACF1	NM_001143860.1	microtubule-actin crosslinking factor 1	Nucleic acid binding; cell cycle arrest; regulation of microtubule-based process; required for mesoderm formation[11]

Symbol	Accession Number	Name	Function/Role
REEP6	NM_001038135.2	receptor accessory protein 6	Regulation of intracellular transport, protein binding, maintenance of nervous system and musculature in embryos; regulate cell adhesion and control of invasive behaviour in trophoblasts[12]
SLBP2	NM_001281909.1	oocyte-specific histone RNA stem-loop-binding protein 2-like	RNA binding[13]

Table 4-2: Characteristics of the RNAs studied. The first three transcripts were used as housekeepers. Blue coloured transcripts are spermatozoa specific transcripts. Information about the function/role data without specific references was taken from either NCBI or UniProt.

[1] Busso *et al.* (2007), Jamsai *et al.* (2008)

[2] Takemoto *et al.* (2016)

[3] Yang *et al.* (2012)

[4] Nasr-Esfahani *et al.* (2004)

[5] Fu *et al.* (2008)

[6] Zheng *et al.* (2007)

[7] Shiyanov *et al.* (1999)

[8] Yang *et al.* (2013)

[9] Tsunematsu *et al.* (2006)

[10] Dathe *et al.* (2004)

[11] Chen *et al.* (2006)

[12] Argasinska *et al.* (2003), Branco *et al.* (2016)

[13] Whitfield *et al.* (2000)

Symbol	Fw and Rev Primer Sequence	Product Size cDNA/gDNA [bp]
GAPDH	GAAACCTGCCAAGTATGATGAG	143 / 143
	CAGCATCGAAGGTAGAAGAGTG	
S28	GCTCCATCATCCGAAACGTG	103 / 103
	CAGTCACAAGTTCAGCGCAG	
CDH1	ATGCTCCCAGATTCAACCCA	130 / 1298
	GTGTAAACAGCCTCCCATGC	
PRM1	AGATGTCGCAGACGAAGGAG	119 / 375
	AGTGCGGTGGTCTTGCTACT	
GTSF1L	GCTACCCTGCTTCAACAACTAG	128 / 128
	GGGTTCTTTCTCCGACATGATG	
HMGB4	AAGCCCGATACCAAGAAGAGAT	111 / 111
	GGCAGAAGAGTAGGAAGGATGA	
SPATA3	AGGGCAAGAGGAAGAAGTCA	104 / 104
	CTGGGAACTCGACTCTGAGC	
SLBP2	ACCTGTTTCCTCCATCCCAG	106 / 106
	TCATAGGCGGTTCAGTTGCT	
CRISP2	CTCTCCACCTGCCAGTAACA	112 / 956
	TCGTTCATTTGGACTGCTGT	
KIF5C	ACACAGCTCTAGAAGTCACAGT	111 / 111
	TCACTCCCAGCCAAATCAAC	
ODF1	TTAAGCTTTACTGTCTTCGCCC	104 / 104
	TGTTGTTCTTCTCAGTTTGGCA	
TEX26	TTGGAGCAAGAATGGCAGAG	111 / 4400
	TCCTCATCGTGGTGGCATAA	

Symbol	Fw and Rev Primer Sequence	Product Size (bp) cDNA/gDNA
ADAMTS6	CACCAGCTTTCTAGATTCAGGC	106 /Exon spanning
	ATCAGCATCATACACCTGACCT	
SPEM	TTGGCATCAACATGGTGACG	110 / 410
	CTTCCCAGGTGAGCGTAACT	
FBXW5	CGAGTGCTTCTTCATCTTCCTG	144 / 235
	AAGACCACAGAGTTGACCACAT	
REEP6	ACATACTGGGTGGTGTACGG	116 / 286
	GCCATGCAGAACAACAGGAA	
DDB1	CTATCGTTGCGCTTGAGTCC	138 / 1364
	ACAGGTTCAAGTCTTCCGCT	
MACF1	GCTTCAGACAGGAGGGATCATA	135 / 135
	AGCAAACCCATGAATGACCAAA	
FAM71D	AGCCTCTCGGAAATCACGAA	100 / 100
	ATGTGGACAGGTGTGAGGTC	

Table 4-3: Primer sequences. The first three primer pairs were used as housekeeper standards. Grey coloured primer pairs are designed according to the first method: Spermatozoal NGS reads were examined using the UCSC genome browser and the primers were assigned according to the spermatozoal reads found, in some cases they lay in one exon and intron spanning primers could not be designed. The second primer set (black) was designed considering the whole gene. All primers were diluted to a final concentration of 10 µM or 10 pmol/µl.

4.3.9 Validation of candidate transcripts and methodology

Mutually expressed candidate transcripts were selected from bovine, ovine, porcine and human spermatozoa using the RNA-Seq results of the initial NGS analysis of Chapter 3, which revealed 23 shared transcripts. Since the field of bioinformatics is evolving rapidly, the data set was re-mapped and re-analysed in Chapter 3 and the newest available assemblies were used. Out of the 23 transcripts (Table 4-1) initially checked, 16 were analysed. Seven RNAs were specific to spermatozoa with the others involved in either fertilisation, pregnancy outcome, placental or embryo development or were involved in DNA repair or RNA transport and nucleotide binding (see Figure 4-6 and Table 4-2). The genome assemblies used for this set of experiments can be seen in Table 4-4. Two different sets of primers were designed. The first set was chosen according to the bovine spermatozoal reads and adjusted to be within the mapped reads in the UCSC genome browser traces (see representative traces in Figure 4-7 and Figure 4-8). If designing primers within mapped reads was unsuccessful, a second primer set was designed irrespective of these visible reads, see representative traces in Figure 4-7. If most of the reads lay in one exon, then the primers were designed accordingly (see Figure 4-8). Therefore, a separate gDNA analysis took place for every sample, because not all primers could be designed to span or flank introns.

Species	Genome Assembly	Released (year)
H. sapiens	Hg18	March 2006
B. taurus	bosTau7	October 2011
O. aries	oviAri3	August 2012
S. scrofa	susScr3	August 2011

Table 4-4: Used genome assembly datasets for this chapter

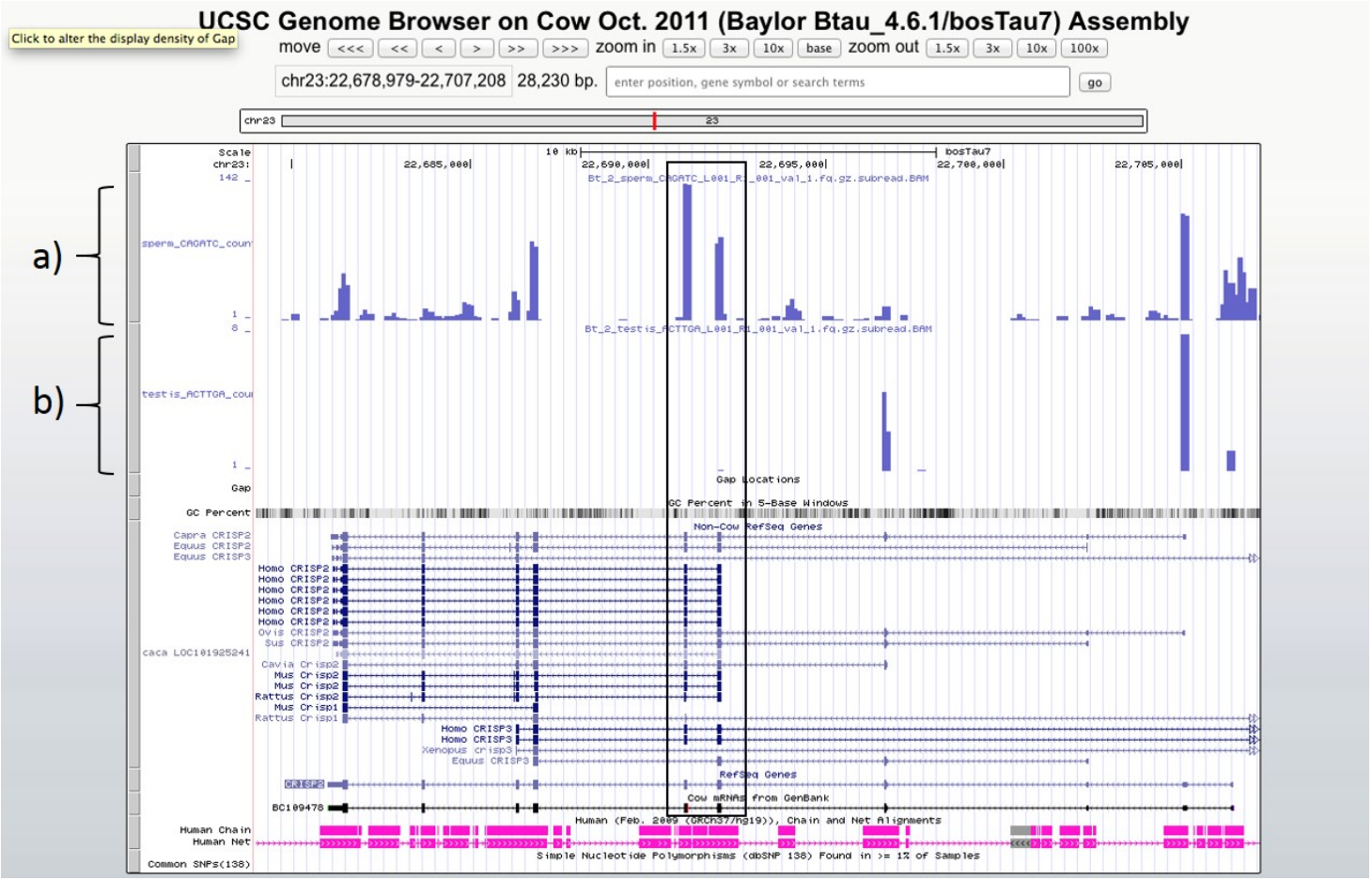


Figure 4-7: Reads for *CRISP2*. a) spermatozoal reads for *CRISPR2*. Not all mapped transcripts were full-length (see panel a), therefore the first set of primers were chosen according to the coverage of spermatozoal reads e.g.:*CRISP2* (exon 5 and 6, see black box); b) testis reads for *CRIPS2*. Testis did not show any reads in exon 5 and 6 for CRISPR2. Reading direction from right to left.

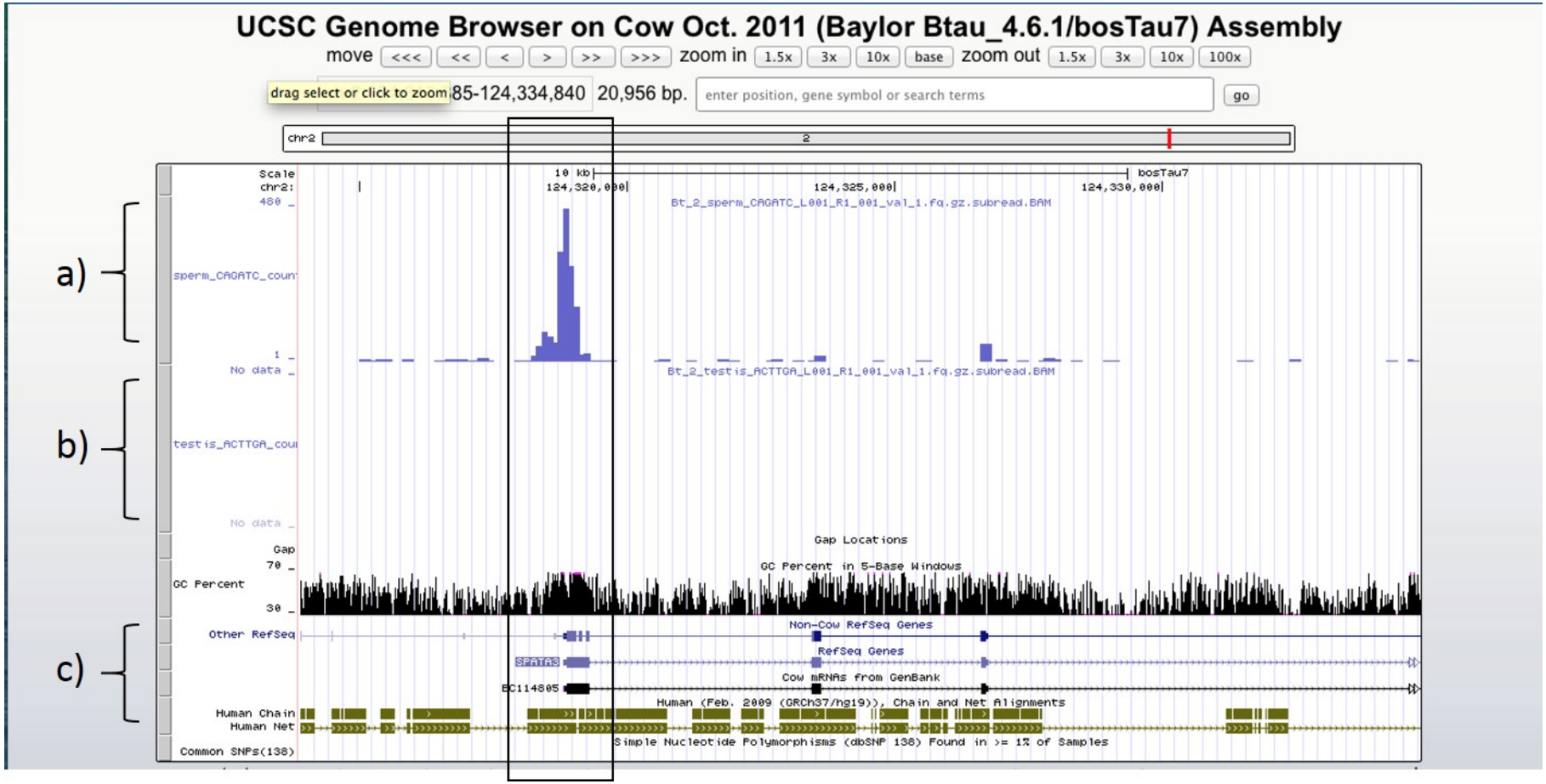


Figure 4-8: Reads for *SPATA3*. a) spermatozoal reads for *SPATA3*. Not all exons were covered with spermatozoal reads, therefore the first set of primers were chosen to cover the spermatozoal reads e.g.: *SPATA3*. Most reads were seen in the promoter region and exon 1 (see black box), therefore the primer was designed lying in one exon; b) testis reads for *SPATA3*. No reads for testis were detected. Reading direction from left to right.

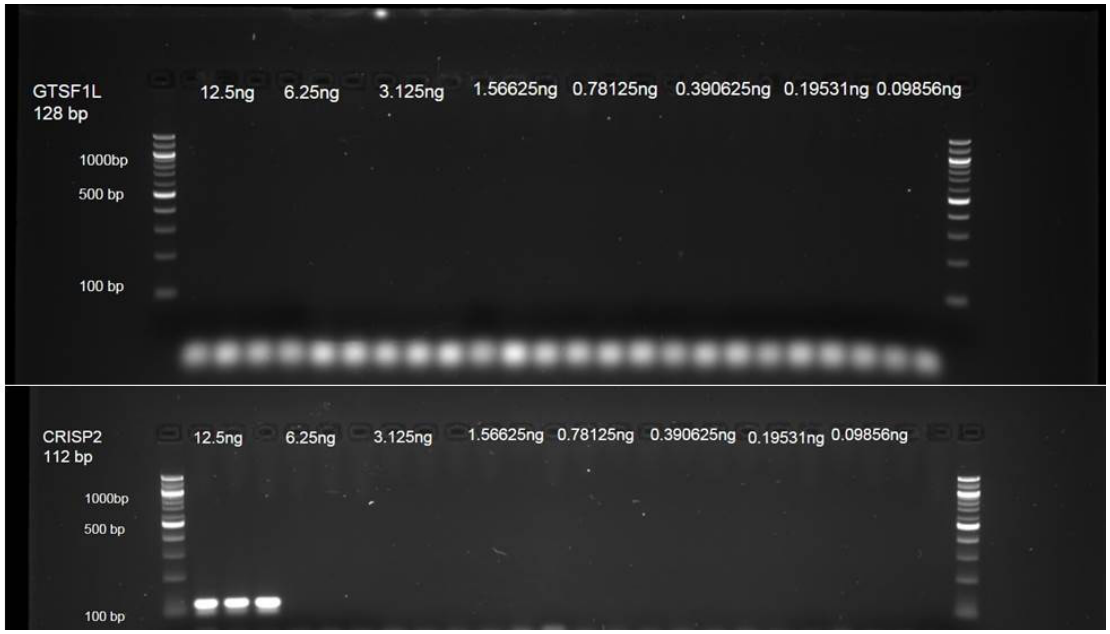


Figure 4-9: Primer set 1 mixed bovine tissue cDNA. a) *Gtsf1l* (128 bp) did not show any signal in the mixed bovine tissue and b) *CRISP2* (112 bp) showed a signal in the highest used cDNA concentration.

Two transcripts are shown here as representatives of all primers for mixed bovine cDNA, all samples were run as triplicates. All primers were tested using mixed bovine tissue cDNA and all non-working primers were re-designed. The first designed primer pair often did not show any signal in the mixed bovine tissue. However, if it was a spermatozoa specific transcript or did not show a signal in the mixed bovine tissue it displayed a strong signal in spermatozoal cDNA. Primers which did not show any signal or showed more than one band were re-designed according to the second method where the requirement for spermatozoal NGS reads was ignored.

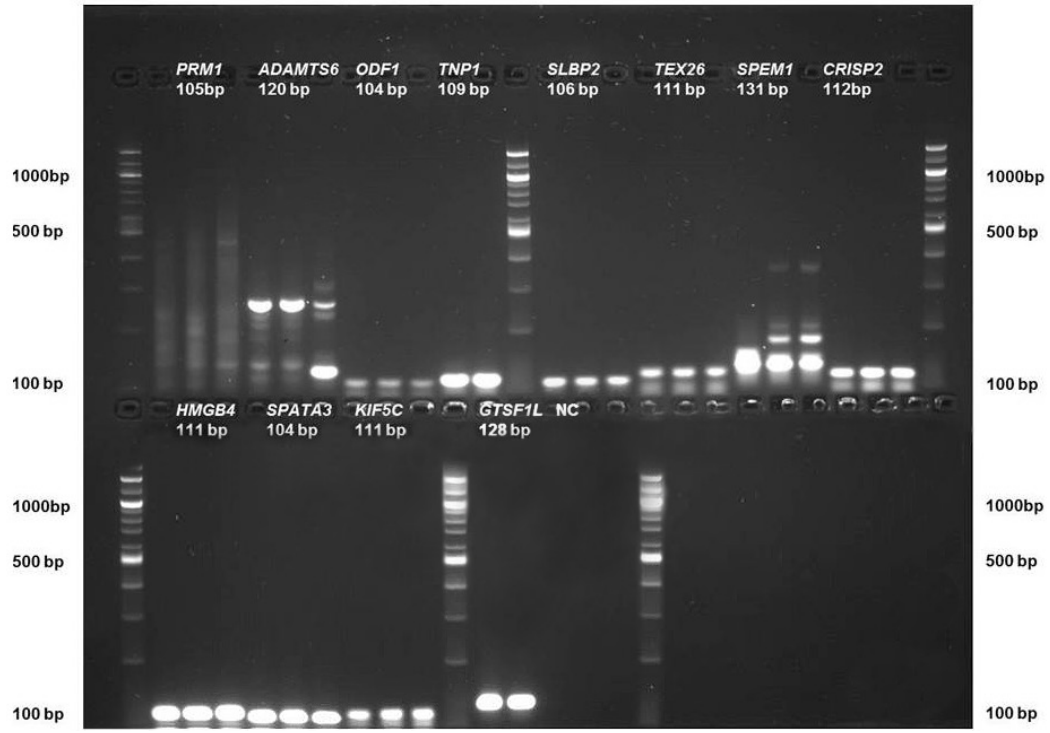


Figure 4-10: Primer set 1 spermatozoal cDNA. a) *PRM1* (105 bp), *ADAMTS6* (120 bp), *ODF1* (104 bp), *TNP1* (109 bp), *SLBP2* (106 bp), *TEX26* (111 bp), *CRISP2* (112 bp) and b) *HMGB4* (111 bp), *SPATA3* (104 bp), *KIFC5* (111 bp) and *GTSF1L* (128 bp). NC= negative water control.

All non-working primers were re-designed. All primers were tested using mixed bovine tissue cDNA. The first designed primer pair often did not show any signal in the mixed tissue. However, if it was a spermatozoa specific transcript or did not show a signal in the mixed bovine tissue it displayed a strong signal in spermatozoal cDNA. Primers which did not show any signal or showed more than one band were re-designed according to the second method where spermatozoal NGS reads were ignored.

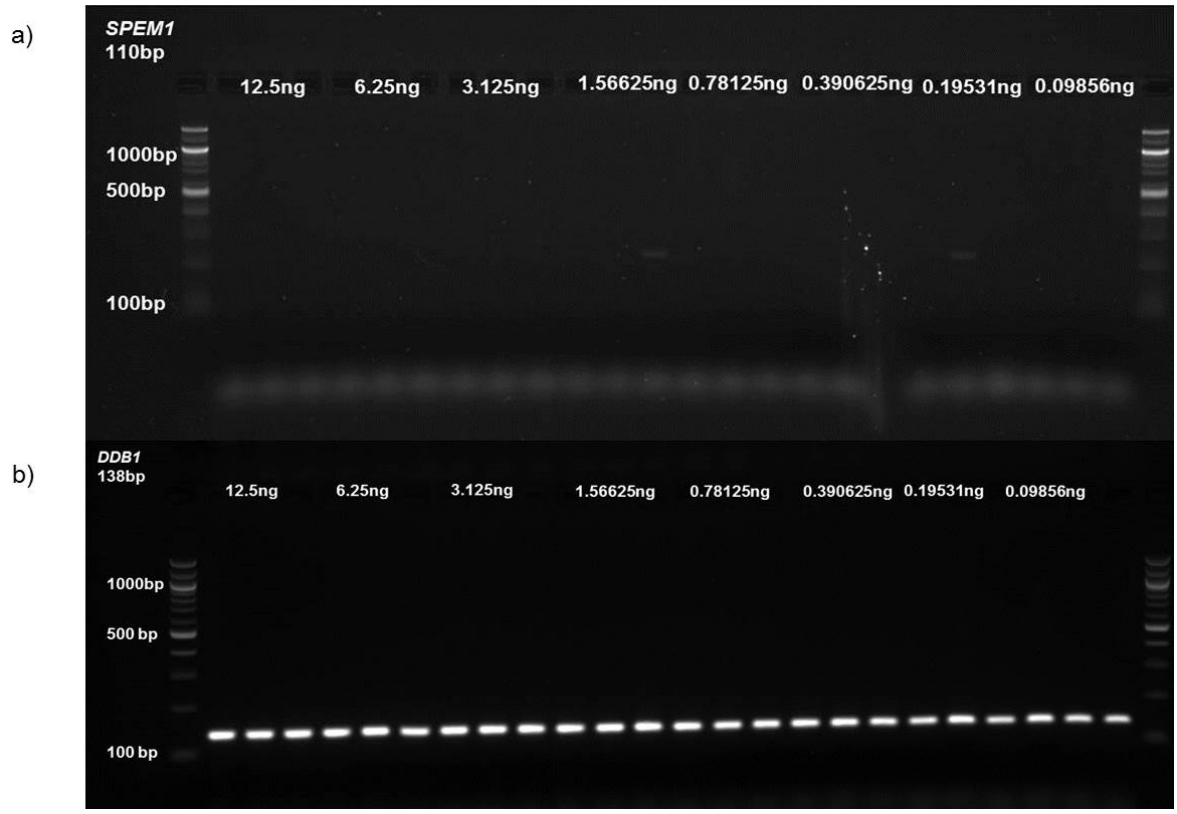


Figure 4-11: Primer set 2 mixed bovine tissue cDNA. a) *SPEM1* (110 bp) and b) spermatozoal cDNA for *DDB1* (138 bp).

Two representatives of the second primer set were selected and shown in this figure for bovine mixed RNA. *SPEM1* is a spermatozoa specific transcript and did not show any signal in the mixed bovine cDNA whereas *DDB1* showed a signal through all mixed cDNA concentrations; *DDB1* as a spermatozoal non-specific transcripts and did not show signals in spermatozoal cDNA. All transcripts showing only one signal either in the mixed bovine cDNA or in the used spermatozoal cDNA and additionally showed only one peak in the melting curve were used for further analysis.

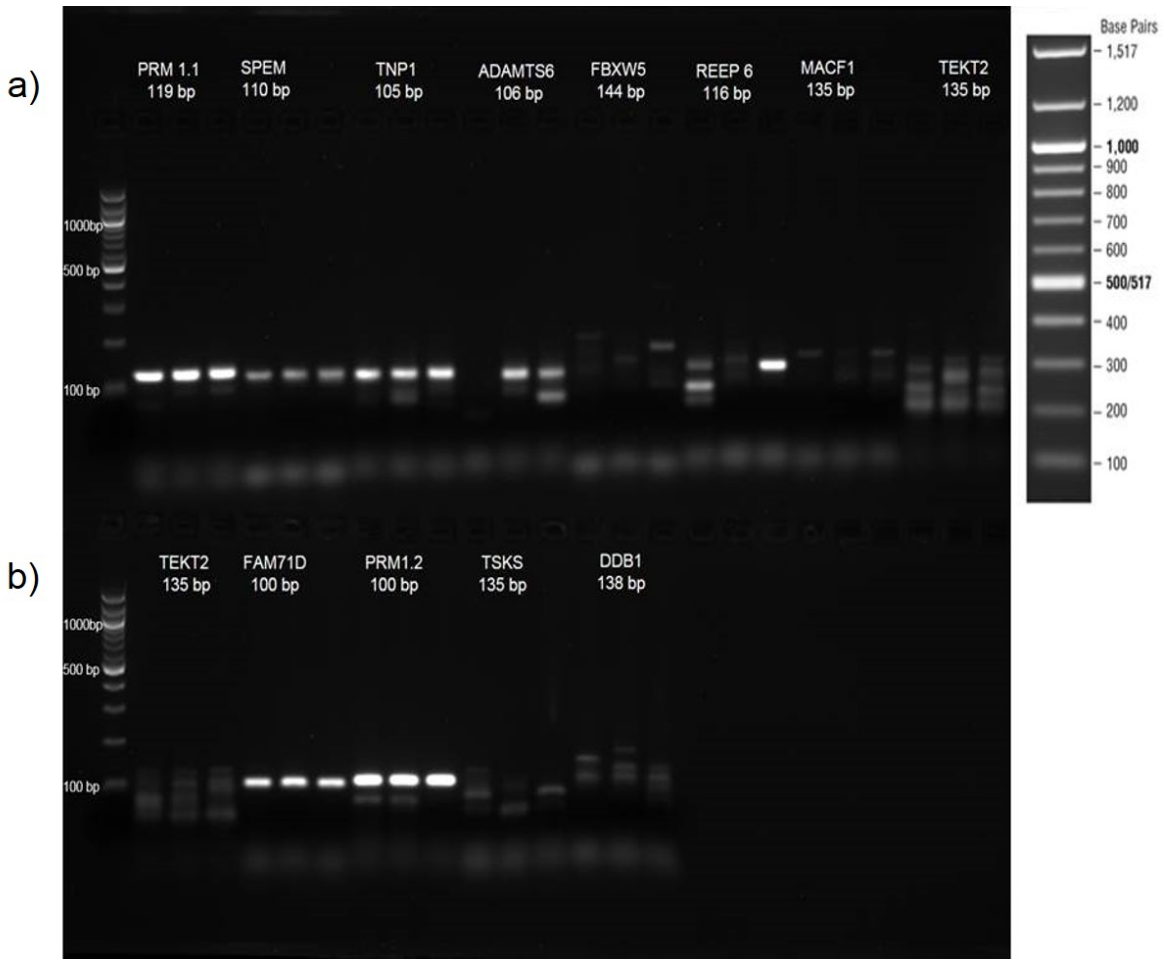


Figure 4-12: Primer set 2 spermatozoal cDNA. a) *PRM1* (119 bp), *SPEM1* (110 bp), *TNP1* (105 bp), *ADAMTS6* (106 bp), *FBXW5* (144 bp), *REEP6* (116 bp), *MACF1* (135 bp), *TEKT* (135 bp) b) TEKT2 (135 bp), *FAM71D* (100 bp), *PRM1* (100 bp), *TSKS* (135 bp), *DDB1* (138 bp).

All transcripts showing only one signal either in the mixed bovine cDNA or in the used spermatozoal cDNA and additionally showed also only one peak in the melting curve were used for further analysis. Primers not fulfilling the selection criteria were not used for further analysis and ignored, e.g. *TSKS*.

4.3.10 Gene Expression Analysis by qPCR

See Chapter 2. Due to sample quantity limitations, it was not possible to measure the concentrations of either the starting RNA templates or the converted cDNAs, but three biological and technical replicates using an equivalent number of embryos at each specific stage were run for all transcripts screened. Examples of the amplification and melting curves for housekeeper transcripts and chosen transcripts can be seen in Figure 4-13.

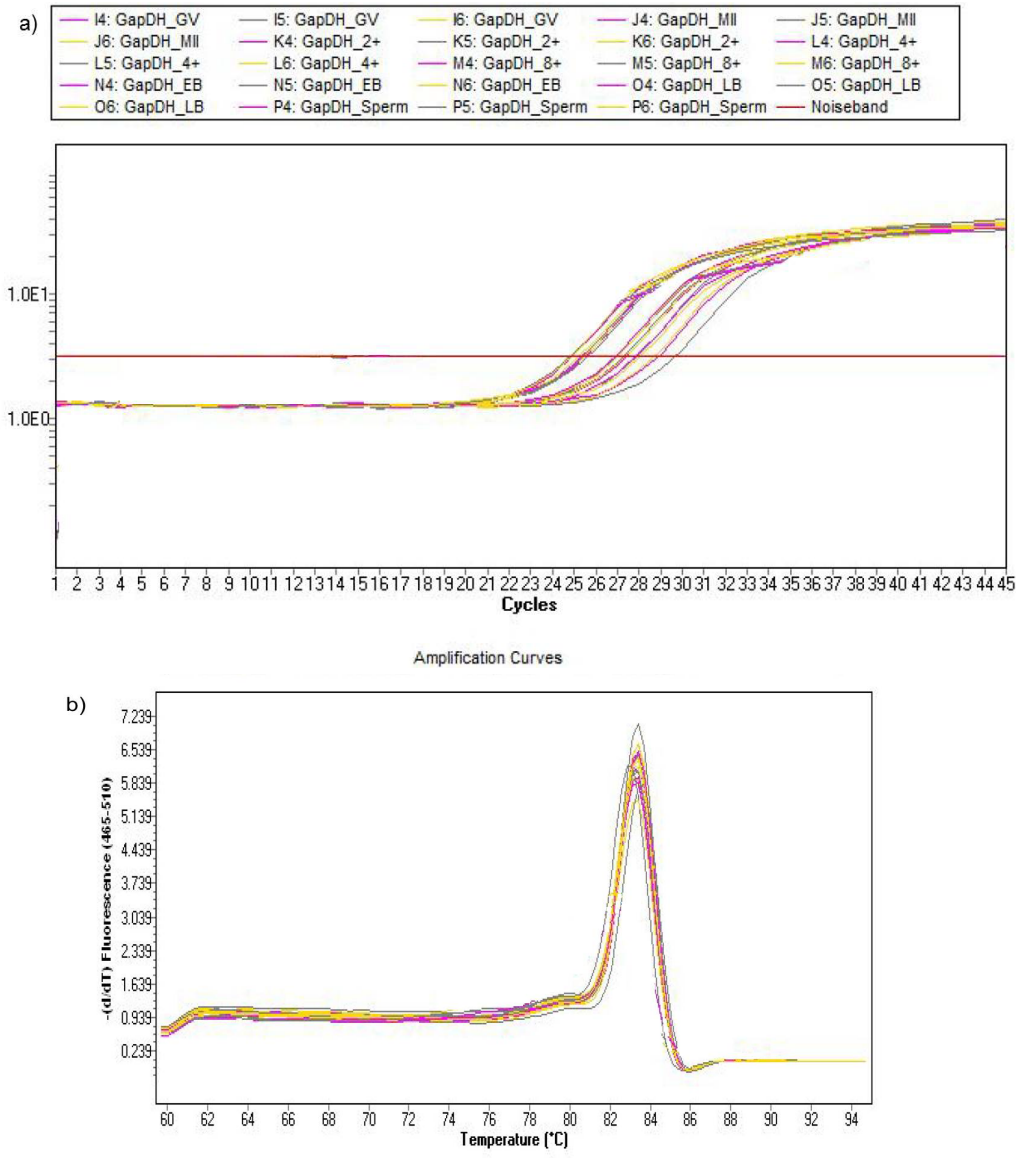


Figure 4-13: RT-qPCR for *GAPDH*. a) amplification curves using *GAPDH*; b) melting curves. Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

4.3.11 Agarose gel electrophoresis

See Chapter 2, section 2.3.7.

4.4 Results

4.4.1 Identification of spermatozoal transcripts transferred to the oocyte

For all screened transcripts three biological and technical replicates were run for every performed RT-qPCR. The existence of all transcripts in GV, MII and the different embryo cell stages were also checked using 'NCBI geoprofiles' (https://www.ncbi.nlm.nih.gov /geoprofiles/?term=) and additional existing literature was reviewed (Ostermeier *et al.*, 2004; Kocabas *et al.*, 2006; Kempisty *et al.*, 2008; Anderson, 2013). The screening results for spermatozoal-derived transcripts in different embryo stages were inconclusive. Although primers that failed the test criteria were not used, more than one PCR product was often seen and the corresponding melting curves showed more than one peak in many cases (data not shown). Because of these limitations, a RT-qPCR analysis could not be calculated and only the gels could be used as an indication of spermatozoal RNA transfer and stability. Transcripts that were undetectable in the MII oocyte stage but with weak to strong signals in 2+ and 4+ cell stage embryos and in early and late blastocysts were *REEP6* (Figure 4-14), *PRM1* (Figure 4-16), *GTSF1L* (Figure 4-17) and *HMGB4* (Appendix V; Figure V-3). Signals, however, were not apparent in all replicates. *DDB1* (Figure 4-15) was the only transcript showing a signal in all samples including both GV and MII oocytes and was therefore excluded from further analysis. Some spermatozoal specific RNAs showed a signal in the GV stage and later in the eight plus cell-stage (*SPEM1;* Appendix, Figure V-1). Experimental conditions were reasonably consistent, since control transcripts (see Appendix V) gave constant signals across all samples and all selected transcripts were detected in spermatozoal RNA with only one melting curve peak detected. Nonetheless, these results were not uniform and as far as the fate of spermatozoal transcripts delivered to the oocyte is concerned, were inconclusive. See Appendix V for all PCR results.

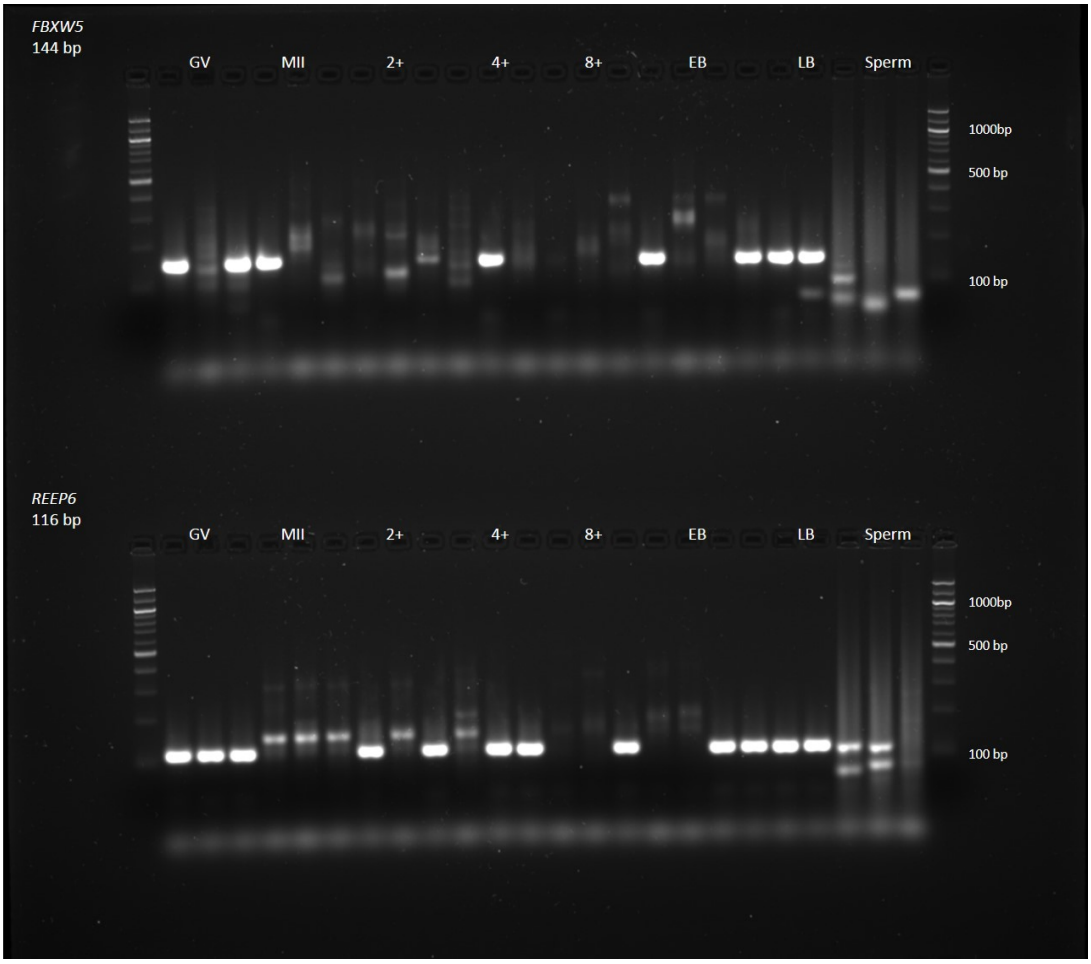


Figure 4-14: Embryonic fate of spermatozoal RNA 1. a) *FBXW5* (144 bp) b) *REEP6* (116bp) shows a signal in GV, none in MII, however, a signal in 2+ cell stage, 4+ cell stage, early and late blastocysts was seen.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

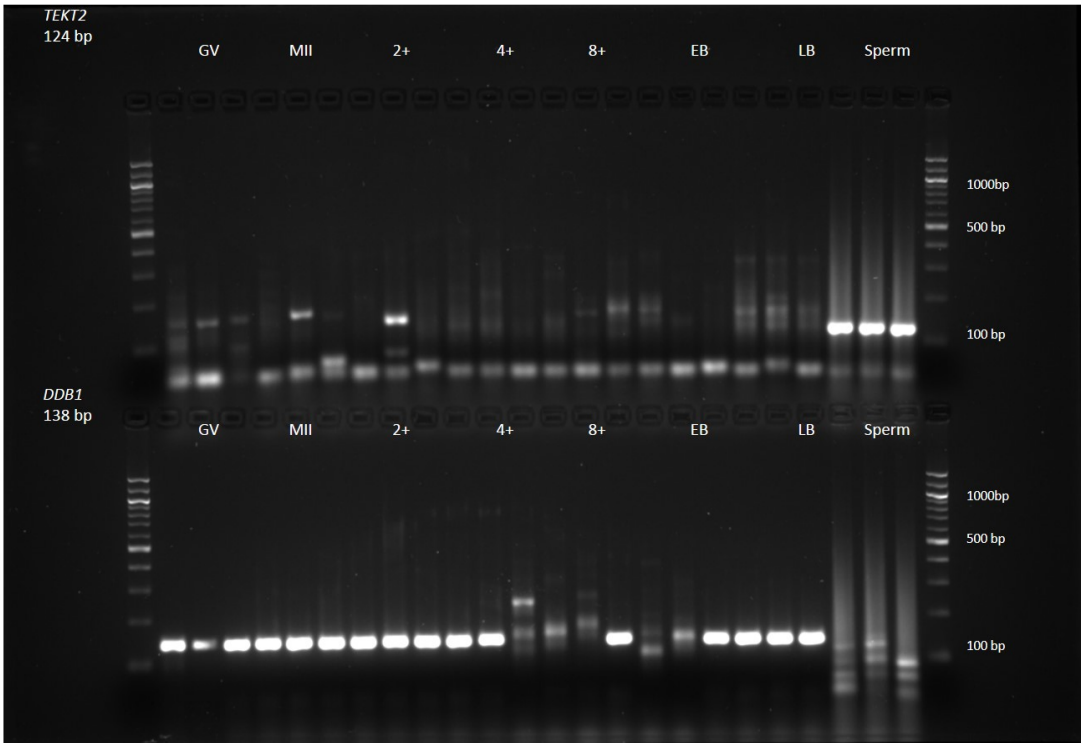


Figure 4-15: Embryonic fate of spermatozoal RNA 2. a) *TEKT2* (124 bp) b) *DDB1* (138 bp).

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

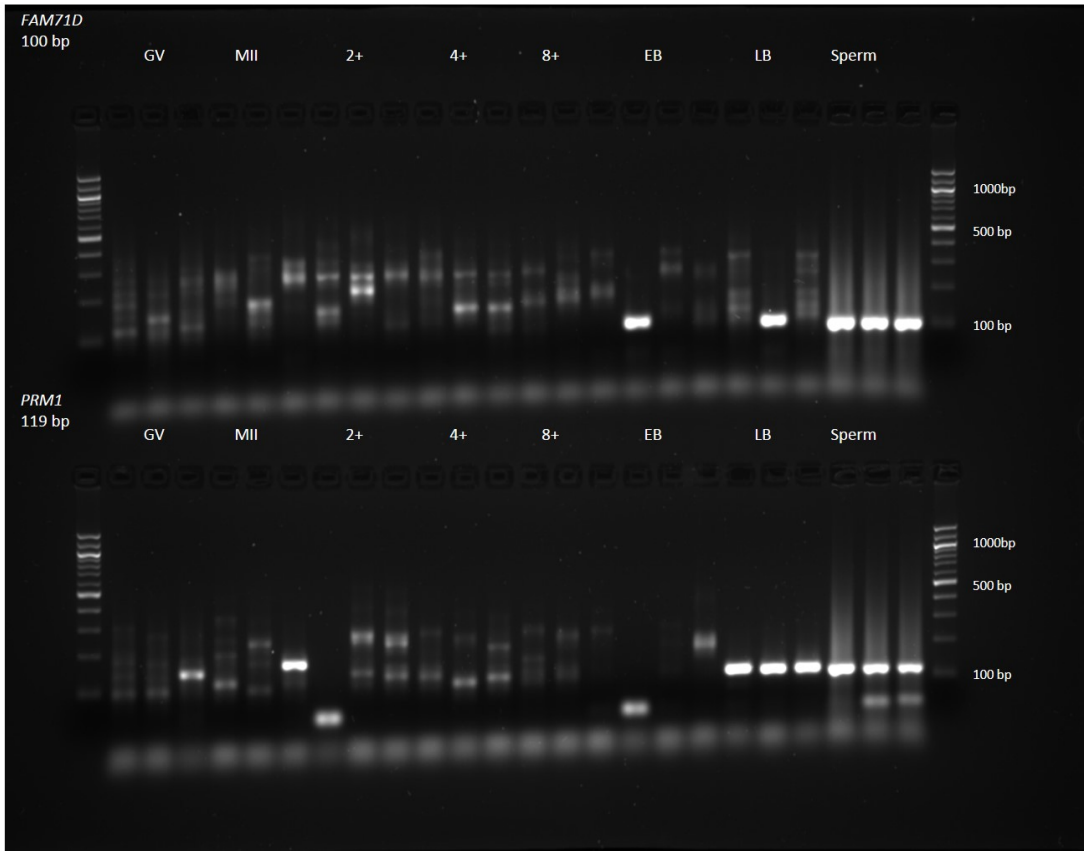


Figure 4-16: Embryonic fate of spermatozoal RNA 3. a) *FAM71D* (100 bp) and b) *PRM1* (119 bp). *PRM1* is the only transcript seen giving a signal at the right height through GV, 2+ cell stage,4+ cell stage, 8+ cell stage. Note that more than one transcript was amplified.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

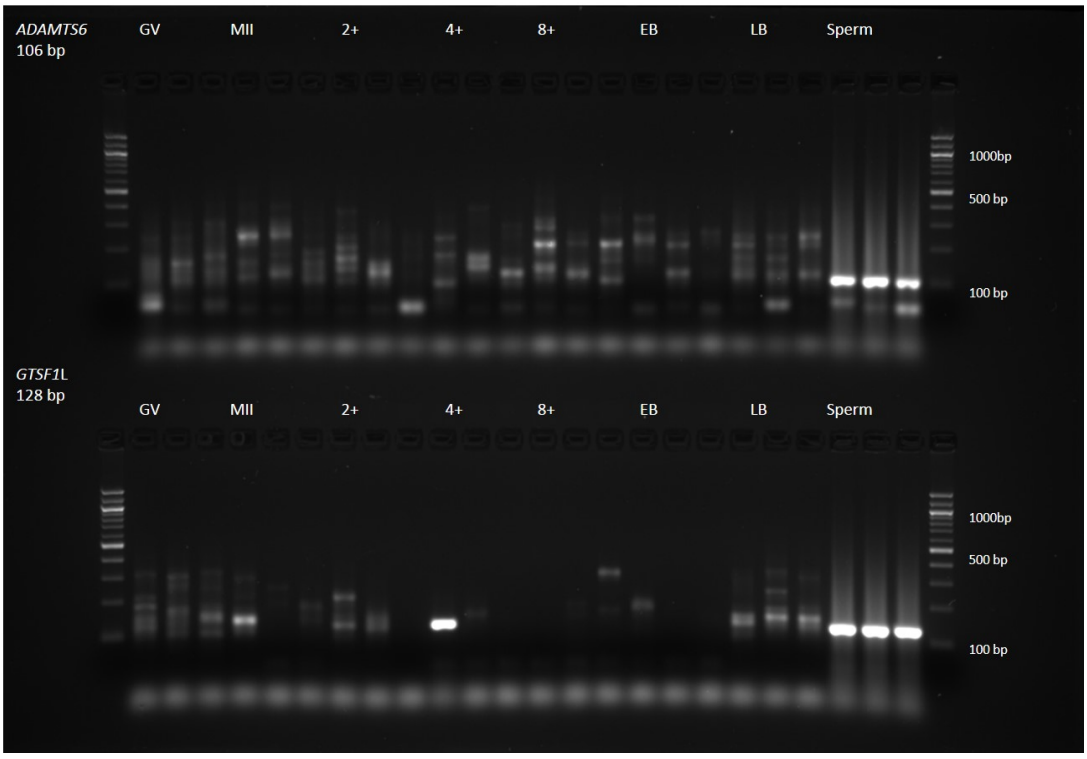


Figure 4-17: Embryonic fate of spermatozoal RNA 4. a) *ADAMTS6* (106 bp) b) *GTSF1L* (128 bp). Most chosen transcripts showed more than one amplified transcript and the spermatozoal RNA transfer can just be speculated.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

4.5 Discussion

During the initial bioinformatical analysis, 23 mutually expressed RNAs from bovine, ovine, porcine and human were identified using the assembly in Table 4-4. Transcripts selected for embryonic analysis were 1) spermatozoal specific or if not, 2) considered to play a role in fertilisation and pregnancy outcome (*CRISP2*, *GTSF1L*, *ODF1*, *PRM1*, *SPATA3*, *SPEM1*, *TEX26*, *REEP6*) or 3) considered to play a role in embryo or placental development (*ADAMTS6*, *FBXW5*, *KIF5C*, *MACF1*) or 4) involved in DNA repair (*DDB1*) or binding (*HMGB4*) or 5) involved in RNA transport and nucleotide binding (*KIF5C*, *MAFC1*, *SLEBP2*) (Argasinska *et al.*, 2003; Busso *et al.*, 2007; Gilbert *et al.*, 2007; Malcher *et al.*, 2013; Takemoto *et al.*, 2016; Yang *et al.*, 2012; Nasr-Esfahani *et al.*, 2004; Fu *et al.*, 2008; Zheng *et al.*, 2007; Shiyanov *et al.*, 1999; Yang *et al.*, 2013; Tsunematsu *et al.*, 2006; Dathe *et al.*, 2004; Chen *et al.*, 2006; Arangasamy *et al.*, 2011; Whitfield *et al.*, 2000). Since not all spermatozoal NGS reads indicated the presence of full-length transcripts, the 'UCSC' genome browser was used to visualise all reads as traces (Figure 4-7 and Figure 4-8). All reads were carefully analysed and the first primer set was designed according to spermatozoal reads found in exons (see examples in Figure 4-7 and Figure 4-8 and Table 4-3). Primers were blasted to ensure that coverage was for only one transcript and were tested in mixed bovine tissue. In most cases, no signals or weak signals were obtained. Additionally, consistent signals across all cDNA concentrations were not obtained. As some transcripts were spermatozoa specific (e.g.: *GTSF1L*, *PRM1*), all primer pairs were tested on spermatozoal cDNA. If the primers gave a signal and only one product was produced, then the primers were used for further experiments. If the primers did not work or did not display a specific product, a second set of primers ignoring the spermatozoal UCSC traces was designed and tested in both mixed cDNA and spermatozoal cDNA. As expected, in most cases, the first set of primer designs provided signals from spermatozoal cDNA, but no or weak signal in mixed bovine cDNA- whereas the second set of primers worked *vice versa* as expected (as the first primer set was designed according to the transcript reads itself). Therefore, the primer efficiency could not be calculated for all primers. As the goal was to follow the fate of spermatozoal RNA delivered to the oocyte and to assess its stability through different embryo stages, good working primers were needed in the first instance. While other studies have looked at the RNA content in spermatozoa and their likely role in male infertility, this is one of the first studies looking into the fate in embryo development of spermatozoal RNA delivered to the oocyte (Lalancette *et al.*, 2008b; Card *et al.*, 2013; Sendler *et al.*, 2013; Ostermeier *et al.*, 2004; Kempisty *et al.*, 2008; Anderson, 2013; Gapp *et al.*, 2014).

PCR products for *REEP6* (Figure 4-14), *PRM1* (Figure 4-16), *GTSF1L* (Figure 4-17) and *HMGB4* (Appendix V, Figure V-3) were not generated from MII oocytes but were seen in some of the developing embryos. Results, however, were inconsistent and the gel results were difficult to interpret. For example, some spermatozoal 'specific' primer sets generated signals in oocytes, which may be due to too many PCR cycles being used in the experiment and therefore resulted in unspecific amplification. RNA from non-fertilising spermatozoa of denuded embryos could have influenced the results despite efforts intended to avoid this possibility. Although the same bull was used and paternal changes and inconsistency may be excluded, different cow oocytes and embryos have been pooled throughout the experiment and a maternal effect cannot be excluded. However, this study did not find clear evidence for spermatozoal RNA donation to the oocyte.

One bovine spermatozoon contains 10–140 fg of total RNA (Card *et al.*, 2013). Therefore, the SYBR RT-qPCR approach used in these experiments may not have been sensitive enough to reliably detect such low levels of transcripts diluted even further in the embryonic cytoplasm. Although all primers were optimised as far as possible, some unpredictable results occurred, perhaps introduced through primer dimers, or after a certain number of cycles. 'TaqMan' probes for RT-qPCR which avoid primer dimers could be another method of choice, as it could increase the specificity of the method. Theoretically, comparisons of existing embryo data could reveal spermatozoa specific transcripts given to the oocyte. The small numbers of samples (20) pooled for each stage may have been insufficient and should be increased. Genomic contamination, however, can be excluded, since all samples were thoroughly tested beforehand.

In future, *in vitro* and *in vivo* knock-down and knock-out experiments in the developing mammal (mouse, bovine) or other vertebrate model (zebrafish, drosophila) embryo could be considered to test the function of paternally delivered RNAs. Since these experiments were performed, new annotations for human and bovine were released and all released annotations were combined with the available 'xenoref' annotation data (see Chapter 2). Additionally, the ovine annotation was re-built using the human annotation, 'xenorefseq' data combined with NGS data produced for this thesis and other sequencing projects of the group to determine and discriminate the best candidate genes (Chapter 2). A revisit of the data revealed an improvement with many more mutually expressed (from 23 to 122; see analysis in Chapter 3) transcripts being revealed. These additional RNAs could be examined and re-visited for further analysis.

4.6 Conclusion

Taken together, the fate and hence potential function of spermatozoal transcripts in the early embryo could not be ascertained with any certainty. On balance, the evidence from these experiments is that RNAs of spermatozoal origin were delivered to the oocyte and could be detected in some post-fertilisation, development stages. However, no clear expression patterns pointing towards potential functions post-fertilisation could be derived from the experimental data. These data should be revisited with new mutually expressed (or species specific) transcripts (Chapter 3) chosen and followed. Future work after identifying potential inherited paternal transcripts would involve silencing and knock-out models of available mammal and vertebrate models to clarify functionality.

Chapter 5: Assessing the Effect of Processing Techniques on the Spermatozoal Transcriptome

5.1. Introduction

15% of couples worldwide are affected by infertility (Agarwal *et al.*, 2015). The responsibility of poor quality male gametes in infertility overall lies at around 20-30% of this burden (Agarwal *et al.*, 2015). Assisted reproductive techniques (ART) try to help couples suffering from infertility to conceive a child (Said and Land, 2011). ART cycles have increased dramatically since the birth of the first IVF child in 1978 (Cohen, 1978; Ménézo *et al.*, 2000).

Cryopreservation and spermatozoal selection methods are vital and widely used methods in the breeding industry and routinely utilised in clinical management of male infertility for patients undergoing chemo-/radiotherapy or assisted reproduction procedures (Li *et al.*, 2010; Medeiros *et al.*, 2002).

Every species' spermatozoa needs to be handled differently and react differently to cryopreservation methods. This is due to physiological and biochemical differences and the variation in anatomy and physiology of the female reproductive tract in relation to spermatozoal transport in the oviduct (Holt, 2000; Lin *et al.*, 1993). Some species like the bovine need only a few motile spermatozoa to fertilise the oocyte, whereas others like the porcine need thousands to promote conception (Holt, 2000). Therefore, there is less allowable tolerance for poor quality spermatozoa during cryopreservation.

Not all crucial spermatozoal characteristics impacting fertilisation potential, however, are targeted by clinical spermatozoal preparation techniques, which are dependent mainly on visual or subjective assessment after differential density gradient centrifugation (DDGC) or swim-up. Considering that ART success rates have hardly changed in the last 20 years, it can be argued that better spermatozoal selection methods are needed, particularly for intracytoplamic sperm injection (ICSI) (Said and Land, 2011). One such alternative spermatozoal selection method makes use of hyaluronic acid (HA) to mimic the natural environment of the oocyte-cumulus oopherous complex, which is surrounded by HA (Parmegiani et al., 2010b). Only mature spermatozoa that have completed plasma remodelling, cytoplasmic removal and nuclear maturation can bind HA using specific receptors, which is thought to act as a natural barrier to oocyte fertilisation by immature spermatozoa (Parmegiani et al., 2010b; Huszar et al., 2003; Huszar et al., 2007).

5.1.1. Cryoinjury and cryopreservation

Cryopreservation is a crucial method to store spermatozoa for assisted reproductive technologies (ART). Wrongly used techniques or solutions can lead to ice crystallisation (mechanical damage) and therefore to osmotic stress of the cell where intracellular water is withdrawn from the cell, as well as an influx of ions (Holt, 2000; Mazur, 1984; Watson, 2000). Osmotic stress also occurs during the process of thawing spermatozoa, whereas the inward flux of water can cause membrane disruption. Like any cell, cryoinjuries to spermatozoa are caused mainly by thermodynamic and structural properties of the plasma membrane (Holt, 2000). Lin *et al.* (1993) reported that the spermatozoal membrane structure is built with an unusual assembly of lipids and is organised into different domains which contain a large amount of docosahexaenoic acid side chains possibly responsible for membrane fluidity and instability (Friend, 1984; Holt, 1984). To counteract this instability, the membrane contains sterols (Holt, 2000). However, the membrane responds to severe temperature changes by altering its physical phase state through lipid phase transition, but cannot modify its lipid content to adapt to the environmental conditions and is therefore not adapted to the severe drops in temperature involved in cryopreservation (Holt, 2000). Cold shock is caused by lipid phase transition and is therefore involved in cryoinjuries through membrane rearrangements involving lipids, proteins and consequently poor calcium influx during thawing at temperatures below 17°C (Holt *et al.*, 1992; Drobnis *et al.*, 1993; Bailey *et al.*, 1994; Robertson and Watson, 1986). In addition to membrane damage, reactive oxygen species (ROS) production is increased and can induce damage to the membrane and the DNA. Diminished motility through mitochondrial damage and deteriorated viability including apoptosis in frozen-thawed spermatozoa have all been reported (Aitken and Baker, 2006; O'connell *et al.*, 2002; Donnelly *et al.*, 2001a; Chohan *et al.*, 2004; Paasch *et al.*, 2004; Li *et al.*, 2010; Linschooten *et al.*, 2011). Therefore, different cryopreservation strategies have been developed aimed at preventing cryoinjuries that could affect and damage the DNA, which would compromise fertilisation, embryonal development and the health of progeny as a consequence (Branco *et al.*, 2010; Aitken and Baker, 2006). It has been shown that resveratrol, which is a stilbenoid, a type of natural phenol decreases DNA damage and ascorbate and catalase supplementation as antioxidants reduce reactive oxygen species (ROS) production during the spermatozoal thawing process (Branco *et al.*, 2010; Li *et al.*, 2010). Since ROS seem to be the parameter with the most influence on cryodamage, many cryoprotective media have been therefore supplemented with antioxidants including catalases to reduce the damage to a minimum. Such media have been shown to successfully increase the quality of frozen-thawed cells and therefore improves motility and

viability (Anger *et al.*, 2003; Gadea *et al.*, 2004; Li *et al.*, 2010; Yoshimoto *et al.*, 2008; Grossfeld *et al.*, 2008). However, no information is available about damage to spermatozoal RNA and protein on a molecular basis after freeze-thawing procedures. The likelihood of increased risks of preterm birth, low birthweight, congenital anomalies and perinatal mortality of ART offspring is statistically increased and transmitted through spermatozoal damage (Hansen *et al.*, 2002; Kalra and Molinaro, 2008). Analysing the impact of freeze thawing on spermatozoal RNA could help to dispel doubts over the preservation of RNAs in frozen stored spermatozoa and support the contention that the freeze-thaw process is not detrimental to spermatozoal RNA carriage (an important consideration in the ART context).

5.1.2. Hyaluronic acid (HA)

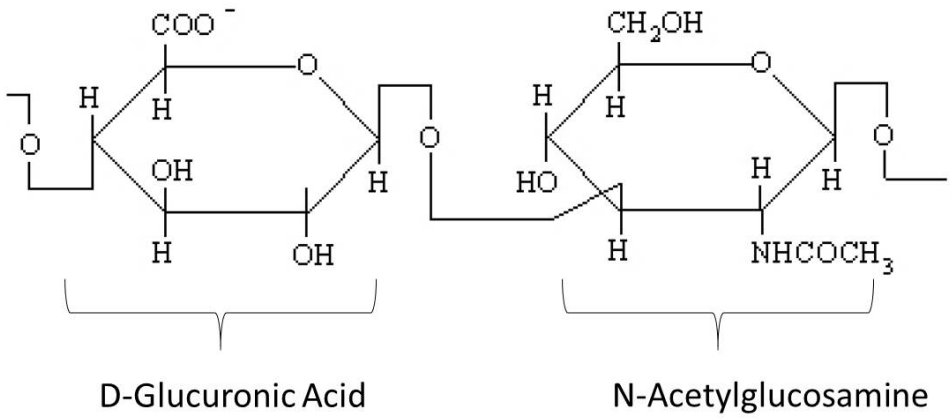


Figure 5-1: Chemical structure of hyaluronic acid. Figure adapted by (Necas *et al.*, 2008).

Hyaluronic acid (HA) (Figure 5-1) is a glycosaminoglycan that was first described in 1934 and is found mainly in the extracellular and pericellular matrix in organisms as a hydrophobic mucopolysaccharide containing two sugar molecules, D-glucuronic acid and N-acetylglucosamine (Meyer and Palmer, 1934; Necas *et al.*, 2008). Its main function is to bind water and to lubricate moveable parts of the body, for example joints. HA is also important for different biological functions including the maintenance of the elastoviscosity of connective tissue, control of tissue hydration, water homeostasis, and the supramolecular assembly of proteoglycans in the extracellular matrix. The glycosaminoglycan is also involved in various receptor-facilitated functions in cell detachment, mitosis, migration and inflammatory events, working as a signalling molecule (Toole, 2001; Turley *et al.*, 2002;

Hascall *et al.*, 2004). HA found its way into clinical usage through its ability to facilitate wound healing and tissue regeneration in surgery (Necas *et al.*, 2008). Additionally, HA is not only involved in high quality spermatozoa selection it potentially contributes to improved embryo quality (Huszar *et al.*, 2003; Cayli *et al.*, 2003; Huszar *et al.*, 2007; Camenisch *et al.*, 2000; Jakab *et al.*, 2005; Parmegiani *et al.*, 2010b; Marei *et al.*, 2012; Marei *et al.*, 2013; Marei *et al.*, 2016). HA selected spermatozoa are less likely to carry chromosomal aneuploidy. Additionally, HA seems essential in cardiac organogenesis in the embryo (Jakab *et al.*, 2005; Huszar *et al.*, 2007; Parmegiani *et al.*, 2010a; Camenisch *et al.*, 2000).

5.1.3. HA as a spermatozoal selection method for clinical usage

Oocytes are surrounded by an extra-cellular matrix or glycocalyx of glycoproteins, including hyaluronic acid which was first reported in the rabbit, mouse and bovine before being identified in humans (Sasso, 1959; Eppig, 1979; Ball *et al.*, 1982; Dandekar *et al.*, 1992; Parmegiani *et al.*, 2010b). A working model proposes that the plasma membrane of mature spermatozoa is densely packed with HA receptors and hyaluronidases which the spermatozoon uses to bind and to digest HA in order to penetrate the oocyte (Morton, 1975; Salustri *et al.*, 1989; Sasso, 1959; Eppig, 1979; Ball *et al.*, 1982; Dandekar *et al.*, 1992; Parmegiani *et al.*, 2010b). This interaction model is already being exploited as a spermatozoal selection method for ART (Parmegiani *et al.*, 2010a). Cayli *et al.* (2003) showed that immature spermatozoa with membrane deficiencies are less able to bind to HA, and may have deficits in their available HA binding sites. The ability of mature spermatozoa to bind HA for ART needed to be investigated further to demonstrate that good quality, viable, mature spermatozoa are indeed chosen by this selection method (Huszar *et al.*, 2007). Normal features of spermatozoal quality might include gametes with low or no DNA fragmentation, an equal Protamine 1 to Protamine 2 ratio, completed protamine histone exchange and low or no cytoplasmic retention. Spermatozoa with one or more of these features that are abnormal may be less able to fertilise an oocyte (Lambard *et al.*, 2004; Huszar *et al.*, 2007). Parmegiani *et al.* (2010b) reported a significant decrease in DNA fragmentation of HA (for ICSI) selected spermatozoa and an increase in embryo quality and embryonic development. Choosing spermatozoa on a more subjective visual basis may select spermatozoa with DNA damage which may therefore result in unsuccessful fertilisations or pregnancy loss (Lopes *et al.*, 1998; Cayli *et al.*, 2003; Seli and Sakkas, 2005; Borini *et al.*, 2006). Furthermore, some ART services offer a limited number of cycles per couple and therefore it is crucial to ensure that as far as possible, only the best quality

gametes are used giving a good fertilisation rate with fewer cycles (Zini *et al.*, 2008). In addition, immature spermatozoa often have decreased levels of the heat shock protein HSPA2, which plays an important role in recruiting spermatozoal surface receptors for spermatozoon-oocyte recognition and this decrease is associated with chromosomal aneuploidies (Redgrove *et al.*, 2012; Kovanci *et al.*, 2001; Parmegiani *et al.*, 2010a). HA bound spermatozoa appear to have a lower incidence of chromosomal aneuploidies, reducing the risk of chromosomal aberrations for offspring following ICSI treatments (Jakab *et al.*, 2005; Ogawa *et al.*, 2000; Morel *et al.*, 2001; Anton *et al.*, 2004). Furthermore, the crucial step of penetrating through the cumulus cells to the oocyte is thought to occur via interactions between spermatozoal HA binding site and the HA-rich oocyte-cumulus complex (Huszar *et al.*, 2007). Huszar *et al.* (2007) indicated that spermatozoa used to fertilise the oocyte selected by HA selection, retain the high levels of genetic integrity of paternal contributions that natural cycles normally provide and therefore removes spermatozoa which would have normally been eradicated through the natural way. Objective spermatozoa selection based on HA binding may offer a substantial improvement to the usual subjective assessment of spermatozoon vitality selection for ART.

5.2. Aim

This chapter aims to assess the effects of spermatozoal processing techniques used in IVF clinics or the breeding industry. Therefore, the first aim is to compare RNA profiles of frozen and fresh spermatozoa from the same donors. This analysis could reveal iatrogenic damage to the spermatozoal transcriptome caused by the cryopreservation processes and help to assess the possible risks posed to ART by the use of frozen stored spermatozoa. Secondly, the RNA profiles of unselected and HA selected semen samples shall be compared to see if molecular differences between them might help to confirm HA binding ability.

5.3. Material and methods

Materials used and methods performed are corresponded to Chapter 3 (Material and methods), except otherwise indicated. This applies especially to the fresh and frozen spermatozoal samples.

5.3.1. Collection of fresh spermatozoal donor samples

Discussed in detail in Chapter 2.

For each replicate for the cryopreservation study, five different human donor samples were pooled and split equally into two aliquots, one of which was immediately processed for freezing in liquid nitrogen and stored at -190°C. The other aliquot was immediately processed for RNA isolation directly after spermatozoa isolation on a density gradient (see Chapter 2). Frozen samples were stored for a minimum of 48 h prior to thawing and RNA isolation was performed as described previously (in Chapter 2). All RNA samples were frozen and stored at -80°C before usage.

5.3.2. Handling of HA selected and unselected samples

5.3.2.1. Collection of HA selected samples

HA selected and unselected samples were kindly provided by Dr. Gabor Huszar, The Sperm Physiology Laboratory, Department of Obstetrics, Gynaecology and Reproductive Sciences, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA.

5.3.2.1 Demography of HA selected and unselected samples

5.3.2.1.1 HA selected

Sample	Sperm concentration (million/ml)	Total sperm count ($\times 10^6$)	Semen volume (ml)	Sperm motility (%)
HA7	212.7	191.43	0.9	57
HA10	43.3	38.97	0.9	67
HA11	208	312	1.5	62
HA18	143.2	143.2	1	52
HA20	163.4	163.4	1	55
Average	154.12	169.80	1.06	58.60
STD	68.60	98.18	0.25	5.94

Table 5-1: Demography of HA selected samples

5.3.2.1.2 HA unselected

Sample	Sperm concentration (million/ml)	Total sperm count ($\times 10^6$)	Semen volume (ml)	Sperm motility (%)
S7	379.5	380.1	0.6	83
S10	81	82	1	58
S11	205.3	208.5	3.2	68
S18	139.8	141.3	1.5	47
S20	263.4	264.4	1	70
Average	213.8	215.26	1.45	65.2
STD	115.22	114.96	1.02	13.52

Table 5-2: Demography of HA unselected samples

5.3.2.2 HA selected and unselected spermatozoal sample processing

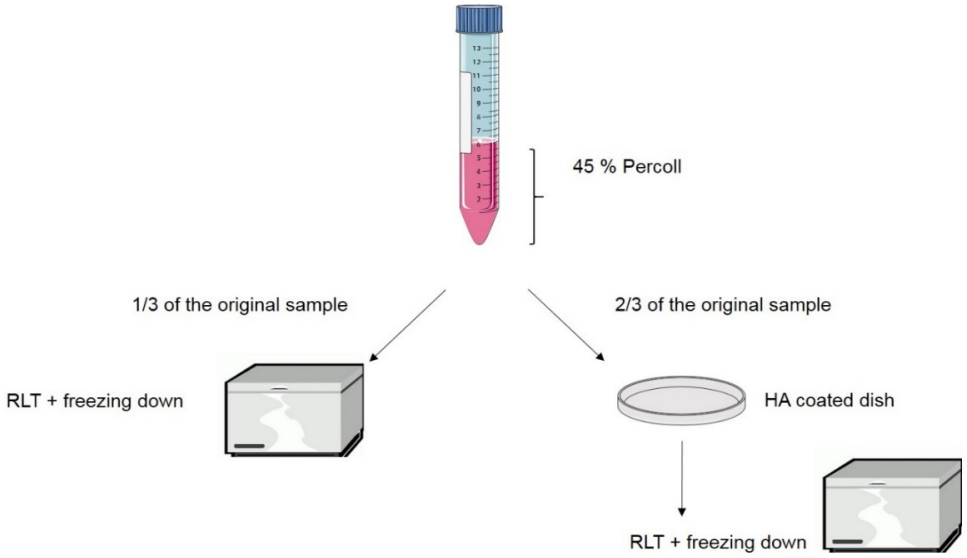


Figure 5-2: HA-selection procedure. Patient samples were pelleted and 1/3 was frozen in RLT buffer containing β-mercaptoethanol. 2/3 was applied to HA coated dishes and the unbound fraction was removed with sperm wash buffer. RLT buffer was applied to the dish to lyse and isolate RNA from the bound fraction. Both fractions were frozen after spermatozoa lysis.

In brief:

Liquefied human spermatozoa were centrifuged on a 45% monolayer using “Isolate gradient” at 600 x g for 25 minutes. One third of the pelleted semen (unselected fraction) was frozen down in 700 µl RLT buffer (lysis buffer), containing 40 µl β-mercaptoethanol. The remaining 2/3 was applied to HA coated dishes for sperm selection (HA selected fraction) for 30 minutes and the binding progress was monitored. The supernatant was carefully removed after the 30 minute binding period and the HA bound sperm were released from the dish by adding lysis buffer containing β-mercaptoethanol. After placing the samples into the ‘Disrupter Genie’ to ensure lysis, the samples were frozen at -80°C and sent to our laboratory on dry ice (Figure 5-2) (Jakab *et al.*, 2005).

5.3.3 Experimental workflow

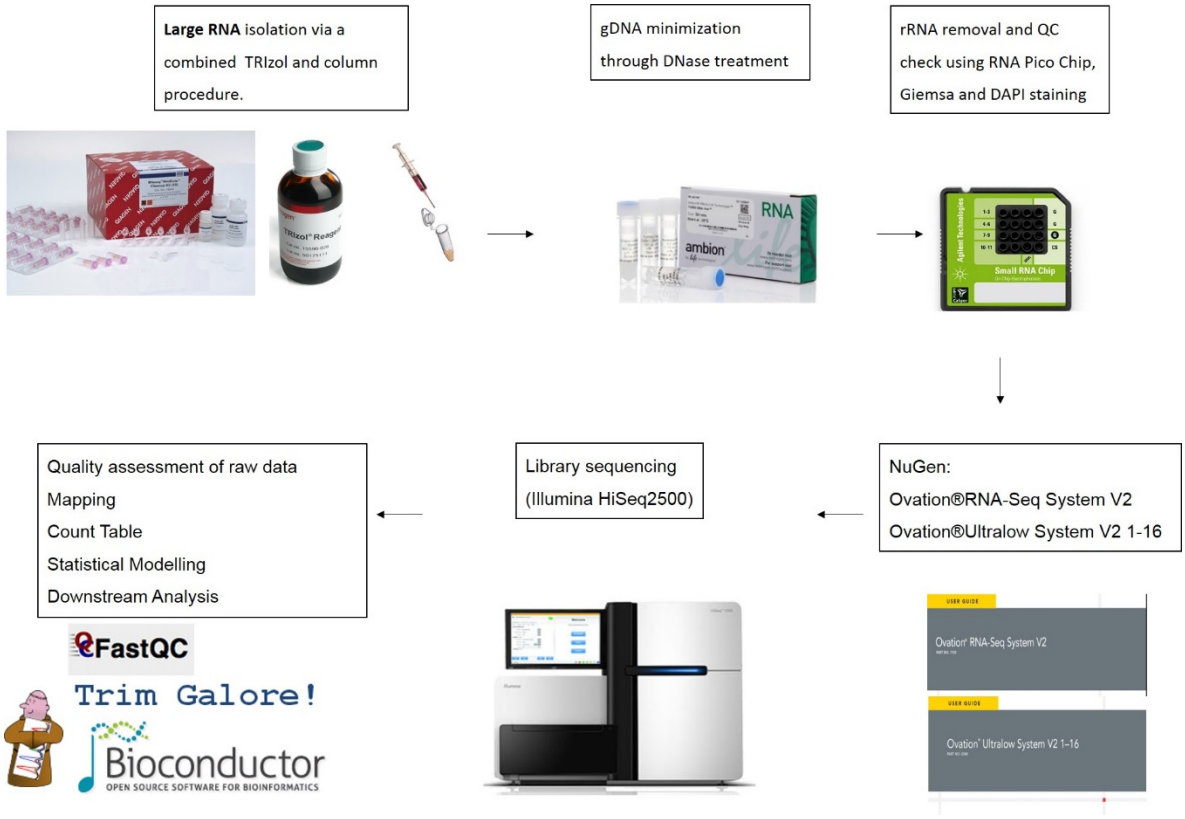


Figure 5-3: Workflow and experimental design for library preparation

5.3.4 RNA extraction and quantification for HA selected and unselected spermatozoal samples

For this approach, the focus lay specifically on the extraction of long RNAs. See Chapter 2, for RNA extraction and Chapter 3 for quantifications and DNase treatment.

5.3.5 Library construction and equimolar pooling for HA selected and unselected spermatozoal samples

5.3.5.1 Library production for RNA sequencing using the NuGEN Ovation system

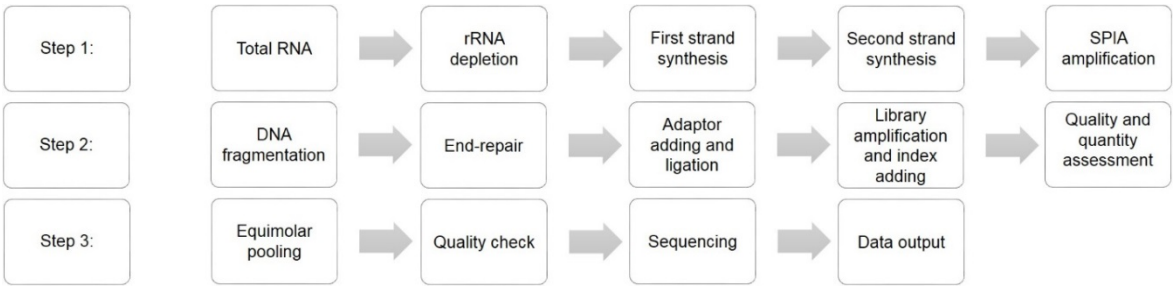


Figure 5-4: Library construction workflow and experimental design

Samples that passed the quality check for gDNA and somatic cell contamination were used for library synthesis (see Chapter 2; see workflow Figure 5-4). rRNA depletion, first and second strand synthesis and SPIA amplification was performed using NuGEN's 'Ovation® RNA-Seq System V2' (Version 7102). Working at the limit of RNA yields, 5 µl (25-50 ng RNA) of each sample was used as starting material. According to the manufacturer's recommendations, the minimum RNA input for NuGEN library construction system is 50 ng. Hence, some of the samples used for this study were below the recommended limit and their RNA could not be quantity control checked using either the Bioanalyzer or the Qubit. Library preparation, however, was continued using the 'Ovation® Ultralow Library Systems V2' (Version 0344) with an input of 50 ng cDNA for all samples after successful RNA transcription and cDNA amplification. If necessary, shearing was performed when the cDNA fragments were above 300 bp and quantification and quality control was completed after every amplification procedure as can be seen in Chapter 3, using Agilent's Bioanalyzer.

5.3.5.2 Equimolar pooling and quality control

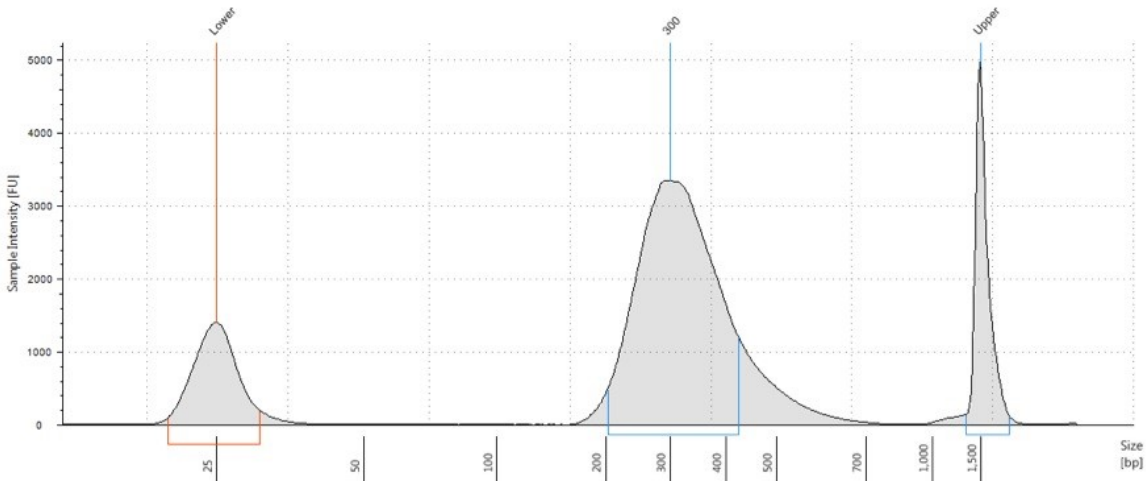


Figure 5-5: Electropherogramm traces of the pooled libraries. Bovine spermatozoal and testis libraries, ovine spermatozoal and testis libraries, porcine spermatozoal and testis libraries and human spermatozoal and testis were pooled (n=8), after indices attachment. Before the libraries were loaded onto the HiSeq2500 instrument, another quality control was performed using the Agilent 2000TapeStation. The sample is peaking at 300 bp and the marker is seen at 25 bp and 1,500 bp. FU=Fluorescent Unit; bp=base pairs

See Chapter 3 for equimolar pooling. The sequencing unit (St. James's Hospital, Leeds, UK) performed a quality control of the samples using the Agilent '2000TapeStation' prior to sequencing. The pooled sample should peak at 300 bp and not show any additional peaks. Peaks at 25 bp and 1,500 bp are from markers, showing the quality control for the run (Figure 5-5).

5.3.5.3 Sequencing

Sequencing was performed at the Sequencing Unit of St. James' Hospital, UK. See Chapter 3 for material and methods.

5.3.6 SeqMonk analysis

Comparison between RNA-Seq libraries was originally undertaken as described in Chapter 2. For HA-selected and unselected samples, an additional analysis was undertaken using the SeqMonk package (http://www.bioinformatics.bbsrc.ac.uk/projects /seqmonk/) (Andrews 2007). SeqMonk has built-in pipelines for data handling analysis using the original BAM files created following read mapping (see Chapter 2) and the relevant annotation files (Andrews, 2007). All data were imported using an RNA-seq pipeline focused on counting reads over exons with libraries scaled to counts per million reads (cpm). Cpm reads were then clustered into HA selected and unselected groups, for comparison by the exact Fisher test, which failed to reveal any statistically-significant differences between the groups (as had 'edgeR' using the equivalent limma package in R). However, closer examination of the data indicated that this was because very wide intra-group variations in expression had led to excessive noise in the analysis. As an alternative strategy, the group-averaged cpm values were filtered such that only values with a cut-off of ≥10 reads were exported to Excel for further inspection. This generated a list of 165 gene features that were further filtered to look only at highly abundant RNAs generating ≥600 cpm in the HA-selected group. Thirty-six (36) features were represented in this filtered set above 200 cpm (See Table in Appendix VI).

5.4 Results

5.4.1 NGS validation of fresh spermatozoal RNA vs. frozen spermatozoal RNA

5.4.1.1 Unassigned, ribosomal/mitochondrial and coding RNA reads

Comparing fresh with frozen spermatozoa indicated that ~14% and ~7.2% of reads, respectively, were not assigned in the fresh and frozen samples. Fewer reads of mitochondrial and ribosomal origin were seen in the fresh (~12.7%) compared with the frozen (~21.5%) sample. Coding RNAs in the fresh and frozen samples were ~33.9% and ~25.9% of all reads, respectively. See Figure 5-6 a) and b) for all percentages.

5.4.1.2 Other RNA types

~2.1% and ~2.4% or all reads, respectively, were designated as 'other RNA' types (long intergenic non-coding RNAs, snorRNA, miRNA) from fresh and frozen sperm. Of these, ~91.7% were identified as intergenic non-coding RNAs in the fresh sample compared with ~94.3% in the frozen, ~2.4% were identified as snorRNA in the fresh sample compared with ~1.1% in the frozen and ~6% were identified as miRNAs in the fresh sample, compared with ~4.6% in the frozen (shown in Figure 5-6c) and d).

5.4.1.3 RNA repeats

In the fresh sample, ~37.2% of reads were found to be repetitive RNAs compared to ~42.9% in the frozen sample with the main differences between fresh and frozen arising from the ribosomal large and short subunits with ~11% more rRNA of the short unit in fresh and ~20% for the short RNA unit in the frozen sample. The RN7SL1 RNA was decreased by 5% in the frozen sample and the other repeats did not differ more than ~1% from each other (see Figure 5-6 e) and f).

A summary of the read coverage and transcript assignment is provided at Table 5-3.

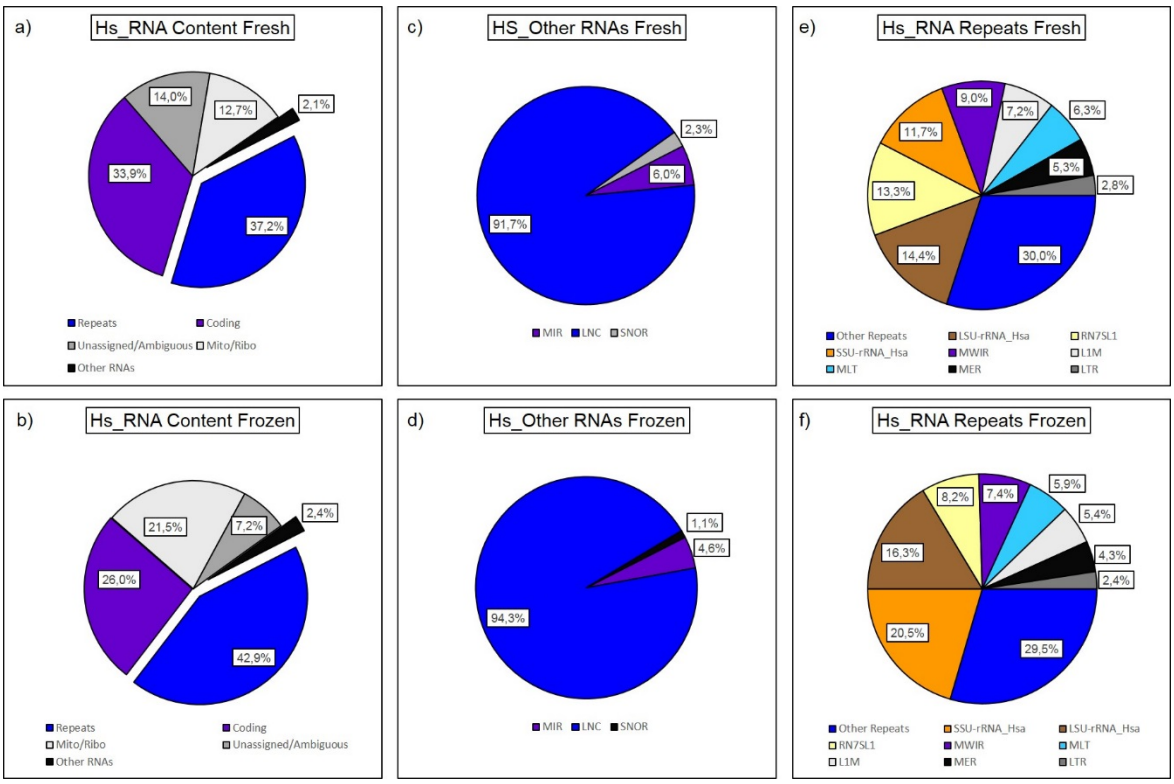


Figure 5-6: RNA content fresh sperm vs. frozen. a) RNA content fresh spermatozoa; b) RNA content frozen spermatozoa; c) Other RNAs fresh spermatozoa; d) Other RNAs frozen spermatozoa; e) RNA repeats spermatozoa: f) RNA repeats frozen spermatozoa. More reads were assigned in the frozen-thawed spermatozoa with ~14% in fresh spermatozoa and remaining ~7.2% in frozen. The main differences were seen in the mitochondrial/ribosomal reads where ~21.5% were seen in frozen spermatozoa and ~12.7% in fresh. The amount of coding reads were seen with ~7% difference and repeats in fresh spermatozoa with ~37.2% to ~42.9% in frozen. No big difference was visible in “other RNAs” with 97.7% of LINC RNAs in fresh and 94.3% in frozen spermatozoa, ~2.3% SNOR RNAs in fresh and ~1.1% in frozen spermatozoa and 6% of MIR in fresh to ~4.6% in frozen spermatozoa. The main difference in the repeats was the rRNA small subunits with 11.7% in fresh compared to ~20.5% in the frozen spermatozoa. All other repeats did not differ.

Species	RNA content	Fresh	Frozen
Human	Repeats	37.17%	42.93%
	Coding	33.93%	25.97%
	Unassigned/Ambiguous	14.05%	7.17%
	Mito/Ribo	12.74%	21.51%
	Other RNAs	2.11%	2.42%

Species	Other RNAs	Fresh	Frozen
Human	MIR	5.95%	4.64%
	LNC	91.70%	94.28%
	SNOR	2.35%	1.08%

Species	RNA Repeats	Fresh	Frozen
Human	Other Repeats	29.98%	29.52%
	LSU-rRNA_Hsa	14.35%	16.34%
	RN7SL1	13.26%	8.17%
	SSU-rRNA_Hsa	11.73%	20.49%
	MWIR	8.99%	7.36%
	L1M	7.23%	5.45%
	MLT	6.29%	5.95%
	MER	5.34%	4.31%
	LTR	2.82%	2.41%

Table 5-3: RNA content fresh sperm vs. frozen percentage.

5.4.2 DE analysis using the Bioconductor package edgeR

5.4.2.1 Data exploration and MDS plots plots using the Bioconductor package edgeR

5.4.2.1.1 Fresh and frozen samples

Biological and technical differences were seen plotting the raw reads of each sample using a MDS plot (see Figure 5-7). The frozen sample of sample 1 showed differences from the fresh sample in the first and the second dimension, whereas the replicate sample 2 clustered together in the fresh and frozen dataset showing dependence.

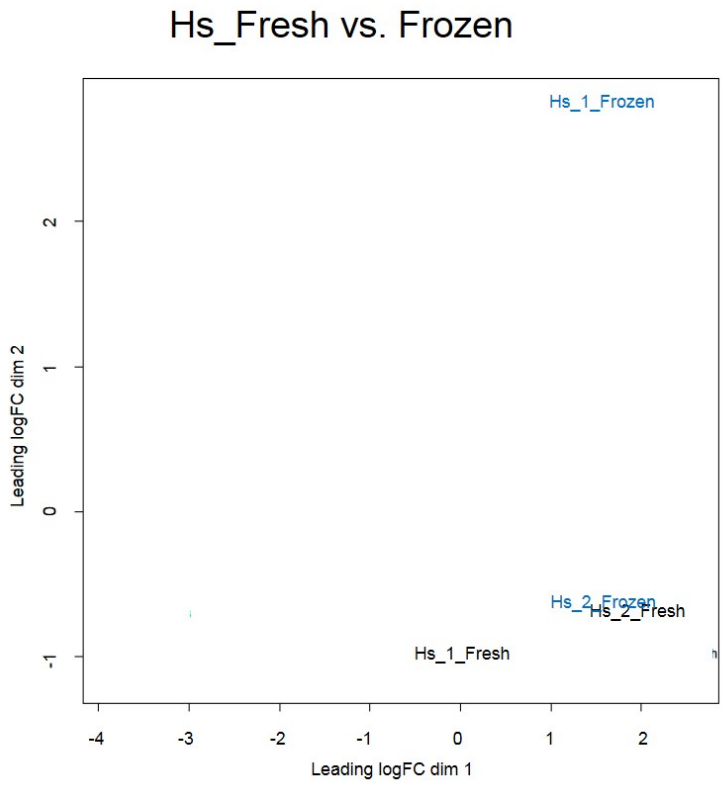


Figure 5-7: MDS plot fresh vs. frozen spermatozoa raw reads. Fresh and frozen replicates and samples are clustering in both dimensions, showing biological and technical dependency. One frozen sample (Hs_1_Frozen) is not clustering with the other samples, which shows differences to the other sample s. n=5 for each sample.

5.4.2.1.2 HA selected and unselected sampels

The technical and biological variation between each HA selected vs. semen (unselected) sample was summarised using a MDS plot prior to the differential expression analysis (see Figure 5-8 and explanation in Chapter 2). The MDS plot (Figure 5-8) showed clustering of selected and unselected samples, suggesting low technical and more significantly, low biological difference between both sample groups. The control, testis sample, showed the highest variation between the selected and unselected samples, clustering at the opposite end of the second dimension. More variation between individuals was seen than between HA selected and unselected spermatozoa.

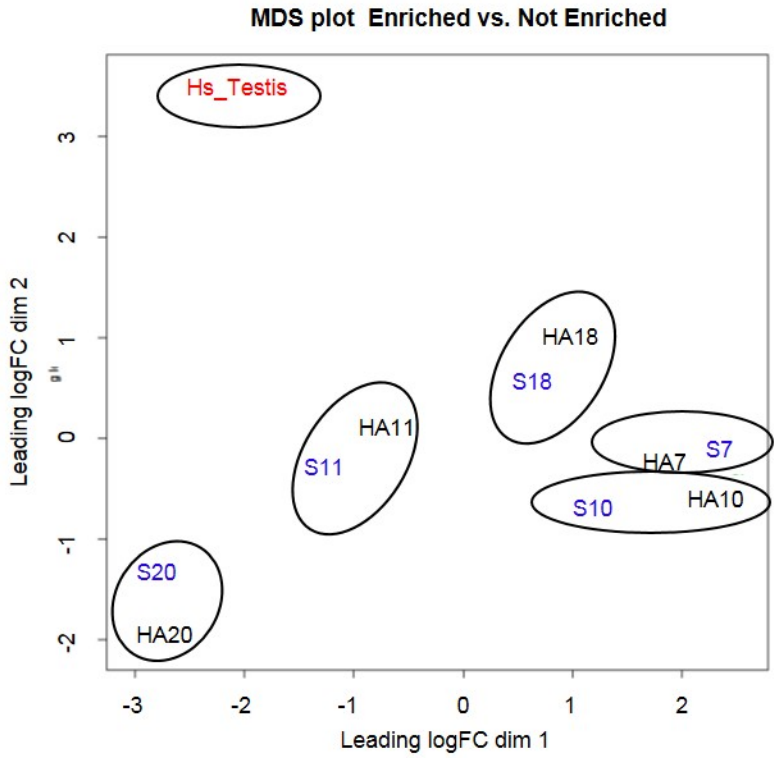


Figure 5-8: MDS plot of selected (HA) vs. unselected samples (S). All replicates form a cluster in the first and second dimension. The closer the samples lie the less biological and technical difference can be seen. n=1 for each library

5.4.2.2 Correlations analysis and DE transcripts

5.4.2.2.1 Fresh and frozen samples

Spearman's correlation showed r_s=0.53 (Figure 5-9), which indicates a distinct correlation between fresh and frozen raw reads. To identify DE transcripts, edgeR as a Bioconductor tool was used. Using edgeR, two differential expressed ribosomal transcripts (18S rRNA and 28S rRNA) between fresh and frozen samples were discovered, all other 12007 genes were shared and not significantly different in both sample sets (Figure 5-9).

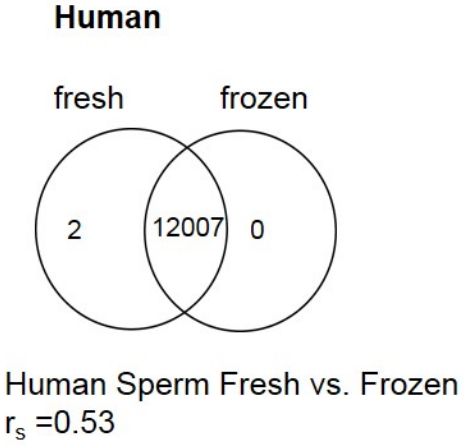


Figure 5-9: Spearman's correlation and DE transcripts of fresh vs. frozen spermatozoa. The correlation of fresh and frozen replicate reads were estimated and a distinct correlations seen. After edgeR analysis both 18S and 28S rRNA transcripts were seen as significant differentially expressed. 12007 transcripts were seen, shared between both sample groups.

5.4.2.2.2 HA selected and unselected samples

edgeR analysis did not reveal any significant differences between the HA-selected samples compared to the unselected fraction. All 13,582 RNAs reported were shared between each group (see Venn diagram Figure 5-10a)), which is also shown by the smear diagram (Figure 3-10b)) where significant differences would be highlighted by the colour red. This was also confirmed by the Spearman's correlation analysis by showing a distinct positive correlation (r_s=0.78) between the two groups.

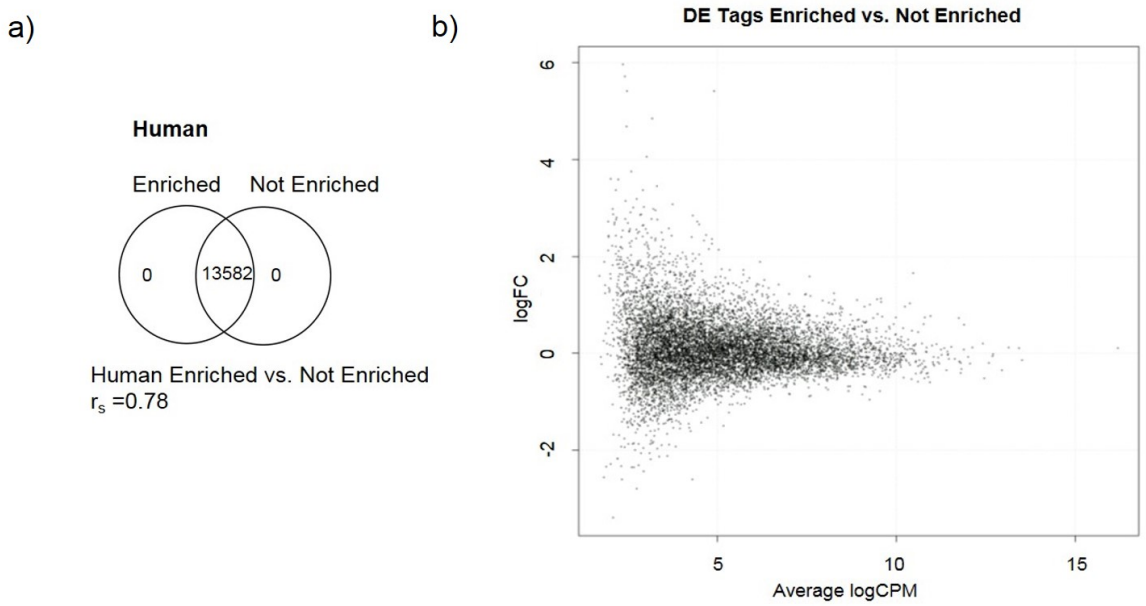


Figure 5-10: DE transcripts, Spearman's correlations and DE transcripts between HA selected and unselected samples. a) Statistical analysis did not highlight a difference between the two sample groups. All 13,582 transcripts were not significantly differentially expressed. Spearman's correlation revealed r_s=0.78, which designates a distinct correlation between both sample populations. b) No significantly differential expressed transcripts were found between the two groups, which is indicated by the black colour of the dots in the smear diagram.

5.4.3 SeqMonk analysis for HA selected and unselected samples

Although the selected and unselected groups were tested for differences following R-based 'edgeR' analysis, a parallel analysis was undertaken using SeqMonk. The SeqMonk RNA-seq pipeline filtering highlighted 36 features in the HA-selected group with cpm ≥600 (see Appendix VI).

A clear trend for increased expression of a number of RNAs was seen, including *MOSPD3* (*Motile Sperm Domain containing 3*), *RBFOX3* (*RNA binding protein, fox-1 homolog 3*), *RAB35*, (*RAB35, member RAS oncogene family*), *MTPAP* (*Mitochondrial poly(A) polymerase*), *PIGQ* (*Phosphatidylinositol glycan anchor biosynthesis class Q*) and *JUNB* (*proto-oncogene, AP-1 transcription factor subunit*) (see Appendix VI) in HA-selected spermatozoa.

Manual examination of the data uncovered a ~x 3 greater abundance of reads for *MOSPD3* RNA (selected spermatozoa: 1,587 cmp vs. unselected spermatozoa: 501 cpm). This trend was confirmed in the previous data extracted via Bioconductor and *MOSPD3* was also found ~3 times higher expressed in selected spermatozoa with 322.87 cpm (average) to 108.73 cpm (average) in the unselected fraction. Visualising the *MOSPD3* data for each sample using the UCSC genome browser selected spermatozoa (Figure 5-11a) and unselected spermatozoa (Figure 5-11b), the black box indicates the scale for all samples) confirmed the trend for greater abundance in selected versus unselected samples.

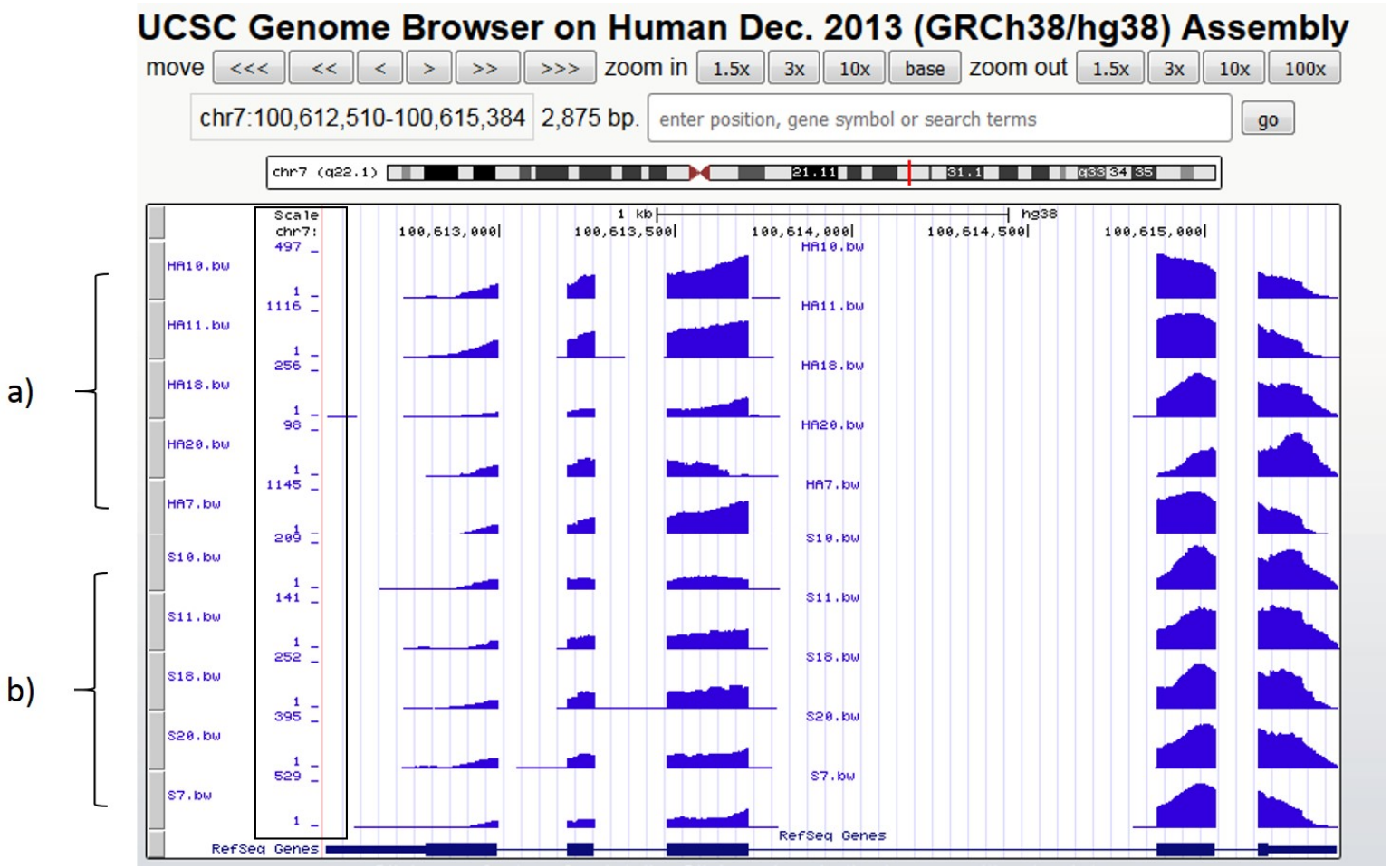


Figure 5-11: UCSC profile of *MOSPD3* for all samples. a) HA selected spermatozoa fraction, followed by b) their paired unselected sample. All HA selected spermatozoa samples show a trend in increased expression of the transcripts compared to their paired unselected sample, see black box for the scale.

5.5 Discussion

Examining the replicate sets using an MDS plot revealed differences in the first fresh and frozen sample set. As individual samples were pooled prior to splitting into fresh and frozen batches for RNA isolation, biological differences can be excluded as a likely cause and an unidentified technical cause is more likely.

Spearman's correlation between fresh and frozen samples shows an r_s=0.53, reflecting the sample difference in the first replicate set. The replicates correlation would be more significant, if the first sample set of fresh and frozen would cluster closer together. The outlying pool cannot be excluded from the analysis, since only two replicates were run. However, two rRNA transcripts showed higher differential expression in the fresh spermatozoa, seen as the only difference between the sample sets. More than two replicates and additionally single sample sequencing would be required to see a clear significance in differences and trends between fresh and frozen samples.

Nonetheless, comparing the RNA content of fresh and frozen pools did not reveal differences between them, despite the outlier. It was noted that rRNA and mtRNA was more likely to be degraded in the frozen samples and therefore less likely to be efficiently removed by the rRNA depletion step carried out prior to library construction. Frozen-thawed spermatozoa appeared to show a small (~8%) reduction in coding RNAs while a ~5% reduction in repeat RNAs was detected in extracts from fresh spermatozoa, although these differences may be due to technical variation. The total RNA isolation approach includes a column-step that is not designed to preserve small RNA content. However, this did not lead to a total loss of miRNAs and revealed ~2% difference in miRNA reads between fresh and frozen samples.

Whilst DNA damage and fragmentation can be examined using a number of techniques, including acridine orange staining and (TdT)-mediated dUTP nick end labelling (TUNEL), RNA/protein damage and changes in DNA methylation changes leading to epigenetic alterations cannot be examined this way (Donnelly *et al.*, 2001b). As there is an increasing body of evidence suggesting that intact DNA is not the only spermatozoon-derived factor for successful reproduction and embryo development (Ostermeier *et al.*, 2004; Miller and Ostermeier, 2006a; Lalancette *et al.*, 2008a; García-Herrero *et al.*, 2010). It is important to take steps ensuring that frozen stored spermatozoa are as similar as possible to fresh spermatozoa and that spermatozoal RNAs and proteins survive the freeze-thaw process intact. Failure to achieve this could compromise embryo development or lead to a sudden termination of pregnancy and increased health risk to progeny. (Takemoto *et al.*, 2016)

Although the majority of spermatozoa are physiologically effected of freeze-thawing processes, by permanent cell membrane damage, morphological changes, reduced motility and decreased mitochondrial function the results of this chapter suggests that the widespread use of freeze-thawing processes does not affect spermatozoal RNA deleteriously (Hammerstedt *et al.*, 1990; Kempisty *et al.*, 2008; Bissonnette *et al.*, 2009; Das *et al.*, 2010; Sendler *et al.*, 2013). Although this is the only study focusing directly on the direct iatrogenic effects of freeze-thawing on spermatozoal RNA composition, surveys of RNA types reported in studies making use of either fresh or frozen bovine, ovine and human spermatozoa have been similar to those that reported here (Card *et al.*, 2013; Anderson, 2013; Kempisty *et al.*, 2008; Lalancette *et al.*, 2008b). Hence, while RNA profiles are rarely identical, studies examining RNAs in human, bovine and porcine spermatozoa are generally comparable and need not be concerned with freezing artefacts, provided the process is carried out correctly (Kempisty *et al.*, 2008; Lalancette *et al.*, 2008b; Card *et al.*, 2013; Anderson, 2013; Jodar *et al.*, 2013; Sendler *et al.*, 2013). More significantly, the reliance of ART on frozen stored semen is also not brought into question by these results.

Despite using a RNA input below the recommended minimum for HA selected and unselected samples (NuGEN's 'Ovation® RNA-Seq System V2', Version 7102), library construction was successful.

ART offspring have an increased risk of preterm birth, low birthweight, congenital anomalies and perinatal mortality that may be transmitted through spermatozoal RNAs (Hansen *et al.*, 2002; Kalra and Molinaro, 2008; Dias and Ressler, 2014; Gapp *et al.*, 2014). HA selection, therefore (and other forms of spermatozoa selection) may help support ART, by reducing the incidence of using spermatozoa with high levels of DNA fragmentation. HA selected spermatozoa also have lower frequencies of chromosomal aneuploidy (Huszar *et al.*, 2007; Parmegiani *et al.*, 2010a). Furthermore, immature spermatozoa that cannot bind to HA may have uncompleted histone to protamine exchange and higher levels of cytoplasmic retention (Huszar *et al.*, 2007). Taken together, HA spermatozoal selection probably mimics natural spermatozoon selection (Huszar *et al.*, 2007). The results of this investigation support the finding that spermatozoa binding to HA represent a more viable population and should therefore be further examined with regard to their transcriptome, including miRNAs. For this experiment, an insufficient separation between the HA-binding and non-binding spermatozoa in the original separation experiments was provided from an external laboratory and given that the right separation of HA selected and unbound samples are used for further analysis significant differences in HA bound to unbound spermatozoa revealing infertility can be found, which can may be developed into a diagnostic tool to

diagnose infertility faster and furthermore be used to find the best spermatozoon as candidate for ICSI treatments (Yagci *et al.*, 2010; Huszar *et al.*, 2007).

Comparison between HA selected and unselected samples showed no significant differences using either Bioconductor's edgeR tool or Fisher's exact test in SeqMonk. The former revealed 13,582 mutual transcripts (see Figure 5-10) among the 10 samples. SeqMonk, however, was set to quantify the data using an RNA-seq pipeline focusing on coding RNAs alone and revealed a clear trend for increased expression of a number of RNAs, including *RBFOX3* (RNA binding protein, fox-1 homolog 3), *RAB35*, (RAB35, member RAS oncogene family), *MTPAP* (mitochondrial poly(A) polymerase), *PIGQ* (phosphatidylinositol glycan anchor biosynthesis class Q) and *JUNB* (proto-oncogene, AP-1 transcription factor subunit) and *MOSPD3* (see Appendix VI) in HA-selected spermatozoa. *MOSPD3* was chosen for further investigation due to time concerns and because it seemed coincidentally relevant to spermatozoal motility. *MOSPD3* contains a *MSP* domain, which functions in the amoeboid motility of nematode spermatozoa by promoting an actin-like polymerisation and depolymerisation process (Italiano *et al.*, 1999; Roberts and Stewart, 2000).

MOSPD3 is located at chromosome 7 in the human and has 5 splice variants varying in size between 225 aa and 235 aa (information taken from protein atlas, http://www.proteinatlas.org/). The protein contains a transmembrane domain and a MSP (Major Sperm Protein) domain. MSP is involved in spermatozoal motility in nematode worms, functioning as part of the cytoskeleton and therefore in the amoeboid movement of the spermatozoon (Pall *et al.*, 2004; Italiano *et al.*, 2001; Roberts and Stewart, 2000). Silencing MSP through RNA interference (RNAi) leads to the loss of motility in nematode spermatozoa, indicating an important function in spermatozoal movement (Buerger, 2010). Furthermore the MSP domain seems to act as a signalling molecule playing a role in nematode oocyte meiotic maturation and ovarian muscle contraction (Kosinski *et al.*, 2005; Miller *et al.*, 2001). Little is known about *MOSPD3* in mammals and the only available study showed that *MOSPD3* depleted homozygote mice are highly lethal with little survivors (Buerger, 2010). This lethality is due to *MOSPD3* playing an important role in cardiac development and function (Buerger, 2010). In this context, a more detailed expression and functional follow-up analysis of MOSPD3 expression was performed by immunocytochemistry and Western blot analysis that could lead to confirmation of a role for the protein in motility and fertilisation events in the human spermatozoa (see Chapter 6).

5.6 Conclusion

Although two replicates each (fresh and frozen) of five pooled donor samples each are insufficient to be conclusive to analyse if cryopreservation introduces significant damage, the results of these experiments suggest freezing is not detrimental to spermatozoal RNA content and that any effect is likely to be relatively minor. Ideally, single sequencing of all the samples should be considered, since pooling of the samples may hide weaker trends relating to freeze thawing. In this regard, small changes in RNA content may still have unintended consequences for ART offspring and it may be that other types of RNA, particularly miRNAs and tRNAs may be more important in this regard and warrant further investigation.

Stringent edgeR analysis of HA selected spermatozoa compared to unselected samples did not show any significant differences in RNA expression between the two fractions, which was confirmed by Spearman's correlations. However, a trend towards the increased expression of *MOSPD3* in HA-selected spermatozoa led to its further investigation. Most research into the function of this protein has been performed in nematodes where MSP is involved in spermatozoal movement, oocyte meiotic maturation and ovarian muscle contraction. MOSPD3 contains a MSP domain and causes a lethal defect in the right heart ventricle in *Mospd3* knock-out mice. Comparatively little is known about its function in human; however, comparisons between HA-bound and unbound spermatozoa may facilitate the development of tools to select the best spermatozoa for ART treatment or to help diagnose infertility earlier. Additional analysis of MOSPD3 expression in HA selected spermatozoa, compared to the unbound fraction was performed in Chapter 6. Providing more input RNA would also permit a wider consideration of other RNA types including miRNAs. Recent studies have shown that alterations in spermatozoal borne miRNAs and other sncRNAs may be involved in infertility events, embryogenesis, early miscarriage (particularly following the use of ART) and in increasing health risk factors of ART offspring in adulthood (Carrell and Hammoud, 2010; Rodgers *et al.*, 2013; Gapp *et al.*, 2014; Dias and Ressler, 2014; Sharma *et al.*, 2016; Chen *et al.*, 2016).

Chapter 6: Revealing the Motile Sperm Protein Domain 3 Protein as a Potential Marker for Spermatozoal Quality

6.1. Introduction

In recent years, the diagnostic potential of human semen RNA profiles has been explored using microarray and RNA-sequencing based approaches (Bonache *et al.*, 2012; García-Herrero *et al.*, 2011; Garrido *et al.*, 2013). Although results to date have shown some promise in identifying particular RNAs or groups of RNAs involved in infertility that may be diagnostically useful, RNA markers as diagnostic tools have some disadvantages when compared with proteins. The main problems include the destructive nature of RNA isolation and the small quantity of extracted RNA in spermatozoa compared with proteins. RNA sequencing analysis of human HA- selected vs. unselected spermatozoa in Chapter 5 revealed a trend for *MOSPD3* upregulation in HA- selected spermatozoa. Therefore, the focus of the work reported in this chapter was on MOSPD3 as a potential protein fertility marker.

6.1.1 Motile Sperm Domain-Protein 3

The *Motile Sperm Protein Domain 3 (MOSPD3)* gene is located on chromosome 7 in humans and is known to encode for a membrane protein with a size of 235 amino acids (aa) (protein atlas) as well as five splice variants with slightly different sizes (225 aa - 235 aa). MOSPD3 is conserved between human and mouse and contains a Major Sperm Protein (MSP) domain and two transmembrane domains, where the presence of the transmembrane domains suggest an involvement in the formation of membrane-associated protein networks (Pall *et al.*, 2004). MOSPD3 belongs to the MOSP family whose function in human is largely unknown. The protein seems to be involved in embryonic development, with a lethal phenotype in transgenic mice homozygous for an integrated gene trap vector (Pall *et al.*, 2004; Kara, 2012; Buerger, 2010). The lethal defect occurred in the cardiac right ventricle (Forrester *et al.*, 1996; McClive *et al.*, 1998; Pall *et al.*, 2004; Kara, 2012; Buerger, 2010). MOSPD3 protein is expressed in embryonic heart, kidney, liver and spleen and is also found in adult heart tissue (Pall *et al.*, 2004; Buerger, 2010).

6.1.2 Function and structure of the Major Sperm Protein

The MSP was first reported as the Major Sperm Protein of *Caenorhabditis elegans*, which expresses over 30 closely related MSP genes (Burke and Ward, 1983; Scott *et al.*, 1989). Tarr and Scott (2005) identified new members of the MSP family in *Ascaris suum* using a bioinformatical approach and MSP genes have been identified in 14 different species of nematodes to date (Scott *et al.*, 1989). The functions in all nematodes seem similar. Unlike flagellated mammalian spermatozoa, nematode spermatozoa are amoeboid cells, with motility driven by the dynamics of an actin-like cytoskeletal protein, the MSP protein. The MSP protein seems to be involved firstly and more importantly in cell motility, allowing nematode spermatozoa to 'crawl' and secondly to promote oocyte maturation and sheath cell contraction in the female by acting as a signaling molecule (Italiano *et al.*, 2001; King *et al.*, 1994; Miller *et al.*, 2001). The protein in *A. suum* is built up through seven strands of beta pleated sheets, which collectively resembles an immunoglobin-like domain and its assembly and disassembly during cell motility is mediated by regional differences in intracellular pH, whereas the *C. elegans* MSP consists of a dimer (Bullock *et al.*, 1996; Singaravelu and Singson, 2011). Dimeric MSP functions as a building block of fibres, polymerising into a higher order structure and driving motility by forming a filament-packed pseudopod, which can change direction through altering its shape (Baker *et al.*, 2002; Roberts and King, 1991; King *et al.*, 1994; Sepsenwol and Taft, 1990). MSP functions as part of the cytoskeleton, driving the locomotion in nematode spermatozoa in a manner which is almost indistinguishable to the actin-based motility seen in other cell types (Roberts and Stewart, 2000). Silencing of MSP function leads to blocking of motility and hence contributes to male infertility in nematodes (Buerger, 2010).

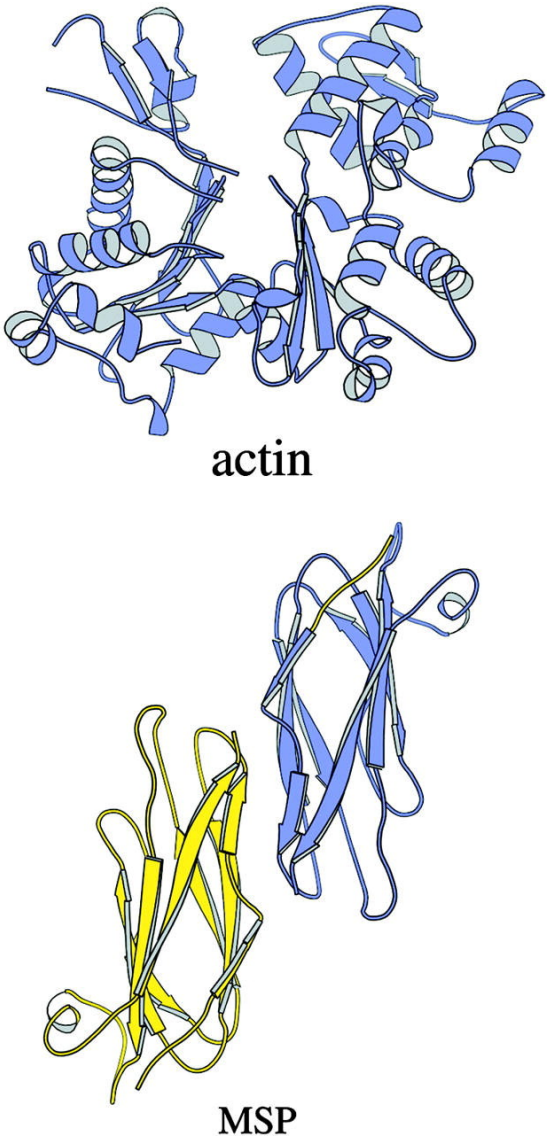


Figure 6-1: Ribbon diagrams of Dictyostelium discoideum G-actin (Matsuura *et al.*, 2000) and the Ascaris suum α-MSP dimer (Bullock *et al.*, 1996) at the same magnification.

6.1.3 Comparison of MSP and Actin based motility

Though MPS and actin appear to fulfil similar roles, their biochemical properties are quite different and they do not share sequence homology with each other nor with other proteins of the cytoskeleton (Roberts and Stewart, 1995). While MSP functions as a dimer, actin is monomeric, and also binds nucleotides (Figure 6-1) (Roberts and Stewart, 2000). The four subunits of the actin filament are arranged as beads on a string, compared to MSP in which the filaments are constructed from two loosely connected helical subfilaments (Roberts and Stewart, 2000; Stewart *et al.*, 1994). While actin enables motility through the formation of an actin-rich pseudopod structure having a characteristic structural polarity in amoeba and white blood cells, for example, the

nematode spermatozoon does not contain enough actin to form pseudopodia. Instead, it uses MPS filament assembly at the leading edge of the locomoting spermatozoon and disassembly near the cell body at the base of the pseudopod to generate a spermatozoal tread milling locomotion process (Nelson *et al.*, 1982; King *et al.*, 1994; Roberts and King, 1991; Roberts and Stewart, 2000). MSP and Actin assembly and disassembly require ATP hydrolysis together with other accessory proteins to arrange a controlled dual ended polymerising building process (Italiano *et al.*, 1996; Roberts and Stewart, 2000). Additional studies have taken a more detailed look into the orchestration of these proteins and found protein-protein interaction domains associated with the actin cytoskeleton and its organisation, as well as nematode spermatozoa specific clusters containing either a LIM domain (composed of two zinc finger domains separated by a two-amino acid residue hydrophobic linker) or a PDZ domain (structural domain of 80-90 aa) (Tarr and Scott, 2004).

Taken together, both processes appear morphologically and dynamically analogous using a similar mechanical principle, although being based on an similar orchestration but different proteins involved (Buerger, 2010; Roberts and Stewart, 2000).

6.1.4 Major Sperm Protein (MSP) domain proteins in other eukaryotic organisms

More light has to be shed onto the general functional information of the MSP-domain proteins, as apart from nematodes, almost nothing is known about these proteins in other animals. The only other functional information of this type of protein comes from the study of vesicle associated proteins (VAP). Though VAP genes encode an N-terminal MSP domain as well as a C-terminal transmembrane domain (see Figure 6-2), MOSPD proteins do not contain the coiled-coil domain found in VAPs (Loewen and Levine, 2005). Homologous proteins (see Figure 6-2) to VAP have been described in *H. sapiens* (Nishimura *et al.*, 2004), *S. cerevisiae* (Kagiwada *et al.*, 1998; Loewen and Levine, 2005), *Drosophila* (Pennetta *et al.*, 2002) and plants (Laurent *et al.*, 2000). VAPs act as mediators in protein-protein interactions and bind to different membrane fusion proteins (Weir *et al.*, 1998). Buerger (2010) speculated that MSP proteins might have a similar function.

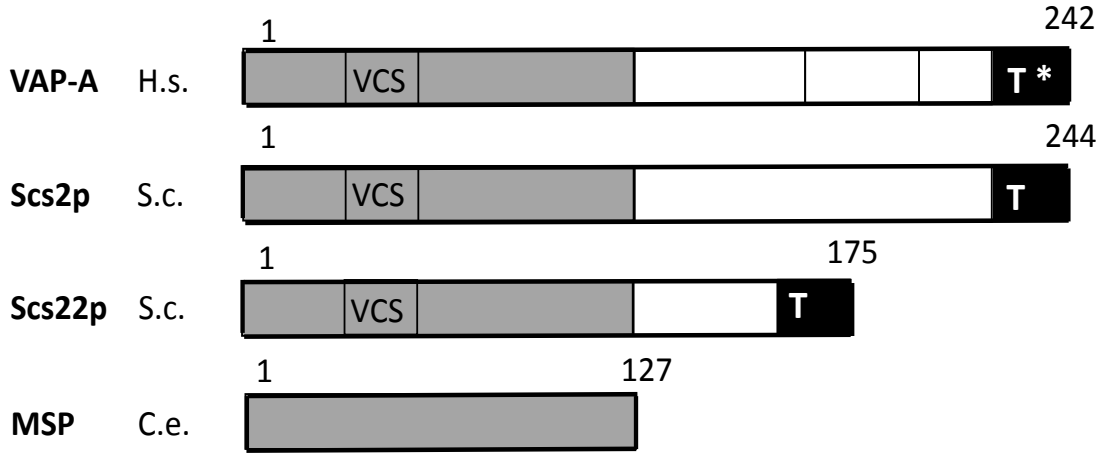


Figure 6-2: Domain structure of VAP homologues. In humans and other metazoa, a MPS domain is imbedded into homolog proteins (grey rectangle), a linker that contains a coiled-coil motif and a carboxyl-terminal transmembrane domain (black rectangle marked T). Scs2 and Scs22 VAP homologues in *S. cerevisiae* and other fungi have a similar domain structure but no coiled-coil in the linker, which is much shorter in Scs22. VCS stands for VAP consensus sequence and seems highly conserved in VAPs, whereas the nematode MSP protein consists of the MSP domain alone. Figure adapted from Loewen and Levine (2005).

6.2 Aims

To examine MOSPD3 expression in relation to spermatozoal motility.

6.3 Material and methods

6.3.1 Sample collection and processing

Fresh normozoospermic human semen samples (n = 3) obtained following standard ethically approved and consenting guidelines were divided into two fractions; one was used to perform a two layered differential density gradient centrifugation (DDGC) and the other fraction was washed with spTalp and centrifuged (see material and methods 2.1.4). Both fractions were used for subsequent Western blot analysis. Frozen normozoospermic and oligoasthenozoospermic samples (n=3) (kindly received from the IVF unit, Seacroft Hospital, Leeds, UK) were used for immunocytochemistry and separated into three layers (90%, 45% and 20%) by DDGC. The interfaces between the layers together with the pellet were collected for further analysis.

6.3.2 Protein extraction and concentration

1 ml RIPA buffer (Table II-6) was added to ~20 x 10^6 pelleted spermatozoa. A flask of 70% confluent MCF-7 cells (a kind gift of Dr. Sandra Bell, LIMM, University of Leeds, UK) was washed several times with 1 x PBS before adding 1 ml of RIPA Buffer (Table II-6) for protein extraction. Mechanical homogenization using a G26 needle was carried out 10 times, followed with sonication 10 x 5 seconds at maximum speed (Soniprep 150, MSE Sanyo). This was followed by a rotating incubation at 4°C for one hour and centrifugation for one minute at 14,000 x g at 4°C (Eppendorf, Centrifuge 5415R). The supernatant was transferred into a sterile reaction tube. Proteins were concentrated using Amicon®Ultra-0.5 Centrifugal Filter Devices 3K; 14,000 x g at 4°C for 20 minutes (Eppendorf, Centrifuge 5415R). Samples were kept at -20°C until usage.

6.3.3 Pierce™BCA protein assay

The bicinchoninic acid (BCA) protein assay was developed by Smith *et al.* (1985) and enables determination of the concentration of proteins in a solution. It is based on the interaction of two reactions: First, the peptide bonds in protein reduce Cu^{2+} ions from the copper (III) sulfate to Cu^{+} and then two molecules of bicinchoninic acid chelate with each Cu^{+} ion forming a purple coloured complex with proteins that absorbs light at the wavelength of 562 nm. The BCA protein assay was performed by diluting the protein assay dye reagent concentrate 1:50 with solution A and the whole solution was diluted 20:1 to the protein lysate and incubated at 37°C for 30 mins and the absorption measured at 562 nm in a spectrophotometer (Varioskan Flash Multimode Reader, Thermo Fisher Scientific). Protein amounts were extrapolated from the measured standard curve.

6.3.4 Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis

Using Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE), proteins can be separated according to their electrophoretic mobility (Laemmli, 1970; Towbin *et al.*, 1979). Electrophoretic mobility is determined by the molecular weight and length of the polypeptide chain (Shapiro *et al.*, 1967). By incubating the proteins with SDS, the linearized proteins receive a stoichiometric negative charge (Laemmli, 1970). For the SDS-PAGE, the Mini-Protean Tetra Electrophoresis System (Bio-Rad Ltd.) was used that contained 10-well, 1.5 mm thick gels. A 12.5% resolving gel (Appendix II, Table II-7) was cast and overlaid with a 4% stacking gel (Appendix II, Table II-8).

10 µg protein from the MCF-7 cell culture, pelleted spermatozoa and the different spermatozoal layers were mixed with 1 x sample buffer (Appendix II, Table II-10), 5% (v/v) β-Mercaptoethanol was added and boiled at 95°C for 5 minutes to linearize the proteins. After denaturation, the samples were loaded onto the gel together with 5 µl of a prestained protein ladder. For the 12.5% separating gel with 4% stacking gel, a vertical tank system (Bio-Rad Ltd.) was used. Proteins were separated in 1 x Tris-Glycine running buffer (Appendix II, Table II-9) and the gel run until proteins were adequately resolved, i.e. at 120 V for 1.5 hours. A Coomassie stain (Brilliant R -250 Blue) was performed afterwards to visualize the separated proteins (Chen *et al.*, 1993).

6.3.5 Protein transfer, blocking and detection

Separated proteins can be transferred onto a Polyvinylidene difluoride (PVDF) membrane, where the target protein can be labelled with an antibody (Towbin *et al.*, 1979; Renart *et al.*, 1979; Matsudaira, 1987). Proteins were transferred to a methanol activated PVDF membrane by preparing a transfer sandwich. The transfer took place at 250 mA for 1.5 h using a mini trans-blot (Bio-Rad Ltd.) and 1x transfer buffer (Appendix II, Table II-11). After blotting, the membrane was blocked in 5% slim milk solved in TBST (Appendix II, Table II-12) for 1 h at RT or overnight at 4°C. The MOSPD3 primary antibody (HPA041137; Concentration: 0.5 mg - 1 mg/mL; SIGMA) was diluted 1/200 in TBST plus 2% slim milk, GapDH 1/5000 (GTX28245; Concentration: 12.8 mg/ml; GeneTex, kind gift of Dr. Lynn McKeown) in TBST plus 5% milk. Membranes were incubated in 750 µl antibody-containing solution in a moisturised chamber at 4 °C overnight or 1 h at room temperature. After incubation, the membrane was washed three times with TBST for 15-20 mins each, before incubation with the secondary peroxidase-coupled antibody in TBST and 2% slim milk, for 1h at RT took place (goat anti-rabbit IgG-HRP diluted 1/2000; 656120, Concentration: 1 mg/ml; ThermoFisher Scientific; donkey anti-mouse IgG (H+L), 715-035-150-JIR, Antibody Concentration: 0.8 mg/ml; Stratech Scientific, diluted 1/10000; which was a kind gift from Prof. Dr. Beech's laboratory). The membrane was again washed three times in TBST for 15-20 min. The drained membrane was incubated with SuperSignal^{Tm}West Femto Maximum Sensitivity Substrate according to the manufacturer's instructions. Protein signals were detected with the Geldoc.

6.3.6 Stripping for re-probing

Membranes where stripped for re-probing purposes using the mild stripping buffer protocol from Abcam® (15 g glycine, 1 g SDS, 10 ml Tween20, pH 2.2 in 1 L of ultrapure water). The membrane was covered with the mild stripping buffer for 10 minutes, before the buffer was replaced with a fresh one for further 10 minutes. After discarding the mild stripping buffer the membrane was incubated with PBS for 10 minutes twice and TBST for 5 minutes twice. After the incubations the membrane was ready for the blocking stage and could be reused.

6.3.7 Statistical quantification for Western Blot analysis

For quantification purposes three blots were assessed using ImageJ software to determine the band intensity, the background was subtracted. Using the intensity values the protein of interest (MOSPD3) was normalized against the control protein (GAPDH) and the different DDGC layers used (90% fraction and 45% fraction) of the spermatozoal samples were taken

into the calculations using a paired one tailed Mann-Whitney U test. Since a normal distribution cannot be guaranteed with n = 3, the t-test was not be used.

6.3.8 Immunocytochemistry

10^5 spermatozoa were spun down using a cytospin centrifuge (Cytospin3, Shandon) onto poly-L-lysine coated glass slides at ~10,000 x g for 15 minutes before a methanol / acetic acid (ratio 3:1) fixation was performed at room temperature for 10 minutes. The slides were washed with PBS three times for 10 minutes each and blocking was performed using 3% BSA. A dilution of 1:20 anti-MOSPD3 (H00064598-M05; Abnova) in 1% BSA incubation took place in a moisture chamber at 4°C over night. Secondary antibody incubation was performed using a goat anti-mouse IgG-TRITC (Ab6786, Concentration: 1 mg at 2 mg/ml; abcam) 1:1000 in 1% BSA in PBS, together with DAPI (see Chapter 2) and was incubated in a moisture chamber at room temperature for 60 min, before 3 x 10 minutes washing in PBS and mounting with DPX was performed at room temperature for 60 minutes. A negative control for all antibodies was performed.

6.4 Results

6.4.1 SDS PAGE gel

10 µg of protein was loaded onto two SDS PAGE gels, with one gel used for Coomassie staining (Figure 6-3) to check that proteins were present, and an identical gel was used for Western Blot analysis.

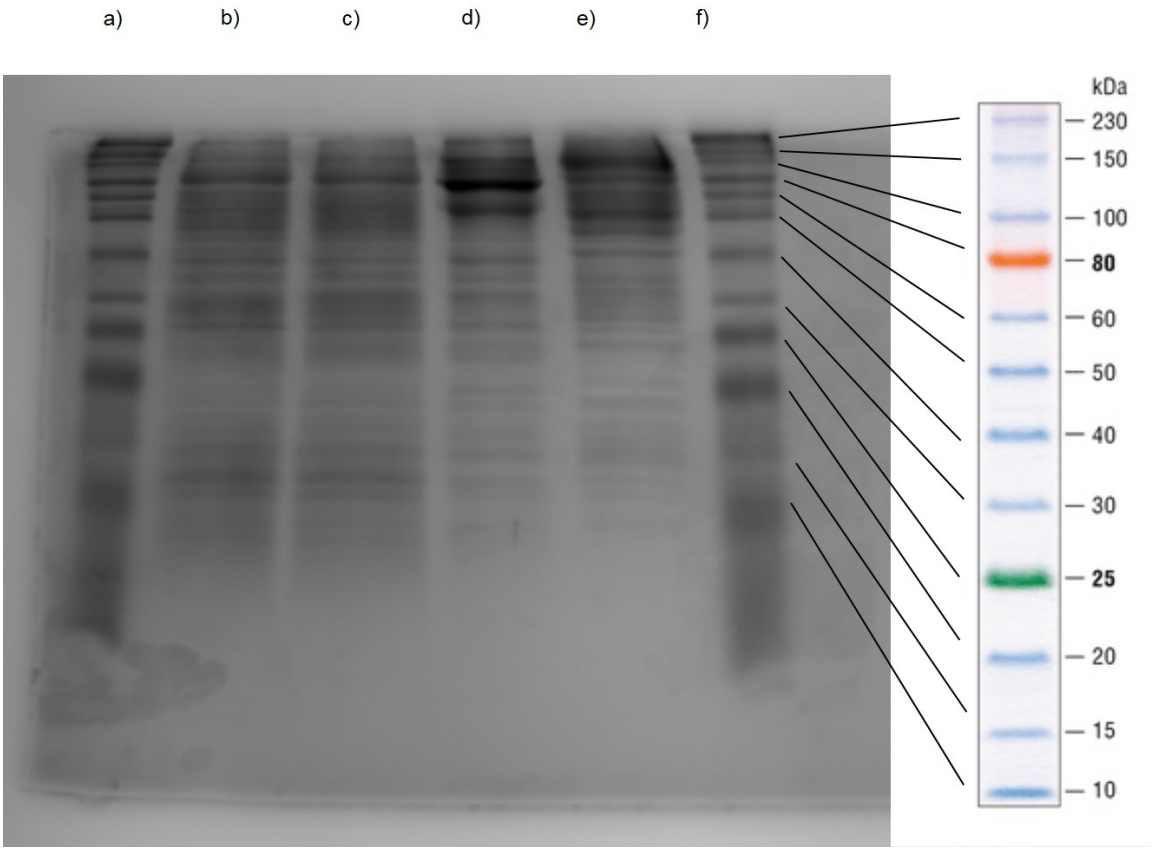


Figure 6-3: Coomassie stain for SDS PAGE gel. a) and f): Pre-stain protein ladder; b) 45 % fraction; c) 90% fraction; d) pelleted spermatozoa; e) MCF7 was used as a positive control. Two SDS PAGE gels were run at the same time under the same conditions, however, one SDS PAGE gel was Coomassie stained to see if the protein separation via electrophoresis worked.

6.4.2 Western Blot analysis

10 µg spermatozoal protein from the 45% interphase fraction (layer between 45% and 90%) (less motile spermatozoa) and 90% layer fraction (motile fraction), pelleted spermatozoa and MCF7 cells were quantified for their expression of MOSPD3 by Western Blot analysis. As can be seen in Figure 6-4 a gradual increase in MOSPD3 at ~26 kDa is visible starting with the 45% layer fraction, continuing with the 90% layer fraction and then the pelleted spermatozoa (pelleted whole semen). The loading control (GAPDH) shows that equal amounts of protein were loaded in each fraction. MCF7 cells show a stronger GAPDH signal, since they may contain more GAPDH than spermatozoa. A primary and secondary antibody Blot control was performed for all antibodies and did not show any non-specific binding.

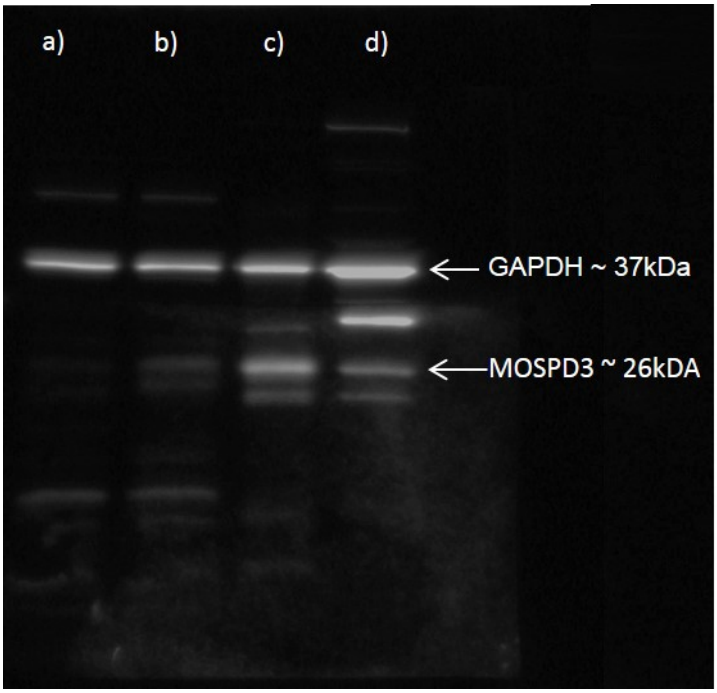


Figure 6-4: Western Blot analysis for MOSPD3. a) 45% fraction; b) 90 fraction; c) pelleted spermatozoa; d) MCF7. Expression of MOSPD3 was analysed by Western blotting of protein extracts (10 µg each) from less motile spermatozoa fractions (45%), motile spermatozoal fractions (90%), pelleted spermatozoa and MCF7 cells. The 45% layer fraction (less motile, interface fraction) shows a decrease of MOSPD3 compared to the 90% layer fraction (motile fraction) and the pelleted spermatozoa. Pelleted spermatozoa and MCF7 cells were used as assay controls. The arrows mark either the ~26 kDa MOSPD3 or the ~37 kDa GAPDH loading control. n=1 for each loaded sample

Values for statistical analysis of MOSPD3 and GAPDH of all three performed Western blots were quantified via densitometry using ImageJ. Using an one-tailed Mann-Whitney U test shows ~86% less MOSPD3 in the 45% layer containing the less motile spermatozoa compared to the 90% layer containing the more motile spermatozoa as shown in Figure 6-5.

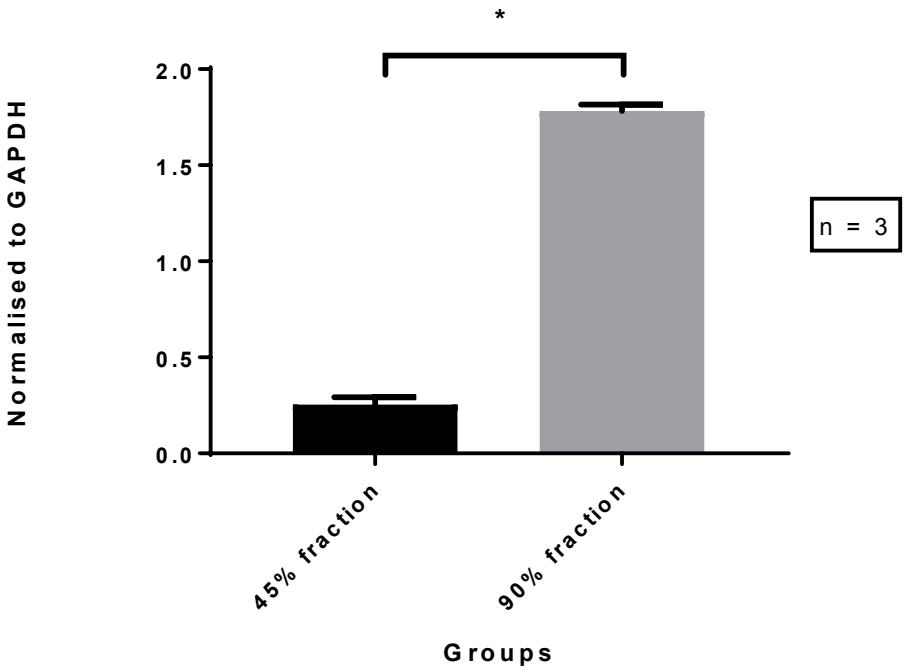


Figure 6-5: Statistical analysis of three Western Blot membranes coated against MOSPD3 and GAPDH. Relative amounts of MOSPD3 are shown, normalised against GAPDH. All values used for a paired one tailed Mann-Whitney U test (n=3) are based on densitometry using ImageJ. A ~86% decrease of MOSPD3 was visible comparing the motile 90% layer spermatozoal fraction with the less motile containing 45% spermatozoal fraction.

6.4.3 Immunocytochemistry

Expression of MOSD3 in motile (90% fraction) and less motile (45% fraction and 20% fraction) spermatozoa was examined by immunocytochemistry. Faint, punctate staining for MOSPD3 was detected in the spermatozoal tail in the less motile spermatozoa fractions (20% and 45%) shown in Figure 6-6 panel b) and c), whereas increased staining of the post acrosomal region and the spermatozoal tail (see arrows in Figure 6-6) was detected in motile spermatozoa (90% fraction) visible in Figure 6-6 panel a).

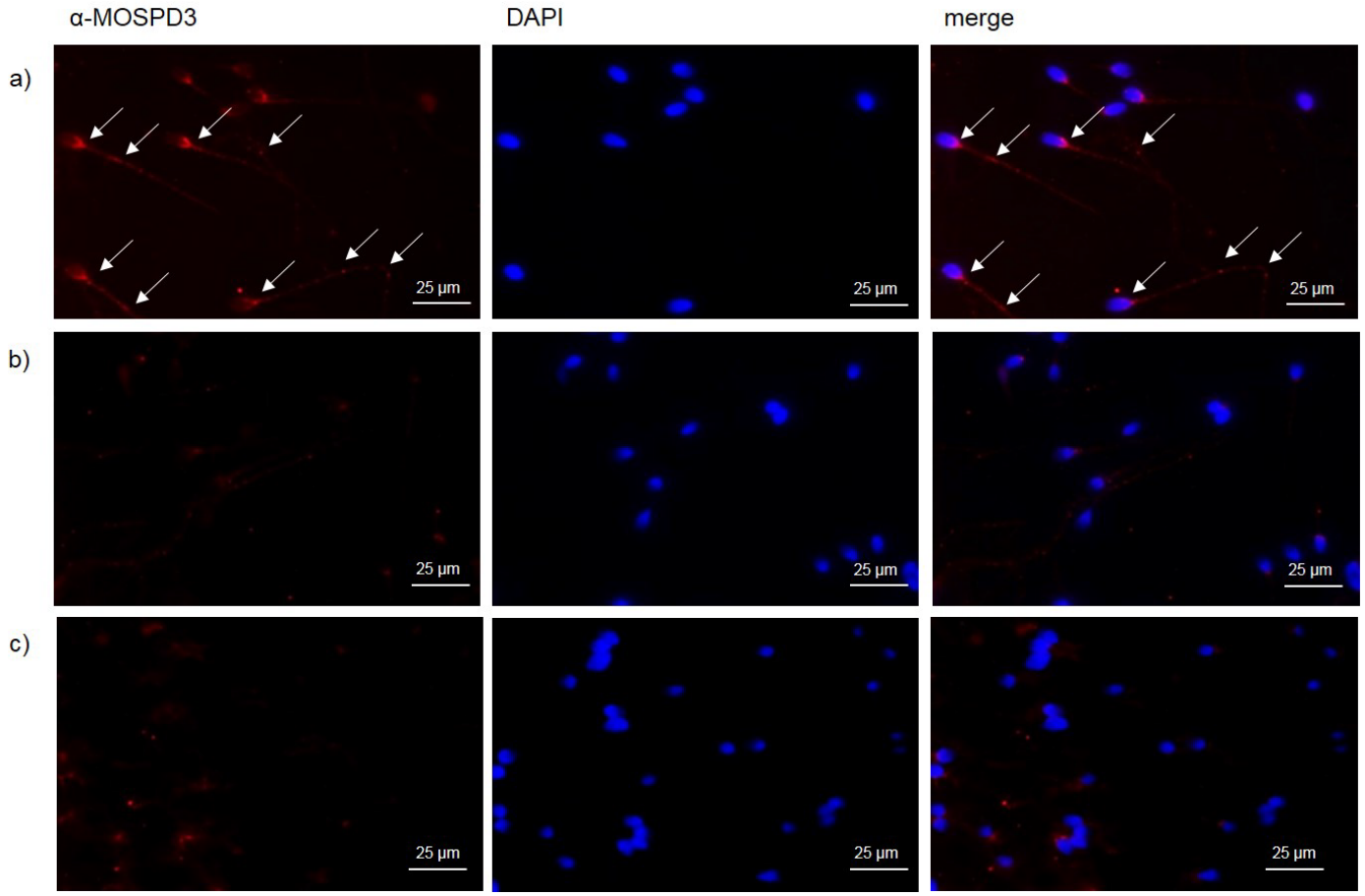


Figure 6-6: **Characterisation of MOSPD3 by immunofluorescence in less motile (45% fraction) and motile (90% fraction) spermatozoa.** Panel a) shows the 90% fraction of pelleted spermatozoa and a strong signal at the post-acrosomal region (arrows) and throughout the tail; whereas the signal in panel b) 45% layer fraction is almost undetectable and similar in panel c) the 20% fraction.

6.5 Discussion

The diagnostic potential of human semen RNA profiles has been explored using microarray and RNA-sequencing based approaches (Miller and Ostermeier, 2006a; Miller *et al.*, 1999; Moldenhauer *et al.*, 2003; Ostermeier *et al.*, 2002; Ostermeier *et al.*, 2005b; Zhao *et al.*, 2006; Garrido *et al.*, 2009; García-Herrero *et al.*, 2011; Platts *et al.*, 2007; Bonache *et al.*, 2012). However, although results to date have shown some promise in identifying particular RNAs or groups of RNAs that may be diagnostically useful, RNA markers as diagnostic tools for infertility have disadvantages when compared with proteins. Problems include the relative instability of extracted RNA compared with protein, the need for relatively expensive detection equipment and complex procedures for analysing and determining RNA levels at the picomolar level. For these reasons, the potential of spermatozoal RNA was explored to highlight possible protein markers of fertility, using available antibodies.

Guided by former RNA-Seq analysis of HA-enriched vs. original samples (see Chapter 5), one protein candidate (MOSPD3) in the enriched fraction was chosen for further consideration. MOSPD3 expression was greater in pelleted spermatozoa and the 90% motile fraction, whereas spermatozoa from the less motile fraction (45%) showed an ~84% decrease in expression. Different bands were visible in all MOSPD3 Western Blot samples which might be due to isoforms or possibly to non-specific binding. Western Blot experiments (Figure 6-4) were confirmed using immunocytochemistry (Figure 6-6). A strong post-acrosomal signal together with punctate signals along the tail in motile spermatozoa were observed. Less motile spermatozoa showed either no signal or a weaker signal in the post-acrosomal region together with weaker signals on the tail. Why the pelleted spermatozoa showed a stronger signal of MOSPD3 remains open to interpretation. The target protein may also have been digested or post-transcriptionally cleaved in less mature spermatozoa (Mahmood and Yang, 2012); however another possibility is that the more motile spermatozoa in pelleted fractions express more of the protein.

Little is known about the function of MOSPD3 in mammals and especially in mammalian spermaotozoa. The present data support the possibility that MOSPD3 is involved in mammalian spermatozoal motility. In addition, The present data confirms the location of MOSPD3 in spermatozoa seen by Li *et al.* (2014). This group used anti-MOSPD3 coated beads to bind to the post-acrosomal region and spermatozoal tail as a tool to recover

low amounts of spermatozoa from forensic samples contaminated by epithelial cells. Additionally, the present data suggests that the increased occurrence of MOSPD3 in motile spermatozoa compared with less motile spermatozoa could, with further research and development be incorporated into a clinical assay for assessing male fertility at the protein level. Study of MOSPD3 knock down models, to investigate whether the spermatozoa are immotile, not able to fertilise the oocyte and to address other unknown functions of MOSPD3 in spermatozoa are warranted considering the involvement of the nematode MSP in oocyte meiotic maturation and in ovarian muscle contraction (Kosinski *et al.*, 2005; Miller *et al.*, 2001). Studying spermatozoa of MOSPD3 knock down mice could be difficult, since Buerger (2010) showed high lethality of MOSPD3 depleted homozygote mice within the first day of birth, with survivors having a thinned right ventricular heart wall. No information about the fertility of these mouse strains is available to date. Another possibility would be to introduce siRNA of *Mospd3* to the spermatozoon or to create a conditional knockout mice line using CRISPR/CAS technology.

6.6 Conclusion

MOSPD3 was chosen on the basis of RNA-sequencing and subsequent interest in the proteins original identification in nematodes and its potential function in spermatozoa. It was expressed more highly in sperm from 90% fractions, where the more motile spermatozoa in a typical ejaculate are usually found. The present data supports the suggestion that spermatozoa binding to HA represent more viable populations and that the chosen protein is a potential marker of spermatozoa viability that could be developed into a diagnostic tool.

Chapter 7: General Discussion

The main objective of this work was to identify common, inter-species transcript 'expression' patterns in mammalian spermatozoa that could be indicative of shared expression networks for potential essential functional requirements in both mammalian spermatogenesis and post-fertilisation events. Secondary objectives included investigation of deleterious effects of freeze-thawing on the RNA profile of human spermatozoa and comparative RNA profiling study of hyaluronic acid selected and unselected human spermatozoa looking for differences indicative of spermatozoal quality.

The main design of this study was based on available NGS-RNA-sequencing approaches, made possible by rapid developments in the technology. NGS is now significantly cheaper, faster and more accurate than Sanger sequencing and is less limiting than microarray based analysis. The increased affordability makes it possible to address different biological and diagnostic questions including the identification of rare or low transcript levels and alternative splice variants without prior knowledge of the RNA itself (Metzker, 2010; Anton and Krawetz, 2012; Sultan *et al.*, 2008). Additionally, NGS sequencing allows simultaneous high resolution and simultaneous assessment of a high number of samples through multiplex arrangements, supported on all high throughput platforms (Anton and Krawetz, 2012). With the wide availability and sophistication of NGS technologies, the need for the right data analysis and for improvements in species' sequence annotation is growing. Recent developments have substantially reduced the critical minimum input mass of RNA, required for sequencing although all of the library construction techniques require some level of PCR based amplification (Hurd and Nelson, 2009). At the beginning of this study, nanograms of starting material were required for NGS experiments, which therefore placed spermatozoal RNA-seq work almost beyond reach. Each mature spermatozoon contains as little as ~12.5 fg (human); ~10 (bovine) and ~5 fg (porcine) total RNA (Card *et al.*, 2013). Complicating matters further, to ensure that results are derived from spermatozoa alone, pure populations are needed for every experiment, as one somatic cell contains 100-200 fold more RNA than a single spermatozoon and RNA introduced by somatic cell contamination of spermatozoa samples could significantly influence and bias the sequencing output (Cappallo-Obermann *et al.*, 2011; Jodar *et al.*, 2013; Goodrich *et al.*, 2013; Cappallo-Obermann and Spiess, 2016). A high number of reported human studies to date have made use of washed semen samples or only one density cushions (usually too low at

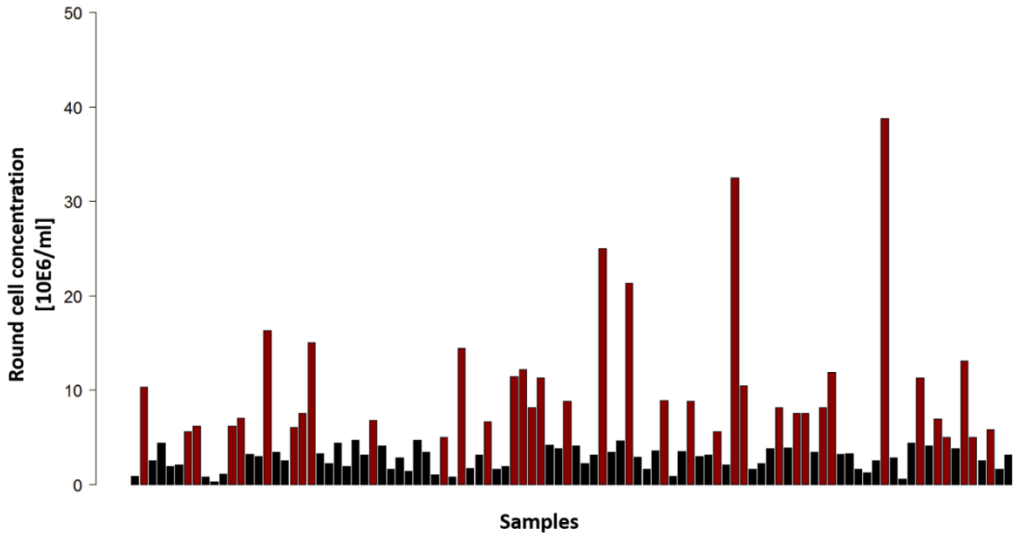


Figure 7-1: Semen samples and their round cell contamination. Samples with > $5x10^6$ /ml round cell contamination are labelled red. Figure adapted from Cappallo-Obermann and Spiess (2016).

density) for DGC, which resulted in somatic cell contamination (see Figure 7-1) (Jodar *et al.*, 2015; Georgiadis *et al.*, 2015; Pacheco *et al.*, 2011; Schuster *et al.*, 2016; Cappallo-Obermann and Spiess, 2016). Evidence for similar somatic cell contamination as reported by Cappallo-Obermann and Spiess (2016) was seen in preliminary studies in Leeds using one or two-step cushions and in swim up and HA-bound populations of human spermatozoa (Appendix VI). To obtain pure populations of spermatozoa and avoid somatic cell contamination, a triple gradient was employed for human and double gradients employed for bovine, ovine and porcine spermatozoa. This more stringent processing eliminated round cell contamination but in so doing, significantly reduced the number of spermatozoa available for RNA isolation. To extract sufficient RNA for any NGS input, large volumes of semen were required and since only frozen animal semen was available, less motile spermatozoa could be obtained. Despite the overall optimisation of the process to obtain a common approach for RNA extraction across all species used in this study, the RNA yield remained low and challenging, due to interspecies differences in spermatozoal resistance to chaotropic agents and chromatin condensation (Kempisty *et al.*, 2008; Das *et al.*, 2010; Goodrich *et al.*, 2013; Shafeeque *et al.*, 2014; Schuster *et al.*, 2016). Before sequencing, somatic cell and RNA quality controls were introduced. Giemsa and DAPI stainings were performed to confirm the absence of somatic cells and in addition all RNAs were run out on a Bioanalyzer. Spermatozoal rRNA is naturally degraded and an intact 18S and 28S rRNA profile is not seen (Cappallo-Obermann *et al.*, 2011). Only samples unambiguously cleared of somatic cells were used in the studies reported here. For all NGS data, UCSC traces

were produced and briefly analysed for the presence of somatic cell markers including CDH1, *GAPDH* and *CD45*. Either were rarely seen or were absent from the sequencing data. Georgiadis *et al.* (2015) reported finding intact 18S and 28S rRNA and only minor levels of degraded RNA and long RNAs using swim up spermatozoa, thus challenging the consensus that spermatozoal RNA is at least partially degraded prior to its isolation (Miller *et al.*, 1999; Ostermeier *et al.*, 2005a; Lalancette *et al.*, 2008a; Kempisty *et al.*, 2008; Bissonnette *et al.*, 2009; Das *et al.*, 2010; Cappallo-Obermann *et al.*, 2011; Sendler *et al.*, 2013).

Giemsa and DAPI stainings were routinely performed throughout the duration of this study (see examples in Appendix VII) and showed that human spermatozoal populations isolated by density gradient centrifugation, hyaluronan selection and even swim up often contained sufficient numbers of somatic cells to influence RNA data outcome. Furthermore, the RNA yields obtained using the swim up method were too small to create NGS-libraries. Georgiadis *et al.* (2015) compared a standard swim up method with density gradient centrifugation using 50% and 90% layers for the gradient centrifugation and processing at 37°C. In both approaches, RNA yields were spectacularly higher than ever reached during the course of the work reported in this thesis (or in any other study published to date). The authors insisted that minor somatic cell contamination did not influence their data outcome and that samples with 0.1% to 17% round cell contamination could be used. In contrast, our experience is that RNA arising from a 0.1% contamination with round cells is sufficient to imbalance the outcome of NGS experiments. Although the modified Trizol® procedure used in the work reported in this thesis may be relatively harsh and could lead to degradation of spermatozoal RNA, all solutions used for density gradient centrifugation were warmed up to 37°C before usage. Different extraction methods were tested and the methods chosen returned the best RNA yields and sufficient material for an NGS based approach. Bioanalyzer traces confirmed that spermatozoal RNAs are essentially free of intact 28S and 18S rRNA. Furthermore, as a standard in all experiments, gDNA was eliminated and in some experiments more than one DNase treatment was carried out and the absence of residual DNA confirmed with intron-spanning primers and a negative RT control. To conclude, yielding sufficient RNA through DGC has its limitations; however, the use of pure spermatozoal populations are necessary because even the smallest somatic cell contamination can bias the data outcome, which was also discussed in Cappallo-Obermann and Spiess (2016).

Mutual pathways in fertilisation and post-fertilisation events and embryo development were confirmed using the initial bioinformatic analyses. The analysis not only revealed 23 mutual transcripts in bovine, ovine, porcine and human, it additionally found many

transcripts involved in infertility and post-fertilisation events (Chapter 4) (Busso *et al.*, 2007; Jamsai and O'Bryan, 2010; Takemoto *et al.*, 2016; Yang *et al.*, 2012; Nasr-Esfahani *et al.*, 2004; Fu *et al.*, 2008; Zheng *et al.*, 2007; Shiyanov *et al.*, 1999; Yang *et al.*, 2013; Tsunematsu *et al.*, 2006; Dathe *et al.*, 2004; Chen *et al.*, 2006; Argasinska *et al.*, 2003; Branco *et al.*, 2016; Whitfield *et al.*, 2000). Three transcripts (*GTSF1L*, *SPATA3* and *FAM71F1*) were recently described as potential male fertility biomarkers and significantly downregulated in non-obstructive azoospermia (Malcher *et al.*, 2013). *GTSF1L* and *SPATA3* were found as common transcripts in all examined species, whereas *FAM71D* was found instead of *FAM71F1*, however, both transcripts are derived from the same family and are highly upregulated in testis (protein atlas). For all experiments, motile spermatozoa were used throughout and a comparison with less motile spermatozoa, including validations on a protein level could be considered to support the findings of Malcher *et al.* (2013). The role of *GTSF1L* in non-obstructive azoospermia is controversial. Malcher *et al.* (2013) found evidence of *GTSF1L* downregulation in azoospermic patients and *Gtsf1l* null mice are similarly described as infertile with additionally supressed retrotransposons observed in null mutant testes (Yoshimura *et al.*, 2009).

Another studied transcript, *CRISP2* was additionally found in all species under study and is localised to the spermatozoal acrosome and the outer dense fibres of the spermatozoal tail and may be involved in the acrosome reaction, gamete interaction and motility and is therefore a potential fertility biomarker (Busso *et al.*, 2005). The use of the ejaculate as a source of information into male reproductive health is justified by its non-invasive nature and by its potential to screen for infertility more rapidly and accurately.

Semen profiling at the molecular level may even be used as a tool for assessing environmental hazards affecting fertility as has been discussed in several publications (Garrido *et al.*, 2009; García-Herrero *et al.*, 2011; Wu *et al.*, 2012; Lima-Souza *et al.*, 2012; Anton and Krawetz, 2012; Malcher *et al.*, 2013; Jodar *et al.*, 2013; Kovac and Lamb, 2014).

The bioinformatics field has also been undergoing rapid changes in the past few years as the technology supporting it has advanced. The initial data sets generated in this study were re-processed several times using improved pipelines with more recent genome annotation files. This renewal of analysis was particularly important in view of the difficulties encountered during the study where four distinct species with widely differing annotation support information needed to be compared. Annotations as well as downstream analysis for human datasets are understandably more advanced as the main focus of the global research community is on human pathology. Different pathway analysis tools were tested to identify the tool with the best coverage for all species,

including DAVID, PANTHER, IPA and Cytoscape. First three, however, are highly anthropocentric and do not cover all the examined species. Therefore, Cytoscape was chosen as the resource of choice as it not only combines most available information and datasets for available networks, but additionally covers datasets for bovine, ovine, porcine and human in its analysis pipelines. Network analysis between species were comparable because they were each analysed using Cytoscape.
Hence, to achieve the widest coverage for all species, new annotations had to be built and re-analysis of the data led to an increase in the number of commonly 'expressed' transcripts from 23 (Chapter 4) to 122 (Chapter 3). Due to time limitations, the experiments in Chapter 4 could not be repeated and had to be considered as the first results from transcript analysis. As indicated, later pathway analysis expanded the number of common transcripts to 122 (Chapter 3), therefore available literature needs to be re-visited to identify potentially novel markers for pre- and post-fertilisation events to widen Chapter 4 for further studies to prove both the transmission of paternal transcripts to the oocyte and the fate of paternal transcripts in the developing embryo.

The most abundant common transcript for all species examined was 7SLRNA. In addition, this RNA was more significantly represented in spermatozoa than testis. The 7SLRNA is a part of the signal recognition particle ribonucleoprotein complex, consisting of 140 nucleotides that is responsible for protein trafficking within the cell and facilitates the secretion of proteins (Ullu *et al.*, 1982; Ullu and Tschudi, 1984; Nagai *et al.*, 2003). The 7SLRNA itself is responsible for the binding and release of the signal peptide including the release of the ribosome, allowing entry into another intracellular compartment or the extracellular space (Ullu and Tschudi, 1984). Once translation of a protein destined for transport to the endoplasmic reticulum (ER) initiates on the ribosome, the signal recognition particle is responsible for guiding the protein to the ER where folding of the nascent protein continues, followed by the transport of fully synthesised proteins to the Golgi apparatus. The Golgi apparatus is a major trafficking and protein directing point of the cell, additionally packing proteins into vesicles to direct them to their final destination either intra- or extracellular. In developing spermatids, the Golgi apparatus is located above the acrosome and vesicles are sent to the acrosome where they fuse with the acrosomal membrane until protein synthesis for the acrosome is completed. Following protein synthesis, the Golgi separates from the acrosome and is digested by autophagy (Moreno *et al.*, 2000). T-SNAREs, V-SNAREs VAMP/synaptobrevin and members of the rab family of small GTPase are found associated with the acrosome exchanging vesicles (exo- and endocytosis) with the Golgi apparatus, and it is assumed that these are involved in acrosomal biogenesis (Ramalho-Santos *et al.*, 2001). These SNARE complexes are additionally needed for the acrosome

reaction in human, with the acrosome building a large secretory complex (Yanagimachi, 1994; Tomes *et al.*, 2002). Prior to oocyte penetration, exocytosis of the acrosome occurs mediated by Ca^{2+}. The spermatozoon undergoes these changes to facilitate penetration of the zona pellucida via its digestion by enzymes released during acrosomal exocytosis (Wassarman and Place, 1999; Tomes *et al.*, 2002; Florman *et al.*, 2004; Harper *et al.*, 2008). Folllowing fusion with the oolemma, oocyte activation is achieved, by the activation of sperm phospholipase C (PLC), GTPases, transient receptor potential cation channels (TRPC) and PI3 kinase, initiating the Ca^{2+} influx and Phosphatidylinositol (3,4,5)-trisphosphate (PIP3) production (Tomes *et al.*, 2002; Florman *et al.*, 2004).

Acrosomal exocytosis differs from secretion in somatic cells as it leads to vesiculation and membrane loss. Tomes *et al.* (2002) speculated that the SNAREs and rab complexes are involved in fusion mechanisms for the outer acrosomal membrane with the plasma membrane although the precise mechanism is unclear to date (Tomes *et al.*, 2002). As 7SLRNAs are involved in membrane trafficking of proteins, it is possible that its abundance in spermatozoal RNA is due to the massive remodelling of spermatids during condensation that leads to spermatozoa production with the maturation of the acrosomal process. Paternal proteins may need to be guided to the right location after fertilisation through the SRP where they may be involved in the activation of the embryonic genome (Gundersen and Shapiro, 1984; Saunders *et al.*, 2002; Saunders *et al.*, 2007; Boerke *et al.*, 2007).

Since the primary structure of the 7SLRNA closely resembles an *Alu* sequence (140 nt), it has been suggested that the highly repetitive *Alu* elements were derived from 7SLRNA through reverse transcription (Ullu and Tschudi, 1984; Sinnett *et al.*, 1991). A major deletion in the 7SL specific sequence of *Alus* occurred in the distant past making them no longer dependent on 7SLRNA genes and giving them autonomy to evolve independently (Ullu and Tschudi, 1984). It has been speculated that 7SLDNA and ALU copies in the genome derived from transposition during the course of evolution (Turker, 1999).

Sciamanna *et al.* (2016) reviewed the function of a reverse transcriptase encoded by the LINE-1 element, suggesting 7SLRNAs may act as a template for reverse transcription. An earlier study suggested the possibility of reverse transcription of 7SLRNA, more precisely into B1 elements (Turker, 1999). Turker (1999), describes B1 elements as *cis*-acting elements, that function as an important part of the methylation centre in the mouse genome and may therefore be involved in epigenetic processes in spermatozoa and the progeny (Quentin, 1994; Morgan *et al.*, 1999; Rakyan *et al.*, 2003; Martínez *et al.*, 2014; Blewitt *et al.*, 2006). Past research has considered the possibility of spermatozoa being able to take up exogenous RNA or DNA molecules and reverse transcribe RNA into cDNA playing a role in the epigenome of the progeny (Giordano *et al.*, 2000; Sciamanna

et al., 2003). The ability of spermatozoa to adsorb and integrate external nucleic acids provides spermatozoa with a novel mechanism for generating biologically active retrogenes and potentially heritable epigenetic phenomena (Sciamanna et al., 2003). Studies by Sciamanna *et al.* (2003) showed evidence for a RT-dependent mechanism functioning in spermatozoa, controlling the origin of new retrogenes being later transferred to the embryo during fertilization and transmitted in a non-Mendelian pathway to the progeny (Sciamanna et al., 2003; Sciamanna et al., 2009).

Preliminary analysis for the activity of a reverse transcriptase in bovine spermatozoa was carried out during the course of this thesis, but yielded inconclusive results (data not shown). Additional studies have shown that a double-stranded endonuclease (DICER) is involved into the biogenesis of small RNAs derived from 7SLRNA (3.1% of total small RNAs), processing the 7SLRNA into 20 nt to 200 nt elements (Ren *et al.*, 2012). Later findings from the same group indicate the interference by 7SLRNA derived small RNAs in the formation of the SRP complex and therefore the inhibition of nascent protein formation (Ren *et al.*, 2013). Some studies suggested that 7SLRNA is involved in autophagy and cell cycle arrest by suppression of phospoprotein 53 (p53) and therefore may be responsible for protein degradation (Grammatikakis *et al.*, 2014). It has been shown that the *Alu* –like sequence of the *Bacillus subtilis* homologous SRP interacts with histone-like proteins (Nakamura *et al.*, 1999). Li *et al.* (2012) observed that histone methylated residues can inhibit the binding of SRPs and therefore inhibit the binding of histone effector proteins. Histone 4 (H4) was found to interact with two different parts of the SRP, SRP68 and SRP72. The capability to bind to the tail of Histone 4 may allow SRP68/72 heterodimers to bind chromatin with high affinity, involved in transcription activation activity (Li *et al.*, 2012). These effects may be important for histone protamine replacement and epigenetic processes.

Assisted reproductive techniques rely on visual or subjective assessment to select the best quality gamete for IVF treatment according to WHO guidelines to date and is often relying on freeze-thawing processes of the patient gametes (WHO, 2010). A relatively new approach is already being used in some clinics for spermatozoa selection and makes use of hyaluronic acid (HA) coated dishes to aid the crucial selection of the best candidate gamete for ICSI (Parmegiani et al., 2010b). HA mimics the natural environment of the oocyte the spermatozoon encounters and has to penetrate for fertilisation via interactions between spermatozoa HA-binding site and the HA-rich oocyte-cumulus complex (Huszar *et al.*, 2007). Immature spermatozoa are not able to bind efficiently to HA and as immature spermatozoa harbour higher levels of DNA fragmentation and chromosopmal aneuploidy, the use of HA-selected spermatozoa may lead to improvements in embryo viability (Cayli et al., 2003; Parmegiani et al., 2010b). In

this study the transcriptomes of HA-selected and unselected spermatozoa were compared by differential expression analysis. However, no significant differences were revealed. This was most likely due to their insufficient separation between the HA-binding and non-binding sperm in the original separation experiments (undertaken by an external laboratory), and confirmed by our own observations (Chapter 5). Ideally a comparison between HA-binding and non-binding spermatozoa should be performed to provide higher resolution. This could help to highlight differences between motile and less motile spermatozoa and could be developed into an HA-independent biomarker tool to detect infertility.

Additionally, further NGS work showed that liquid nitrogen freeze-thawing processed spermatozoa does not influence the large RNA content in humans. To improve spermatozoal selection for ART with as less DNA and RNA damage as possible HA-selection after thawing should be considered, especially since pregnancies initiated through ART have an increased likelihood of premature birth and health risks in the progeny (Hansen *et al.*, 2002; Kalra and Molinaro, 2008). These risk factors may be induced through epigenetic changes and stimulated through sncRNAs. No studies have investigated the sncRNA content of damage after freeze-thawing processes to this date, since the main focus of this study only included large RNAs (Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Bohacek *et al.*, 2015). Therefore important details of potential epigenetic control in early embryogenesis remain to be uncovered.

While the trends were the same, examining the average reads for *MOSPD3* using either analysis pipeline revealed a threefold lower expression in unselected versus HA-selected spermatozoa. While edge R correctly rejected this difference as being statistically significant due to the wide inter-sample variation in RNA levels for this gene, the fact that the protein has a Major Sperm Protein domain (MSP domain), originally identified in nematode sperm led us to take a closer look at its expression (Burke and Ward, 1983; Italiano *et al.*, 2001; Stewart *et al.*, 1994; Pall *et al.*, 2004; Tarr and Scott, 2004; Kara, 2012). Little, however, is known about the function of MOSPD3 protein in mammals, especially in relation to spermatozoal motility and therefore further investigations using Western Blot analysis and immunocytochemistry were performed. MOSPD3 was found upregulated in highly motile spermatozoa compared with less motile spermatozoa in both immunocytochemistry and Western Blot analysis. Hence, the evidence from these experiments suggests that MOSPD3 could play a role in human spermatozoal motility and could therefore be adapted for a clinical tool able to detect and diagnose infertility.

Hamazaki *et al.* (2015) examined highly upregulated promoter associated non-coding (panc) RNAs in the 2 cell-stage of mouse embryos. One of the transcripts targeted is

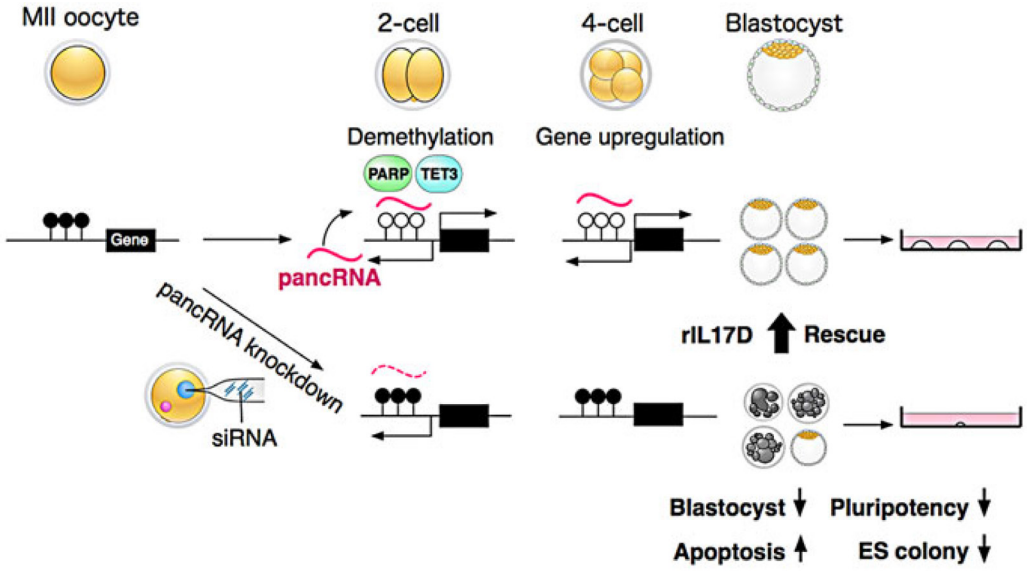


Figure 7-2: Epigenetic changes mediated through pancRNAs in the developing embryo. Figure adapted by Hamazaki *et al.* (2015).

Mospd3 and the promoter region for *Mospd3* in spermatozoa, the MII and 1 cell-stage embryo was shown to be highly methylated. Promoter methylation in 2 cell-stage embryos was decreased, indicating possible gene activation through epigenetic changes. Zygotic gene activation does not, however, start until the 2 cell-stage in the mouse embryo (Aoki *et al.*, 1997). In several species, the zygotic paternal genome is thought to mirror the genome of the fertilising spermatozoon and remains globally methylated before DNA replication (Kishigami *et al.*, 2006). The suppression of *pancMospd3* using siRNA led to a successful formation of most embryos to the late blastocyst stage (Figure 7-2). However, compared to the control group, most of the blastocysts did not hatch (Hamazaki *et al.*, 2015). Therefore, Hamazaki *et al.* (2015) concluded that *pancMospd3* may be involved in epigenetic reprogramming of promoter regions for gene activation and is involved in embryonic developmental processes.

Further studies showed that *Mospd3* knockout mice display neonatal lethality with defects of heart development; however as in Hamazaki *et al.* (2015) knockdown of *pancMospd3* did not cause any detectable developmental defects in blastocyst formation, but led to slower growth of ESC (Figure 7-2) (Pall *et al.*, 2004). Hamazaki *et al.* (2015) discussed the role of gene activation associated pancRNAs. pancRNAs may specify and establish the genomic position for an epigenetic change, paired with the activation of further yet unknown components, which are involved in genome wide methylation (Figure 7-2) (Branco *et al.*, 2012; Hajkova *et al.*, 2010). Additionally, Hamazaki *et al.* (2015) tried to identify factors involved in the activation of pncRNAs and found a strand specific enrichment of an asymmetric distributed CT-motif upstream of

pancRNA and mRNA- partnered genes (Hamazaki *et al.*, 2015). The role and involvement of this motif remains unclear to date. During preimplantation development, DNA methylation and histone change seem to overlap with the expression of pancRNAs, therefore epigenetic changes may be initiated through pancRNAs through affecting both and leading to gene silencing (Hamazaki *et al.*, 2015). Furthermore, it needs to be considered that paternal transferred sncRNAs may play a role and are involved in epigenetical mechanisms as seen in recent publications (Rodgers *et al.*, 2013; Dias and Ressler, 2014; Gapp *et al.*, 2014; Sharma *et al.*, 2016).

Outlook

Analysing RNA isolated from bovine, ovine, porcine and human spermatozoa, found 23 mutual transcripts in the initial analysis and 122 in a reanalysis of the data. Open source annotations were used for the first set of bioinformatical analysis and the number of mutual (shared) transcripts increased to 122 after performing the reanalysis using self-built annotations. Further bioinformatical and network analysis of these additional RNAs is required to identify functionality in pre- and post-fertilisation events. These could not be undertaken as part of this study due to time limitations. Femtograms of paternal RNAs may be delivered to the oocyte and their fate in the developing embryo should be revisited by more advanced NGS and TaqMan technologies.

Since the starting material input required has decreased, it should be possible to examine sncRNA in bovine, ovine, porcine and human to identify the role of mutual and conserved scnRNA in the developing embryo and in health risks which may be given to the offspring. Mutual transcripts may be knocked-out if possible or silenced in mice or bovine embryos to identify functionality.

Bibliography

Aboyoun, P., Pages, H., Lawrence, M., Aboyoun, P., Pages, H. and Lawrence, M. 2014. Representation and manipulation of genomic intervals.

Abraham, K. and Bhargava, P. 1963. Nucleic acid metabolism of mammalian spermatozoa. *Biochemical Journal.* 86(2), p298.

Abu-Halima, M., Backes, C., Leidinger, P., Keller, A., Lubbad, A.M., Hammadeh, M. and Meese, E. 2014. MicroRNA expression profiles in human testicular tissues of infertile men with different histopathologic patterns. *Fertility and sterility.* 101(1), pp.78-86. e72.

Ade, C., Roy-Engel, A.M. and Deininger, P.L. 2013. Alu elements: an intrinsic source of human genome instability. *Current opinion in virology.* 3(6), pp.639-645.

Adesnik, M., Salditt, M., Thomas, W.t. and Darnell, J. 1972. Evidence that all messenger RNA molecules (except histone messenger RNA) contain poly (A) sequences and that the poly (A) has a nuclear function. *Journal of molecular biology.* 71(1), pp.21-30.

Agarwal, A., Mulgund, A., Hamada, A. and Chyatte, M.R. 2015. A unique view on male infertility around the globe. *Reproductive Biology and Endocrinology.* 13(1), p1.

Ahmadi, A. and Ng, S.-C. 1999. Developmental capacity of damaged spermatozoa. *Human Reproduction.* 14(9), pp.2279-2285.

Aitken, R.J. and Baker, M.A. 2006. Oxidative stress, sperm survival and fertility control. *Molecular and cellular endocrinology.* 250(1), pp.66-69.

Alzohairy, A.M., Gyulai, G., Jansen, R.K. and Bahieldin, A. 2013. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid.* 69(1), pp.1-15.

Amanai, M., Brahmajosyula, M. and Perry, A.C.F. 2006. A Restricted Role for Sperm-Borne MicroRNAs in Mammalian Fertilization. *Biology of Reproduction.* 75(6), pp.877-884.

Ambion®. 2012. *Working with RNA: the basics. Available at:.* [Online].

Anderson, E.J. 2013. *Are Full-Length mRNA In Bos taurus Spermatozoa Transferred to the Oocyte During Fertilization? University of Rhode Island. Available at: http://digitalcommons.uri.edu/theses/84/*. thesis.

Anderson, L.M., Riffle, L., Wilson, R., Travlos, G.S., Lubomirski, M.S. and Alvord, W.G. 2006. Preconceptional fasting of fathers alters serum glucose in offspring of mice. *Nutrition.* 22(3), pp.327-331.

Andrews, S. 2007. *SeqMonk; Available at: http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/screenshots/.* [Online].

Andrews, S. 2010. *FastQC: A quality control tool for high throughput sequencing data. Package available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.* [Online].

Anger, J.T., Gilbert, B.R. and Goldstein, M. 2003. Cryopreservation of sperm: indications, methods and results. *The Journal of urology.* 170(4), pp.1079-1084.

Anton, E., Blanco, J., Egozcue, J. and Vidal, F. 2004. Sperm FISH studies in seven male carriers of Robertsonian translocation t (13; 14)(q10; q10). *Human Reproduction.* 19(6), pp.1345-1351.

Anton, E. and Krawetz, S.A. 2012. Spermatozoa as biomarkers for the assessment of human male infertility and genotoxicity. *Systems biology in reproductive medicine.* 58(1), pp.41-50.

Aoki, F., Worrad, D.M. and Schultz, R.M. 1997. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Developmental biology.* 181(2), pp.296-307.

Arangasamy, A., Kasimanickam, V., DeJarnette, J. and Kasimanickam, R. 2011. Association of CRISP2, CCT8, PEBP1 mRNA abundance in sperm and sire conception rate in Holstein bulls. *Theriogenology.* 76(3), pp.570-577.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. and Hannon, G.J. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science.* 316(5825), pp.744-747.

Argasinska, J., Rana, A.A., Gilchrist, M.J., Lachani, K., Young, A. and Smith, J.C. 2003. Loss of REEP4 causes paralysis of the Xenopus embryo. *International Journal of Developmental Biology.* 53(1), pp.37-43.

Askew, D.S. and Xu, F. 1999. New insights into the function of noncoding RNA and its potential role in disease pathogenesis. *Histology and histopathology.* 14(1), pp.235-241.

Avendaño, C., Franchi, A., Jones, E. and Oehninger, S. 2009. Pregnancy-specific β-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. *Human Reproduction.* 24(2), pp.270-277.

Babak, T., Blencowe, B.J. and Hughes, T.R. 2007. Considerations in the identification of functional RNA structural elements in genomic alignments. *BMC bioinformatics.* 8(1), p33.

Bailey, J., Buhr, M. and Robertson, L. 1994. Relationships among in vivo fertility, computer-analysed motility and in vitro Ca2+ flux in bovine spermatozoa. *Canadian Journal of Animal Science.* 74(1), pp.53-58.

Baker, A.M., Roberts, T.M. and Stewart, M. 2002. 2.6 Å resolution crystal structure of helices of the motile major sperm protein (MSP) of Caenorhabditis elegans. *Journal of molecular biology.* 319(2), pp.491-499.

Ball, G., Bellin, M., Ax, R. and First, N. 1982. Glycosaminoglycans in bovine cumulus-oocyte complexes: morphology and chemistry. *Molecular and cellular endocrinology.* 28(1), pp.113-122.

Bartel, D.P. 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell.* 136(2), pp.215-233.

Belan, E. 2013. LINEs of evidence: noncanonical DNA replication as an epigenetic determinant. *Biology direct.* 8(1), p1.

Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E. and Spadafora, C. 2006. Expression of LINE-1 retroposons is essential for murine preimplantation development. *Molecular reproduction and development.* 73(3), pp.279-287.

Berndston, W.E. 1977. Methods for quantifying mammalian spermatogenesis: a review. *Journal of Animal Science.* 44(5), pp.818-833.

Bhargava, P., Bishop, M. and Work, T. 1959. The chemical composition of bull semen with special reference to nucleic acids, free nucleotides and free amino acids. *Biochemical Journal.* 73(2), p242.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z. and Galon, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics.* 25(8), pp.1091-1093.

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

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H. and Whitelaw, E. 2006. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* 2(4), pe49.

Boerke, A., Dieleman, S.J. and Gadella, B.M. 2007. A possible role for sperm RNA in early embryo development. *Theriogenology.* 68, Supplement 1(0), pp.S147-S155.

Bohacek, J., Farinelli, M., Mirante, O., Steiner, G., Gapp, K., Coiret, G., Ebeling, M., Duran-Pacheco, G., Iniguez, A. and Manuella, F. 2015. Pathological brain

plasticity and cognition in the offspring of males subjected to postnatal traumatic stress. *Molecular psychiatry.* 20(5), pp.621-631.
Bonache, S., Mata, A., Ramos, M.D., Bassas, L. and Larriba, S. 2012. Sperm gene expression profile is related to pregnancy rate after insemination and is predictive of low fecundity in normozoospermic men. *Human reproduction.* 27(6), pp.1556-1567.
Borini, A., Tarozzi, N., Bizzaro, D., Bonu, M., Fava, L., Flamigni, C. and Coticchio, G. 2006. Sperm DNA fragmentation: paternal effect on early post-implantation embryo development in ART. *Human Reproduction.* 21(11), pp.2876-2881.
Bouhallier, F., Allioli, N., Lavial, F., Chalmel, F., Perrard, M.-H., Durand, P., Samarut, J., Pain, B. and Rouault, J.-P. 2010. Role of miR-34c microRNA in the late steps of spermatogenesis. *Rna.* 16(4), pp.720-731.
Brachet, J. and Chantrenne, H. 1956. The function of the nucleus in the synthesis of cytoplasmic proteins. In: *Cold Spring Harbor symposia on quantitative biology*: Cold Spring Harbor Laboratory Press, pp.329-337.
Branco, C.S., Garcez, M.E., Pasqualotto, F.F., Erdtman, B. and Salvador, M. 2010. Resveratrol and ascorbic acid prevent DNA damage induced by cryopreservation in human semen. *Cryobiology.* 60(2), pp.235-237.
Branco, M.R., Ficz, G. and Reik, W. 2012. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews Genetics.* 13(1), pp.7-13.
Branco, M.R., King, M., Perez-Garcia, V., Bogutz, A.B., Caley, M., Fineberg, E., Lefebvre, L., Cook, S.J., Dean, W. and Hemberger, M. 2016. Maternal DNA Methylation Regulates Early Trophoblast Development. *Developmental cell.* 36(2), pp.152-163.
Braun, R.E. 1998. Post–transcriptional control of gene expression during spermatogenesis. *Seminars in Cell & Developmental Biology.* 9(4), pp.483-489.
Braun, R.E. 2000. Temporal control of protein synthesis during spermatogenesis. *International journal of andrology.* 23(S2), pp.92-94.
Brenner, S., Jacob, F. and Meselson, M. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature.* 190(4776), pp.576-581.
Brink, R.A. 1956. A genetic change associated with the R locus in maize which is directed and potentially reversible. *Genetics.* 41(6), p872.
Brosius, J. and Tiedge, H. 1996. Reverse transcriptase: mediator of genomic plasticity. *Molecular Evolution of Viruses—Past and Present.* Springer, pp.91-107.
Buerger, K. 2010. *Functional analysis of the mospd gene family. University of Edinburgh. Available at: https://www.era.lib.ed.ac.uk/handle/1842/4435*. thesis.
Bujan, L. 1998. [Environment and spermatogenesis]. *Contraception, fertilite, sexualite (1992).* 26(1), pp.39-48.
Bullock, T.L., Roberts, T.M. and Stewart, M. 1996. 2.5 Å Resolution Crystal Structure of the Motile Major Sperm Protein (MSP) of Ascaris suum. *Journal of molecular biology.* 263(2), pp.284-296.
Burke, D.J. and Ward, S. 1983. Identification of a large multigene family encoding the major sperm protein of Caenorhabditis elegans. *Journal of molecular biology.* 171(1), pp.1-29.
Busso, D., Cohen, D., Hayashi, M., Kasahara, M. and Cuasnicu, P. 2005. Human testicular protein TPX1/CRISP-2: localization in spermatozoa, fate after capacitation and relevance for gamete interaction. *Molecular human reproduction.* 11(4), pp.299-305.
Busso, D., Goldweic, N.M., Hayashi, M., Kasahara, M. and Cuasnicú, P.S. 2007. Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion. *Biology of reproduction.* 76(4), pp.701-708.
Camenisch, T.D., Spicer, A.P., Brehm-Gibson, T., Biesterfeldt, J., Augustine, M.L., Calabro, A., Kubalak, S., Klewer, S.E. and McDonald, J.A. 2000. Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and

hyaluronan-mediated transformation of epithelium to mesenchyme. *The Journal of clinical investigation.* 106(3), pp.349-360.

Cappallo-Obermann, H., Schulze, W., Jastrow, H., Baukloh, V. and Spiess, A.-N. 2011. Highly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly. *Molecular Human Reproduction.* 17(11), pp.669-678.

Cappallo-Obermann, H. and Spiess, A.-N. 2016. Comment on “Absence of sperm RNA elements correlates with idiopathic male infertility”. *Science Translational Medicine.* 8(353), pp.353tc351-353tc351.

Card, C.J., Anderson, E.J., Zamberlan, S., Krieger, K.E., Kaproth, M. and Sartini, B.L. 2013. Cryopreserved Bovine Spermatozoal Transcript Profile as Revealed by High-Throughput Ribonucleic Acid Sequencing. *Biology of Reproduction.* 88(2), pp.49, 41-49.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P. and Werner, A. 2009. Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. *Nucleic Acids Research.* 37(7), pp.2274-2282.

Carlsen, E., Giwercman, A., Keiding, N. and Skakkebæk, N.E. 1992. Evidence for decreasing quality of semen during past 50 years. *British medical journal.* 305(6854), pp.609-613.

Carlson, M., Aboyoun, P. and Page, H. 2016. *An Introduction to the GenomicRanges Package. Package is available online at: https://bioconductor.org/packages/devel/bioc/vignettes/GenomicRanges/inst/doc/GenomicRangesIntroduction.pdf* [Online].

Carninci, P ., Kasukawa, T., Katayama, S., Gough, J., Frith, M., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B. and Wells, C. 2005. The transcriptional landscape of the mammalian genome. *Science.* 309(5740), pp.1559-1563.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H. and Zamore, P.D. 2010. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell.* 143(7), pp.1084-1096.

Carreau, S., Lambard, S., Said, L., Saad, A. and Galeraud-Denis, I. 2007. RNA dynamics of fertile and infertile spermatozoa. *Biochemical Society Transactions.* 35(3), pp.634-636.

Carrell, D.T. and Hammoud, S.S. 2010. The human sperm epigenome and its potential role in embryonic development. *Molecular Human Reproduction.* 16(1), pp.37-47.

Caspersson, T. 1941. Studien über den Eiweißumsatz der Zelle. *Die Naturwissenschaften.* Springer, pp.33-48.

Caspersson, T. and Schultz, J. 1939. Pentose nucleotides in the cytoplasm of growing tissues. *Nature.* 143(3623), pp.602-603.

Cayli, S., Jakab, A., Ovari, L., Delpiano, E., Celik-Ozenci, C., Sakkas, D., Ward, D. and Huszar, G. 2003. Biochemical markers of sperm function: male fertility and sperm selection for ICSI. *Reproductive biomedicine online.* 7(4), pp.462-468.

Chan, D., Delbès, G., Landry, M., Robaire, B. and Trasler, J.M. 2012. Epigenetic alterations in sperm DNA associated with testicular cancer treatment. *Toxicological Sciences.* 125(2), pp.532-543.

Chen, H.-J., Lin, C.-M., Lin, C.-S., Perez-Olle, R., Leung, C.L. and Liem, R.K. 2006. The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes & Development.* 20(14), pp.1933-1945.

Chen, H., Cheng, H. and Bjerknes, M. 1993. One-step Coomassie brilliant blue R-250 staining of proteins in polyacrylamide gel. *Analytical biochemistry.* 212(1), pp.295-296.

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G.-h., Peng, H., Zhang, X. and Zhang, Y. 2016. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science.* 351(6271), pp.397-400.

Chen, Y., McCarthy, D., Robinson, M. and K., S.G. 2014. *edgeR: differential expression analysis of digital gene expression data. Package is available online: https://bioconductor.org/packages/release/bioc/html/edgeR.html.* [Online].

Chiang, M., Steuerwald, N., Lambert, H., Main, E. and Steinleitner, A. 1994. Detection of human leukocyte antigen class I messenger ribonucleic acid transcripts in human spermatozoa via reverse transcription-polymerase chain reaction. *Fertility and Sterility.* 61(2), p276.

Chohan, K., Griffin, J. and Carrell, D. 2004. Evaluation of chromatin integrity in human sperm using acridine orange staining with different fixatives and after cryopreservation. *Andrologia.* 36(5), pp.321-326.

Chuang, J.C. and Jones, P.A. 2007. Epigenetics and MicroRNAs. *Pediatr Res.* 61(5 Part 2), pp.24R-29R.

Cohen, M.E. 1978. Brave New Baby and the Law: Fashioning Remedies for the Victims of In Vitro Fertilization, The. *Am. JL & Med.* 4, p319.

Cole, S.E., LaRiviere, F.J., Merrikh, C.N. and Moore, M.J. 2009. A convergence of rRNA and mRNA quality control pathways revealed by mechanistic analysis of nonfunctional rRNA decay. *Molecular cell.* 34(4), pp.440-450.

Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & development.* 11(21), pp.2755-2766.

Collins, J.E., Wali, N., Sealy, I.M., Morris, J.A., White, R.J., Leonard, S.R., Jackson, D.K., Jones, M.C., Smerdon, N.C. and Zamora, J. 2015. High-throughput and quantitative genome-wide messenger RNA sequencing for molecular phenotyping. *BMC genomics.* 16(1), p578.

Concepcion, C.P., Han, Y.-C., Mu, P., Bonetti, C., Yao, E., D'andrea, A., Vidigal, J.A., Maughan, W.P., Ogrodowski, P. and Ventura, A. 2012. Intact p53-dependent responses in miR-34–deficient mice. *PLoS Genet.* 8(7), pe1002797.

Conley, A.B., Miller, W.J. and Jordan, I.K. 2008. Human cis natural antisense transcripts initiated by transposable elements. *Trends in Genetics.* 24(2), pp.53-56.

Cox, L., Larman, M., Saunders, C., Hashimoto, K., Swann, K. and Lai, F. 2002. Sperm phospholipase Czeta from humans and cynomolgus monkeys triggers Ca2+ oscillations, activation and development of mouse oocytes. *Reproduction.* 124(5), pp.611-623.

Crick, F.H. 1958. On protein synthesis. In: *Symp Soc Exp Biol*, p.8.

Curry, E., Ellis, S. and Pratt, S. 2008. Detection of porcine sperm microRNAs using a heterologous microRNA microarray and reverse transcriptase polymerase chain reaction. *Molecular reproduction and development.* 76(3), pp.218-219.

Curry, E., Safranski, T.J. and Pratt, S.L. 2011. Differential expression of porcine sperm microRNAs and their association with sperm morphology and motility. *Theriogenology.* 76(8), pp.1532-1539.

Cuzin, F. and Rassoulzadegan, M. 2010. Non-Mendelian epigenetic heredity: gametic RNAs as epigenetic regulators and transgenerational signals. *Essays in biochemistry.* 48, pp.101-106.

Dadoune, J.-P. 2009. Spermatozoal RNAs: What about their functions? *Microscopy Research and Technique.* 72(8), pp.536-551.

Dandekar, P., Aggeler, J. and Talbot, P. 1992. Structure, distribution and composition of the extracellular matrix of human oocytes and cumulus masses. *Human Reproduction.* 7(3), pp.391-398.

Das, P.J., McCarthy, F., Vishnoi, M., Paria, N., Gresham, C., Li, G., Kachroo, P., Sudderth, A.K., Teague, S. and Love, C.C. 2013. Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-seq. *PLoS One.* 8(2).

Das, P.J., Paria, N., Gustafson-Seabury, A., Vishnoi, M., Chaki, S.P., Love, C.C., Varner, D.D., Chowdhary, B.P. and Raudsepp, T. 2010. Total RNA isolation from stallion sperm and testis biopsies. *Theriogenology.* 74(6), pp.1099-1106.e1092.

Dathe, V., Pröls, F. and Brand-Saberi, B. 2004. Expression of kinesin kif5c during chick development. *Anatomy and Embryology.* 207(6), pp.475-480.

Daxinger, L. and Whitelaw, E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet.* 13(3), pp.153-162.

De Kretser, D. 1997. Male infertility. *The Lancet.* 349(9054), pp.787-790.

De Loos, F., Van Beneden, T., Kruip, T. and Van Maurik, P. 1992. Structural aspects of bovine oocyte maturation in vitro. *Molecular reproduction and development.* 31(3), pp.208-214.

de Rooij, D. 2001. Proliferation and differentiation of spermatogonial stem cells. *Reproduction.* 121(3), pp.347-354.

De Sousa, P.A., Watson, A.J. and Schultz, R.M. 1998. Transient expression of a translation initiation factor is conservatively associated with embryonic gene activation in murine and bovine embryos. *Biology of reproduction.* 59(4), pp.969-977.

Delbes, G., Hales, B.F. and Robaire, B. 2007. Effects of the chemotherapy cocktail used to treat testicular cancer on sperm chromatin integrity. *Journal of andrology.* 28(2), p241.

Depa-Martynów, M., Kempisty, B., Lianeri, M., Jagodziński, P.P. and Jedrzejczak, P. 2007. Association between fertilin beta, protamines 1 and 2 and spermatid-specific linker histone H1-like protein mRNA levels, fertilization ability of human spermatozoa, and quality of preimplantation embryos. *Folia Histochem Cytobiol.* 45(Suppl 1), pp.S79-85.

Di Giammartino, Dafne C., Nishida, K. and Manley, James L. 2011. Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell.* 43(6), pp.853-866.

Dias, B.G. and Ressler, K.J. 2014. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci.* 17(1), pp.89-96.

Díez-Sánchez, C., Ruiz-Pesini, E., Montoya, J., Pérez-Martos, A., Enríquez, J.A. and López-Pérez, M.J. 2003a. Mitochondria from ejaculated human spermatozoa do not synthesize proteins. *FEBS Letters.* 553(1–2), pp.205-208.

Díez-Sánchez, C., Ruiz-Pesini, E., Montoya, J., Pérez-Martos, A., Enríquez, J.A. and López-Pérez, M.J. 2003b. Mitochondria from ejaculated human spermatozoa do not synthesize proteins. *FEBS letters.* 553(1-2), pp.205-208.

Doma, M.K. and Parker, R. 2006. Revenge of the NRD: Preferential Degradation of Nonfunctional Eukaryotic rRNA. *Developmental Cell.* 11(6), pp.757-758.

Donnelly, E.T., McClure, N. and Lewis, S.E. 2001a. Cryopreservation of human semen and prepared sperm: effects on motility parameters and DNA integrity. *Fertility and sterility.* 76(5), pp.892-900.

Donnelly, E.T., Steele, E.K., McClure, N. and Lewis, S.E.M. 2001b. Assessment of DNA integrity and morphology of ejaculated spermatozoa from fertile and infertile men before and after cryopreservation. *Human Reproduction.* 16(6), pp.1191-1199.

Drobnis, E.Z., Crowe, L.M., Berger, T., Anchordoguy, T.J., Overstreet, J.W. and Crowe, J.H. 1993. Cold shock damage is due to lipid phase transitions in cell membranes: a demonstration using sperm as a model. *Journal of Experimental Zoology.* 265(4), pp.432-437.

Du, Y., Wang, X., Wang, B., Chen, W., He, R., Zhang, L., Xing, X., Su, J., Wang, Y. and Zhang, Y. 2014. Deep sequencing analysis of microRNAs in bovine sperm. *Molecular reproduction and development.* 81(11), pp.1042-1052.

Eddy, E. and O'Brien, D. 1994. *The spermatozoon. In 'The Physiology of Reproduction'. 2nd Edn.(Eds E. Knobil and JD Neill.) pp. 29–77*. Raven Press: New York.

Eddy, S.R. 1999. Noncoding RNA genes. *Current opinion in genetics & development.* 9(6), pp.695-699.

Eddy, S.R. 2001. Non–coding RNA genes and the modern RNA world. *Nature Reviews Genetics.* 2(12), pp.919-929.

Eppig, J.J. 1979. FSH stimulates hyaluronic acid synthesis by oocyte–cumulus cell complexes from mouse preovulatory follicles.
Evenson, D.P. and Wixon, R. 2006. Clinical aspects of sperm DNA fragmentation detection and male infertility. *Theriogenology.* 65(5), pp.979-991.
Evers, J.L. 2002. Female subfertility. *The Lancet.* 360(9327), pp.151-159.
Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P. and Torres-Padilla, M.-E. 2013. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature structural & molecular biology.* 20(3), pp.332-338.
Fagerlind, M., Stålhammar, H., Olsson, B. and Klinga-Levan, K. 2015. Expression of miRNAs in Bull Spermatozoa Correlates with Fertility Rates. *Reproduction in Domestic Animals.* 50(4), pp.587-594.
Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., Laurent III, G.S., Kenny, P.J. and Wahlestedt, C. 2008. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nature medicine.* 14(7), pp.723-730.
Faghihi, M.A. and Wahlestedt, C. 2009. Regulatory roles of natural antisense transcripts. *Nature Reviews Molecular Cell Biology.* 10(9), pp.637-643.
Fang, P., Zeng, P., Wang, Z., Liu, M., Xu, W., Dai, J., Zhao, X., Zhang, D., Liang, D., Chen, X., Shi, S., Zhang, M., Wang, L., Qiao, Z. and Shi, H. 2014. Estimated Diversity of Messenger RNAs in Each Murine Spermatozoa and Their Potential Function During Early Zygotic Development. *Biology of Reproduction.*
Ferlin, A., Raicu, F., Gatta, V., Zuccarello, D., Palka, G. and Foresta, C. 2007. Male infertility: role of genetic background. *Reproductive biomedicine online.* 14(6), pp.734-745.
Feugang, J.M., Rodriguez-Osorio, N., Kaya, A., Wang, H., Page, G., Ostermeier, G.C., Topper, E.K. and Memili, E. 2010. Transcriptome analysis of bull spermatozoa: implications for male fertility. *Reproductive BioMedicine Online.* 21(3), pp.312-324.
Fischer, B.E., Wasbrough, E., Meadows, L.A., Randlet, O., Dorus, S., Karr, T.L. and Russell, S. 2012. Conserved properties of Drosophila and human spermatozoal mRNA repertoires. *Proceedings of the Royal Society of London B: Biological Sciences.* 279(1738), pp.2636-2644.
Florman, H.M., Jungnickel, M.K. and Sutton, K.A. 2004. Regulating the acrosome reaction. *International Journal of Developmental Biology.* 52(5-6), pp.503-510.
Florman, H.M., Jungnickel, M.K. and Sutton, K.A. 2007. What can we learn about fertilization from cystic fibrosis? *Proceedings of the National Academy of Sciences.* 104(27), pp.11123-11124.
Forrester, L.M., Nagy, A., Sam, M., Watt, A., Stevenson, L., Bernstein, A., Joyner, A.L. and Wurst, W. 1996. An induction gene trap screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro. *Proceedings of the National Academy of Sciences.* 93(4), pp.1677-1682.
Fraga, C.G., Motchnik, P.A., Wyrobek, A.J., Rempel, D.M. and Ames, B.N. 1996. Smoking and low antioxidant levels increase oxidative damage to sperm DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.* 351(2), pp.199-203.
Freiman, R.N. 2009. Specific variants of general transcription factors regulate germ cell development in diverse organisms. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms.* 1789(3), pp.161-166.
Friend, D.S. 1984. Membrane organization and differentiation in the guinea-pig spermatozoon. *Ultrastructure of Reproduction.* Springer, pp.75-85.
Fu, J.-J., Zeng, W.-M., Li, Y.-H. and Guo, M.-R. 2008. Establishment and characterization of inducible cell line with stable expression of Spata3. *Chin. J. Cell Biol.* 30, pp.247-250.
Funaro, M. and Paduch, D.A. 2014. Novel Markers of Male Infertility. *Human Fertility: Methods and Protocols.* pp.233-250.

Gadea, J., Sellés, E., Marco, M.A., Coy, P., Matás, C., Romar, R. and Ruiz, S. 2004. Decrease in glutathione content in boar sperm after cryopreservation: Effect of the addition of reduced glutathione to the freezing and thawing extenders. *Theriogenology.* 62(3), pp.690-701.

Galeraud-Denis, I., Lambard, S. and Carreau, S. 2007. Relationship between chromatin organization, mRNAs profile and human male gamete quality. *Asian journal of andrology.* 9(5), pp.587-592.

Gallie, D.R. 1991. The cap and poly (A) tail function synergistically to regulate mRNA translational efficiency. *Genes & development.* 5(11), pp.2108-2116.

Gandini, L., Lombardo, F., Paoli, D., Caponecchia, L., Familiari, G., Verlengia, C., Dondero, F. and Lenzi, A. 2000. Study of apoptotic DNA fragmentation in human spermatozoa. *Human Reproduction.* 15(4), pp.830-839.

Gandini, L., Lombardo, F., Salacone, P., Paoli, D., Anselmo, A.P., Culasso, F., Dondero, F. and Lenzi, A. 2003. Testicular cancer and Hodgkin's disease: evaluation of semen quality. *Human Reproduction.* 18(4), pp.796-801.

Ganguly, I., Gaur, G., Kumar, S., Mandal, D., Kumar, M., Singh, U., Kumar, S. and Sharma, A. 2013. Differential expression of protamine 1 and 2 genes in mature spermatozoa of normal and motility impaired semen producing crossbred Frieswal (HF× Sahiwal) bulls. *Research in veterinary science.* 94(2), pp.256-262.

Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J., Farinelli, L., Miska, E. and Mansuy, I.M. 2014. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci.* advance online publication.

García-Herrero, S., Garrido, N., Martínez-Conejero, J.A., Remohí, J., Pellicer, A. and Meseguer, M. 2011. Differential transcriptomic profile in spermatozoa achieving pregnancy or not via ICSI. *Reproductive BioMedicine Online.* 22(1), pp.25-36.

García-Herrero, S., Meseguer, M., Martínez-Conejero, J.A., Remohí, J., Pellicer, A. and Garrido, N. 2010. The transcriptome of spermatozoa used in homologous intrauterine insemination varies considerably between samples that achieve pregnancy and those that do not. *Fertility and Sterility.* 94(4), pp.1360-1373.

García-López, J., Alonso, L., Cárdenas, D.B., Artaza-Alvarez, H., de Dios Hourcade, J., Martínez, S., Brieño-Enríquez, M.A. and Del Mazo, J. 2015. Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *RNA.* 21(5), pp.946-962.

Garrido, N., García-Herrero, S. and Meseguer, M. 2013. Assessment of sperm using mRNA microarray technology. *Fertility and Sterility.* 99(4), pp.1008-1022.

Garrido, N., Martínez-Conejero, J.A., Jauregui, J., Horcajadas, J.A., Simón, C., Remohí, J. and Meseguer, M. 2009. Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome. *Fertility and Sterility.* 91(4, Supplement), pp.1307-1310.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y. and Gentry, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology.* 5(10), p1.

Georgiadis, A.P., Kishore, A., Zorrilla, M., Jaffe, T.M., Sanfilippo, J.S., Volk, E., Rajkovic, A. and Yatsenko, A.N. 2015. High Quality RNA in Semen and Sperm: Isolation, Analysis and Potential Application in Clinical Testing. *The Journal of Urology.* 193(1), pp.352-359.

Georgiou, I., Noutsopoulos, D., Dimitriadou, E., Markopoulos, G., Apergi, A., Lazaros, L., Vaxevanoglou, T., Pantos, K., Syrrou, M. and Tzavaras, T. 2009. Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes. *Human molecular genetics.* 18(7), pp.1221-1228.

Gilbert, I., Bissonnette, N., Boissonneault, G., Vallée, M. and Robert, C. 2007. A molecular analysis of the population of mRNA in bovine spermatozoa. *Reproduction.* 133(6), pp.1073-1086.

Giordano, R., Magnano, A.R., Zaccagnini, G., Pittoggi, C., Moscufo, N., Lorenzini, R. and Spadafora, C. 2000. Reverse Transcriptase Activity in Mature Spermatozoa of Mouse. *The Journal of Cell Biology.* 148(6), pp.1107-1114.

Giwercman, A. and Petersen, P.M. 2000. Cancer and male infertility. *Best Practice & Research Clinical Endocrinology & Metabolism.* 14(3), pp.453-471.

Goldstein, M. 2002. Surgical Management of Male infertility and other Scrotal disorder. Vol. I. *Campbell's urology, Patrick C. Walsh, Alan B Retik, Vaughan (eds).* 8, pp.313-316.

Goodrich, R., Johnson, G. and Krawetz, S.A. 2007. The Preparation of Human Spermatozoal RNA for Clinical Analysis. *Systems Biology in Reproductive Medicine.* 53(3), pp.161-167.

Goodrich, R.J., Anton, E. and Krawetz, S.A. 2013. Isolating mRNA and small noncoding RNAs from human sperm. *Spermatogenesis: Methods and Protocols.* pp.385-396.

Govindaraju, A., Uzun, A., Robertson, L., Atli, M.O., Kaya, A., Topper, E., Crate, E.A., Padbury, J., Perkins, A. and Memili, E. 2012. Dynamics of microRNAs in bull spermatozoa. *Reprod Biol Endocrinol.* 10, p82.

Grammatikakis, I., Panda, A.C., Abdelmohsen, K. and Gorospe, M. 2014. Long noncoding RNAs (lncRNAs) and the molecular hallmarks of aging. *Aging.* 6(12), pp.992-1009.

Grandjean, V., Fourré, S., De Abreu, D.A.F., Derieppe, M.-A., Remy, J.-J. and Rassoulzadegan, M. 2015. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Scientific reports.* 5.

Grootegoed, J.A., Siep, M. and Baarends, W.M. 2000. Molecular and cellular mechanisms in spermatogenesis. *Best Practice & Research Clinical Endocrinology & Metabolism.* 14(3), pp.331-343.

Grossfeld, R., Sieg, B., Struckmann, C., Frenzel, A., Maxwell, W. and Rath, D. 2008. New aspects of boar semen freezing strategies. *Theriogenology.* 70(8), pp.1225-1233.

Grunewald, S., Paasch, U., Glander, H.J. and Anderegg, U. 2005. Mature human spermatozoa do not transcribe novel RNA. *Andrologia.* 37(2-3), pp.69-71.

Gundersen, G.G. and Shapiro, B.M. 1984. Sperm surface proteins persist after fertilization. *The Journal of cell biology.* 99(4), pp.1343-1353.

Gur, Y. and Breitbart, H. 2006. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. *Genes & Development.* 20(4), pp.411-416.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P. and Surani, M.A. 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science.* 329(5987), pp.78-82.

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

Hamazaki, N., Uesaka, M., Nakashima, K., Agata, K. and Imamura, T. 2015. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. *Development.* 142(5), pp.910-920.

Hammerstedt, R., Graham, J.K. and Nolan, J.P. 1990. Cryopreservation of mammalian sperm: what we ask them to survive. *J Androl.* 11(1), pp.73-88.

Hammoud, S.S., Low, D.H., Yi, C., Carrell, D.T., Guccione, E. and Cairns, B.R. 2014. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell.* 15(2), pp.239-253.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T. and Cairns, B.R. 2009. Distinctive chromatin in human sperm packages genes for embryo development. *Nature.* 460(7254), pp.473-478.

Hansen, K.D., Brenner, S.E. and Dudoit, S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research.* 38(12), pp.e131-e131.

Hansen, M., Kurinczuk, J.J., Bower, C. and Webb, S. 2002. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *New England Journal of Medicine.* 346(10), pp.725-730.

Harakalova, M., Nijman, I.J., Medic, J., Mokry, M., Renkens, I., Blankensteijn, J.D., Kloosterman, W., Baas, A.F. and Cuppen, E. 2011. Genomic DNA pooling strategy for next-generation sequencing-based rare variant discovery in abdominal aortic aneurysm regions of interest—challenges and limitations. *Journal of cardiovascular translational research.* 4(3), pp.271-280.

Harper, C.V., Cummerson, J.A., White, M.R., Publicover, S.J. and Johnson, P.M. 2008. Dynamic resolution of acrosomal exocytosis in human sperm. *Journal of cell science.* 121(13), pp.2130-2135.

Harrison, K., Sherrin, D., Hawthorne, T., Breen, T., West, G. and Wilson, L. 1990. Embryotoxicity of micropore filters used in liquid sterilization. *Journal of in vitro fertilization and embryo transfer.* 7(6), pp.347-350.

Hascall, V.C., Majors, A.K., de la Motte, C.A., Evanko, S.P., Wang, A., Drazba, J.A., Strong, S.A. and Wight, T.N. 2004. Intracellular hyaluronan: a new frontier for inflammation? *Biochimica et Biophysica Acta (BBA)-General Subjects.* 1673(1), pp.3-12.

He, S., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R. and Hugenholtz, P. 2010. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat Meth.* 7(10), pp.807-812.

Hecht, N.B. and Williams, J.L. 1978. Synthesis of RNA by Separated Heads and Talls from Bovine Spermatozoa. *Biology of Reproduction.* 19(3), pp.573-579.

Heller, C.G. and Clermont, Y. 1963. Spermatogenesis in Man: An Estimate of Its Duration. *Science.* 140(3563), pp.184-186.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T. and Ohta, K. 2008. Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature.* 456(7218), pp.130-134.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R. and Zamir, A. 1965. Structure of a ribonucleic acid. *Science.* 147(3664), pp.1462-1465.

Holstein, A.-F., Schulze, W. and Davidoff, M. 2003. Understanding spermatogenesis is a prerequisite for treatment. *Reprod Biol Endocrinol.* 1(107), pb14.

Holt, W. 1984. Membrane heterogeneity in the mammalian spermatozoon. *International review of cytology.* 87, pp.159-194.

Holt, W., Head, M. and North, R. 1992. Freeze-induced membrane damage in ram spermatozoa is manifested after thawing: observations with experimental cryomicroscopy. *Biology of reproduction.* 46(6), pp.1086-1094.

Holt, W.V. 2000. Basic aspects of frozen storage of semen. *Animal Reproduction Science.* 62(1–3), pp.3-22.

Hosken, D.J. and Hodgson, D.J. 2014. Why do sperm carry RNA? Relatedness, conflict, and control. *Trends in ecology & evolution.* 29(8), pp.451-455.

Huang, D.W., Sherman, B.T. and Lempicki, R.A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols.* 4(1), pp.44-57.

Hurd, P.J. and Nelson, C.J. 2009. Advantages of next-generation sequencing versus the microarray in epigenetic research. *Briefings in Functional Genomics & Proteomics.* 8(3), pp.174-183.

Huszar, G., Jakab, A., Sakkas, D., Ozenci, C.-C., Cayli, S., Delpiano, E. and Ozkavukcu, S. 2007. Fertility testing and ICSI sperm selection by hyaluronic acid binding: clinical and genetic aspects. *Reproductive biomedicine online.* 14(5), pp.650-663.

Huszar, G., Ozenci, C.C., Cayli, S., Zavaczki, Z., Hansch, E. and Vigue, L. 2003. Hyaluronic acid binding by human sperm indicates cellular maturity, viability, and unreacted acrosomal status. *Fertility and Sterility.* 79, Supplement 3, pp.1616-1624.

Huszar, G. and Vigue, L. 1990. Spermatogenesis-related change in the synthesis of the creatine kinase B-type and M-type isoforms in human spermatozoa. *Molecular reproduction and development.* 25(3), pp.258-262.

Ing, N., Forrest, D., Love, C. and Varner, D. 2014. Dense spermatozoa in stallion ejaculates contain lower concentrations of mRNAs encoding the sperm specific calcium channel 1, ornithine decarboxylase antizyme 3, aromatase, and estrogen receptor alpha than less dense spermatozoa. *Theriogenology.* 82(2), pp.347-353.

Inoue, A., Matoba, S. and Zhang, Y. 2012. Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. *Cell research.* 22(12), pp.1640-1649.

Italiano, J.E., Roberts, T.M., Stewart, M. and Fontana, C.A. 1996. Reconstitution in vitro of the motile apparatus from the amoeboid sperm of Ascaris shows that filament assembly and bundling move membranes. *Cell.* 84(1), pp.105-114.

Italiano, J.E., Stewart, M. and Roberts, T.M. 1999. Localized Depolymerization of the Major Sperm Protein Cytoskeleton Correlates with the Forward Movement of the Cell Body in the Amoeboid Movement of Nematode Sperm. *The Journal of Cell Biology.* 146(5), pp.1087-1096.

Italiano, J.E., Stewart, M. and Roberts, T.M. 2001. How the assembly dynamics of the nematode major sperm protein generate amoeboid cell motility. *International review of cytology.* 202, pp.1-34.

Jakab, A., Sakkas, D., Delpiano, E., Cayli, S., Kovanci, E., Ward, D., Ravelli, A. and Huszar, G. 2005. Intracytoplasmic sperm injection: a novel selection method for sperm with normal frequency of chromosomal aneuploidies. *Fertility and sterility.* 84(6), pp.1665-1673.

Jamsai, D. and O'Bryan, M.K. 2010. Mouse models in male fertility research. *Asian journal of andrology.* 13(1), pp.139-151.

Jamsai, D., Reilly, A., Smith, S., Gibbs, G., Baker, H., McLachlan, R., de Kretser, D. and O'Bryan, M. 2008. Polymorphisms in the human cysteine-rich secretory protein 2 (CRISP2) gene in Australian men. *Human reproduction.* 23(9), pp.2151-2159.

Jensen, T.K., Carlsen, E., Jørgensen, N., Berthelsen, J.G., Keiding, N., Christensen, K., Petersen, J.H., Knudsen, L.B. and Skakkebæk, N.E. 2002. Poor semen quality may contribute to recent decline in fertility rates. *Human Reproduction.* 17(6), pp.1437-1440.

Jensen, T.K., Jacobsen, R., Christensen, K., Nielsen, N.C. and Bostofte, E. 2009. Good semen quality and life expectancy: a cohort study of 43,277 men. *American journal of epidemiology.* 170(5), pp.559-565.

Jimenez-Chillaron, J.C., Ramon-Krauel, M., Ribo, S. and Diaz, R. 2016. Transgenerational epigenetic inheritance of diabetes risk as a consequence of early nutritional imbalances. *Proceedings of the Nutrition Society.* 75(01), pp.78-89.

Jodar, M., Kalko, S., Castillo, J., Ballescà, J.L. and Oliva, R. 2012. Differential RNAs in the sperm cells of asthenozoospermic patients. *Human reproduction.* 27(5), pp.1431-1438.

Jodar, M., Selvaraju, S., Sendler, E., Diamond, M.P., Krawetz, S.A. and Network, f.t.R.M. 2013. The presence, role and clinical use of spermatozoal RNAs. *Human Reproduction Update.* 19(6), pp.604-624.

Jodar, M., Sendler, E. and Krawetz, S.A. 2016. The protein and transcript profiles of human semen. *Cell and tissue research.* 363(1), pp.85-96.

Jodar, M., Sendler, E., Moskovtsev, S.I., Librach, C.L., Goodrich, R., Swanson, S., Hauser, R., Diamond, M.P. and Krawetz, S.A. 2015. Absence of sperm RNA elements correlates with idiopathic male infertility. *Science translational medicine.* 7(295), pp.295re296-295re296.

Johnson, G., Sendler, E., Lalancette, C., Hauser, R., Diamond, M. and Krawetz, S. 2011. Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Molecular Human Reproduction.* 17(12), pp.721-726.

Johnson, G.D., Mackie, P., Jodar, M., Moskovtsev, S. and Krawetz, S.A. 2015. Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic Acids Research.*

Johnson, L. 1986. Spermatogenesis and aging in the human. *Journal of andrology.* 7(6), p331.

Jungwirth, A., Giwercman, A., Tournaye, H., Diemer, T., Kopa, Z., Dohle, G., Krausz, C. and Infertility, E.W.G.o.M. 2012. European Association of Urology guidelines on Male Infertility: the 2012 update. *European urology.* 62(2), pp.324-332.

Kagiwada, S., Hosaka, K., Murata, M., Nikawa, J.-i. and Takatsuki, A. 1998. The Saccharomyces cerevisiae SCS2 gene product, a homolog of a synaptobrevin-associated protein, is an integral membrane protein of the endoplasmic reticulum and is required for inositol metabolism. *Journal of bacteriology.* 180(7), pp.1700-1708.

Kalra, S.K. and Molinaro, T.A. 2008. The association of in vitro fertilization and perinatal morbidity. In: *Seminars in reproductive medicine*: © Thieme Medical Publishers, pp.423-435.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J. and Hofacker, I.L. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science.* 316(5830), pp.1484-1488.

Kara, M. 2012. *Characterisation and functional analysis of a novel msp domain–containing protein, MOSPD1. University of Edinburgh. Available at: https://www.era.lib.ed.ac.uk/handle/1842/6475*. thesis.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C., Suzuki, M. and Kawai, J. 2005. Antisense transcription in the mammalian transcriptome. *Science.* 309(5740), pp.1564-1566.

Kawano, M., Kawaji, H., Grandjean, V., Kiani, J. and Rassoulzadegan, M. 2012. Novel Small Noncoding RNAs in Mouse Spermatozoa, Zygotes and Early Embryos. *PLoS One.* 7(9), pe44542.

Kempisty, B., Antosik, P., Bukowska, D., Jackowska, M., Lianeri, M., Jaśkowski, J.M. and Jagodziński, P.P. 2008. Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. *Reproduction, Fertility and Development.* 20(4), pp.513-518.

Kempisty, B., Depa-Martynow, M., Lianeri, M., Jedrzejczak, P., Darul-Wasowicz, A. and Jagodzinski, P.P. 2007. Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men. *Folia Histochem Cytobiol.* 45(Suppl 1), pp.S109-S113.

Khatri, P., Sirota, M. and Butte, A.J. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol.* 8(2), pe1002375.

Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R. and Frank, W. 2010. Transcriptional control of gene expression by microRNAs. *Cell.* 140(1), pp.111-122.

Kigami, D., Minami, N., Takayama, H. and Imai, H. 2003. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biology of reproduction.* 68(2), pp.651-654.

Kim, V.N. 2006. Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes & Development.* 20(15), pp.1993-1997.

Kindler, S., Wang, H., Richter, D. and Tiedge, H. 2005. RNA transport and local control of translation. *Annual review of cell and developmental biology.* 21, p223.

King, K.L., Stewart, M. and Roberts, T.M. 1994. Supramolecular assemblies of the Ascaris suum major sperm protein (MSP) associated with amoeboid cell motility. *Journal of Cell Science.* 107(10), pp.2941-2949.

Kishigami, S., Van Thuan, N., Hikichi, T., Ohta, H., Wakayama, S., Mizutani, E. and Wakayama, T. 2006. Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. *Developmental Biology.* 289(1), pp.195-205.

Kleene, K.C. 2005. Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. *Developmental biology.* 277(1), pp.16-26.
Kocabas, A.M., Crosby, J., Ross, P.J., Otu, H.H., Beyhan, Z., Can, H., Tam, W.-L., Rosa, G.J.M., Halgren, R.G., Lim, B., Fernandez, E. and Cibelli, J.B. 2006. The transcriptome of human oocytes. *Proceedings of the National Academy of Sciences.* 103(38), pp.14027-14032.
Kosinski, M., McDonald, K., Schwartz, J., Yamamoto, I. and Greenstein, D. 2005. C. elegans sperm bud vesicles to deliver a meiotic maturation signal to distant oocytes. *Development.* 132(15), pp.3357-3369.
Kovac, J.R. and Lamb, D.J. 2014. Male infertility biomarkers and genomic aberrations in azoospermia. *Fertility and Sterility.* 101(5), pe31.
Kovanci, E., Kovacs, T., Moretti, E., Vigue, L., Bray-Ward, P., Ward, D.C. and Huszar, G. 2001. FISH assessment of aneuploidy frequencies in mature and immature human spermatozoa classified by the absence or presence of cytoplasmic retention. *Human Reproduction.* 16(6), pp.1209-1217.
Krawetz, S.A. 2005. Paternal contribution: new insights and future challenges. *Nat Rev Genet.* 6(8), pp.633-642.
Krawetz, S.A., Kruger, A., Lalancette, C., Tagett, R., Anton, E., Draghici, S. and Diamond, M.P. 2011. A survey of small RNAs in human sperm. *Human Reproduction.* 26(12), pp.3401-3412.
Kretser, D. 1989. Illustrated Pathology of Human Spermatogenesis AF Holstein, EC Roosen-Runge, C. Schirren. *Andrologia.* 21(3), pp.197-197.
Krueger, F. 2013. *Trim Galore! a tool to apply quality and adamter trimming to FastQ files for high throughput sequencing data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.* [Online].
Kumar, G., Patel, D. and Naz, R. 1993. c-MYC mRNA is present in human sperm cells. *Cellular & molecular biology research.* 39(2), p111.
Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature.* 227, pp.680-685.
Lai, Y.M., Yang, F.-P. and Pao, C. 1996. Human papillomavirus deoxyribonucleic acid and ribonucleic acid in seminal plasma and sperm cells. *Fertility and Sterility.* 65(5), p1026.
Lalancette, C., Miller, D., Li, Y. and Krawetz, S.A. 2008a. Paternal contributions: New functional insights for spermatozoal RNA. *Journal of Cellular Biochemistry.* 104(5), pp.1570-1579.
Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T. and Krawetz, S.A. 2009. Identification of human sperm transcripts as candidate markers of male fertility. *Journal of Molecular Medicine.* 87(7), pp.735-748.
Lalancette, C., Thibault, C., Bachand, I., Caron, N. and Bissonnette, N. 2008b. Transcriptome Analysis of Bull Semen with Extreme Nonreturn Rate: Use of Suppression-Subtractive Hybridization to Identify Functional Markers for Fertility. *Biology of Reproduction.* 78(4), pp.618-635.
Lamarck, J.-B.-P. 1809. *Philosophie zoologique.*
Lambard, S., Galeraud-Denis, I., Martin, G., Levy, R., Chocat, A. and Carreau, S. 2004. Analysis and significance of mRNA in human ejaculated sperm from normozoospermic donors: relationship to sperm motility and capacitation. *Molecular Human Reproduction.* 10(7), pp.535-541.
Lapidot, M. and Pilpel, Y. 2006. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO reports.* 7(12), pp.1216-1222.
LaRiviere, F.J., Cole, S.E., Ferullo, D.J. and Moore, M.J. 2006. A late-acting quality control process for mature eukaryotic rRNAs. *Molecular cell.* 24(4), pp.619-626.
Laurent, F., Labesse, G. and de Wit, P. 2000. Molecular cloning and partial characterization of a plant VAP33 homologue with a major sperm protein domain. *Biochemical and biophysical research communications.* 270(1), pp.286-292.

Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. and Casari, G. 2004. In search of antisense. *Trends in Biochemical Sciences.* 29(2), pp.88-94.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 9(8), pe1003118.

Lee, Y.S., Shibata, Y., Malhotra, A. and Dutta, A. 2009. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & Development.* 23(22), pp.2639-2649.

Lefièvre, L., Bedu-Addo, K., Conner, S.J., Machado-Oliveira, G.S., Chen, Y., Kirkman-Brown, J.C., Afnan, M.A., Publicover, S.J., Ford, W.C.L. and Barratt, C.L. 2007. Counting sperm does not add up any more: time for a new equation? *Reproduction.* 133(4), pp.675-684.

Lewis, S. and Aitken, R. 2005. DNA damage to spermatozoa has impacts on fertilization and pregnancy. *Cell and tissue research.* 322(1), pp.33-41.

Li, J., Zhou, F., Zhan, D., Gao, Q., Cui, N., Li, J., Iakhiaeva, E., Zwieb, C., Lin, B. and Wong, J. 2012. A novel histone H4 arginine 3 methylation-sensitive histone H4 binding activity and transcriptional regulatory function for signal recognition particle subunits SRP68 and SRP72. *Journal of Biological Chemistry.* 287(48), pp.40641-40651.

Li, L., Lu, X. and Dean, J. 2013. The maternal to zygotic transition in mammals. *Molecular aspects of medicine.* 34(5), pp.919-938.

Li, T., Vu, T.H., Lee, K.-O., Yang, Y., Nguyen, C.V., Bui, H.Q., Zeng, Z.-L., Nguyen, B.T., Hu, J.-F. and Murphy, S.K. 2002. An imprinted PEG1/MEST antisense expressed predominantly in human testis and in mature spermatozoa. *Journal of Biological Chemistry.* 277(16), pp.13518-13527.

Li, X.-B., Wang, Q.-S., Feng, Y., Ning, S.-H., Miao, Y.-Y., Wang, Y.-Q. and Li, H.-W. 2014. Magnetic bead-based separation of sperm from buccal epithelial cells using a monoclonal antibody against MOSPD3. *International journal of legal medicine.* 128(6), pp.905-911.

Li, Z., Lin, Q., Liu, R., Xiao, W. and Liu, W. 2010. Protective effects of ascorbate and catalase on human spermatozoa during cryopreservation. *Journal of andrology.* 31(5), pp.437-444.

Lian, J., Zhang, X., Tian, H., Liang, N., Wang, Y., Liang, C., Li, X. and Sun, F. 2009. Altered microRNA expression in patients with non-obstructive azoospermia. *Reproductive Biology and Endocrinology.* 7(1), p13.

Liang, X., Zhou, D., Wei, C., Luo, H., Liu, J., Fu, R. and Cui, S. 2012. MicroRNA-34c enhances murine male germ cell apoptosis through targeting ATF1. *PLoS One.* 7(3), pe33861.

Liao, Y., Smyth, G.K. and Shi, W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research.* 41(10), pp.e108-e108.

Lima-Souza, A., Anton, E., Mao, S., Ho, W.J. and Krawetz, S.A. 2012. A platform for evaluating sperm RNA biomarkers: dysplasia of the fibrous sheath—testing the concept. *Fertility and Sterility.* 97(5), pp.1061-1066.e1063.

Lin, D.S., Connor, W.E., Wolf, D.P., Neuringer, M. and Hachey, D.L. 1993. Unique lipids of primate spermatozoa: desmosterol and docosahexaenoic acid. *Journal of Lipid Research.* 34(3), pp.491-499.

Linschooten, J.O., Laubenthal, J., Cemeli, E., Baumgartner, A., Anderson, D., Sipinen, V.E., Brunborg, G., Haenen, G.R.M.M., Fthenou, E., Briedé, J.J., van Schooten, F.J. and Godschalk, R.W.L. 2011. Incomplete protection of genetic integrity of mature spermatozoa against oxidative stress. *Reproductive Toxicology.* 32(1), pp.106-111.

Linschooten, J.O., Van Schooten, F.J., Baumgartner, A., Cemeli, E., van Delft, J., Anderson, D. and Godschalk, R.W.L. 2009. Use of spermatozoal mRNA profiles to study gene–environment interactions in human germ cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.* 667(1–2), pp.70-76.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B. and Kasschau, K.D. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature.* 430(6998), pp.471-476.

Liu, D., Brockman, J.M., Dass, B., Hutchins, L.N., Singh, P., McCarrey, J.R., MacDonald, C.C. and Graber, J.H. 2007. Systematic variation in mRNA 3′-processing signals during mouse spermatogenesis. *Nucleic acids research.* 35(1), pp.234-246.

Liu, W.-M., Pang, R.T.K., Chiu, P.C.N., Wong, B.P.C., Lao, K., Lee, K.-F. and Yeung, W.S.B. 2012. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proceedings of the National Academy of Sciences.* 109(2), pp.490-494.

Loewen, C.J. and Levine, T.P. 2005. A highly conserved binding site in vesicle-associated membrane protein-associated protein (VAP) for the FFAT motif of lipid-binding proteins. *Journal of Biological Chemistry.* 280(14), pp.14097-14104.

Lopes, S., Sun, J.-G., Jurisicova, A., Meriano, J. and Casper, R.F. 1998. Sperm deoxyribonucleic acid fragmentation is increased in poor-quality semen samples and correlates with failed fertilization in intracytoplasmic sperm injection. *Fertility and sterility.* 69(3), pp.528-532.

MacLaughlin, J. and Terner, C. 1973. Ribonucleic acid synthesis by spermatozoa from the rat and hamster. *Biochemical Journal.* 133(4), pp.635-639.

Mahmood, T. and Yang, P.-C. 2012. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences.* 4(9), p429.

Malcher, A., Rozwadowska, N., Stokowy, T., Kolanowski, T., Jedrzejczak, P., Zietkowiak, W. and Kurpisz, M. 2013. Potential biomarkers of nonobstructive azoospermia identified in microarray gene expression analysis. *Fertility and Sterility.* 100(6), pp.1686-1694.e1687.

Marcon, E., Babak, T., Chua, G., Hughes, T. and Moens, P. 2008. miRNA and piRNA localization in the male mammalian meiotic nucleus. *Chromosome Research.* 16(2), pp.243-260.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, pp.387-402.

Marei, W., Ghafari, F. and Fouladi-Nashta, A. 2012. Role of hyaluronic acid in maturation and further early embryo development of bovine oocytes. *Theriogenology.* 78(3), pp.670-677.

Marei, W.F., Raheem, K.A., Salavati, M., Tremaine, T., Khalid, M. and Fouladi-Nashta, A.A. 2016. Hyaluronan and hyaluronidase, which is better for embryo development? *Theriogenology.*

Marei, W.F., Salavati, M. and Fouladi-Nashta, A.A. 2013. Critical role of hyaluronidase-2 during preimplantation embryo development. *Molecular human reproduction.* pgat032.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. and Akoulitchev, A. 2007. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature.* 445(7128), pp.666-670.

Martínez, D., Pentinat, T., Ribó, S., Daviaud, C., Bloks, V.W., Cebrià, J., Villalmanzo, N., Kalko, S.G., Ramón-Krauel, M. and Díaz, R. 2014. In utero undernutrition in male mice programs liver lipid metabolism in the second-generation offspring involving altered Lxra DNA methylation. *Cell metabolism.* 19(6), pp.941-951.

Martins, R.P. and Krawetz, S.A. 2005. RNA in human sperm. *Asian Journal of Andrology.* 7(2), pp.115-120.

Mathur, P.P. and D'Cruz, S.C. 2011. The effect of environmental contaminants on testicular function. *Asian J Androl.* 13(4), pp.585-591.

Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry.* 262(21), pp.10035-10038.

Matsuura, Y., Stewart, M., Kawamoto, M., Kamiya, N., Saeki, K., Yasunaga, T. and Wakabayashi, T. 2000. Structural basis for the higher Ca 2+-activation of the

regulated actin-activated myosin ATPase observed with Dictyostelium/Tetrahymena actin chimeras. *Journal of molecular biology.* 296(2), pp.579-595.
Mattick, J.S. 2001. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports.* 2(11), pp.986-991.
Mattick, J.S. and Makunin, I.V. 2006. Non-coding RNA. *Human molecular genetics.* 15(suppl 1), pp.R17-R29.
Mazur, P. 1984. Freezing of living cells: mechanisms and implications. *American Journal of Physiology-Cell Physiology.* 247(3), pp.C125-C142.
McCarthy, D.J., Chen, Y. and Smyth, G.K. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research.* pgks042.
McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J. and Forrester, L. 1998. Gene trap integrations expressed in the developing heart: Insertion site affects splicing of the PT1-ATG vector. *Developmental dynamics.* 212(2), pp.267-276.
McIver, S.C., Roman, S.D., Nixon, B. and McLaughlin, E.A. 2012. miRNA and mammalian male germ cells. *Human Reproduction Update.* 18(1), pp.44-59.
McKeegan, P.J. 2015. *Metabolic regulation during early embryo development.* . thesis, University of York. Hull York Medical School (York). Available at: http://etheses.whiterose.ac.uk/id/eprint/8954.
Medeiros, C., Forell, F., Oliveira, A. and Rodrigues, J. 2002. Current status of sperm cryopreservation: why isn't it better? *Theriogenology.* 57(1), pp.327-344.
Meirelles, F.V., Caetano, A.R., Watanabe, Y.F., Ripamonte, P., Carambula, S.F., Merighe, G.K. and Garcia, S.M. 2004. Genome activation and developmental block in bovine embryos. *Animal Reproduction Science.* 82–83, pp.13-20.
Memili, E. and First, N.L. 1999a. Control of Gene Expression at the Onset of Bovine Embryonic Development. *Biology of Reproduction.* 61(5), pp.1198-1207.
Memili, E. and First, N.L. 1999b. Zygotic and embryonic gene expression in cow: a review of timing and mechanisms of early gene expression as compared with other species. *Zygote.* 8(01), pp.87-96.
Ménézo, Y., Veiga, A. and Pouly, J. 2000. Assisted reproductive technology (ART) in humans: facts and uncertainties. *Theriogenology.* 53(2), pp.599-610.
Mercer, T.R., Dinger, M.E. and Mattick, J.S. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics.* 10(3), pp.155-159.
Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F. and Mattick, J.S. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences.* 105(2), pp.716-721.
Meschede, D. and Horst, J. 1997. The molecular genetics of male infertility. *Molecular Human Reproduction.* 3(5), pp.419-430.
Metzker, M.L. 2010. Sequencing technologies—the next generation. *Nature reviews genetics.* 11(1), pp.31-46.
Meyer, K. and Palmer, J.W. 1934. The polysaccharide of the vitreous humor. *Journal of Biological Chemistry.* 107(3), pp.629-634.
Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S. and Thomas, P.D. 2009. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic acids research.* pgkp1019.
Miller, D. 1997. RNA in the ejaculate spermatozoon: a window into molecular events in spermatogenesis and a record of the unusual requirements of haploid gene expression and post-meiotic equilibration. *Molecular Human Reproduction.* 3(8), pp.669-676.
Miller, D. 2000. Analysis and significance of messenger RNA in human ejaculated spermatozoa. *Molecular reproduction and development.* 56(S2), pp.259-264.
Miller, D. 2006. Ensuring continuity of the paternal genome: potential roles for spermatozoal RNA in mammalian embryogenesis. *Society of Reproduction and Fertility supplement.* 65, pp.373-389.

Miller, D. 2014. Sperm RNA as a Mediator of Genomic Plasticity. *Advances in Biology.* 2014.
Miller, D. 2015. Confrontation, Consolidation, and Recognition: The Oocyte's Perspective on the Incoming Sperm. *Cold Spring Harbor Perspectives in Medicine.* 5(8).
Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R. and Krawetz, S.A. 1999. A complex population of RNAs exists in human ejaculate spermatozoa: implications for understanding molecular aspects of spermiogenesis. *Gene.* 237(2), pp.385-392.
Miller, D. and Ostermeier, G.C. 2006a. Spermatozoal RNA: why is it there and what does it do? *Gynécologie Obstétrique & Fertilité.* 34(9), pp.840-846.
Miller, D. and Ostermeier, G.C. 2006b. Towards a better understanding of RNA carriage by ejaculate spermatozoa. *Human reproduction update.* 12(6), pp.757-767.
Miller, D., Tang, P.-Z., Skinner, C. and Lilford, R. 1994. Differential RNA fingerprinting as a tool in the analysis of spermatozoal gene expression. *Human Reproduction.* 9(5), pp.864-869.
Miller, M.A., Nguyen, V.Q., Lee, M.-H., Kosinski, M., Schedl, T., Caprioli, R.M. and Greenstein, D. 2001. A Sperm Cytoskeletal Protein That Signals Oocyte Meiotic Maturation and Ovulation. *Science.* 291(5511), pp.2144-2147.
Mital, P., Hinton, B.T. and Dufour, J.M. 2011. The blood-testis and blood-epididymis barriers are more than just their tight junctions. *Biology of Reproduction.* 84(5), pp.851-858.
Mizushima, S., Takagi, S., Ono, T., Atsumi, Y., Tsukada, A., Saito, N. and Shimada, K. 2009. Phospholipase Cζ mRNA expression and its potency during spermatogenesis for activation of quail oocyte as a sperm factor. *Molecular reproduction and development.* 76(12), pp.1200-1207.
Modi, D., Shah, C., Sachdeva, G., Gadkar, S., Bhartiya, D. and Puri, C. 2005. Ontogeny and cellular localization of SRY transcripts in the human testes and its detection in spermatozoa. *Reproduction.* 130(5), pp.603-613.
Moldenhauer, J.S., Ostermeier, G.C., Johnson, A., Diamond, M.P. and Krawetz, S.A. 2003. Diagnosing male factor infertility using microarrays. *Journal of andrology.* 24(6), p783.
Mondal, T., Rasmussen, M., Pandey, G.K., Isaksson, A. and Kanduri, C. 2010. Characterization of the RNA content of chromatin. *Genome research.* 20(7), pp.899-907.
Montjean, D., Grange, P., Gentien, D., Rapinat, A., Belloc, S., Cohen-Bacrie, P., Menezo, Y. and Benkhalifa, M. 2012. Sperm transcriptome profiling in oligozoospermia. *Journal of Assisted Reproduction and Genetics.* 29(1), pp.3-10.
Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H. 1999. Exon shuffling by L1 retrotransposition. *Science.* 283(5407), pp.1530-1534.
Morel, F., Roux, C. and Bresson, J.-L. 2001. FISH analysis of the chromosomal status of spermatozoa from three men with 45, XY, der (13; 14)(q10; q10) karyotype. *Molecular human reproduction.* 7(5), pp.483-488.
Moreno, R.D., Ramalho-Santos, J., Sutovsky, P., Chan, E.K.L. and Schatten, G. 2000. Vesicular Traffic and Golgi Apparatus Dynamics During Mammalian Spermatogenesis: Implications for Acrosome Architecture. *Biology of Reproduction.* 63(1), pp.89-98.
Moretti, E., Capitani, S., Figura, N., Pammolli, A., Federico, M., Giannerini, V. and Collodel, G. 2009. The presence of bacteria species in semen and sperm quality. *Journal of Assisted Reproduction and Genetics.* 26(1), pp.47-56.
Morgan, H.D., Sutherland, H.G., Martin, D.I. and Whitelaw, E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nature genetics.* 23(3), pp.314-318.
Morton, D. 1975. Acrosomal enzymes: immunochemical localization of acrosin and hyaluronidase in ram spermatozoa. *Journal of reproduction and fertility.* 45(2), pp.375-378.

Mruk, D.D. and Cheng, C.Y. 2004. Sertoli-Sertoli and Sertoli-germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. *Endocrine Reviews.* 25(5), pp.747-806.

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. and Jovine, L. 2003. Structure, function and evolution of the signal recognition particle. *The EMBO journal.* 22(14), pp.3479-3485.

Nakamura, K., Yahagi, S.-i., Yamazaki, T. and Yamane, K. 1999. Bacillus subtilis Histone-like Protein, HBsu, Is an Integral Component of a SRP-like Particle That Can Bind theAlu Domain of Small Cytoplasmic RNA. *Journal of Biological Chemistry.* 274(19), pp.13569-13576.

Nasr-Esfahani, M., Salehi, M., Razavi, S., Mardani, M., Bahramian, H., Steger, K. and Oreizi, F. 2004. Effect of protamine-2 deficiency on ICSI outcome. *Reproductive biomedicine online.* 9(6), pp.652-658.

Naz, R.K. 1998. Effect of Actinomycin D and Cycloheximide on Human Sperm Function. *Systems Biology in Reproductive Medicine.* 41(2), pp.135-142.

Necas, J., Bartosikova, L., Brauner, P. and Kolar, J. 2008. Hyaluronic acid (hyaluronan): a review. *Veterinarni medicina.* 53(8), pp.397-411.

Nelson, G.A., Roberts, T.M. and Ward, S. 1982. Caenorhabditis elegans spermatozoan locomotion: amoeboid movement with almost no actin. *The Journal of Cell Biology.* 92(1), pp.121-131.

Ng, S.-F., Lin, R.C., Laybutt, D.R., Barres, R., Owens, J.A. and Morris, M.J. 2010. Chronic high-fat diet in fathers programs [bgr]-cell dysfunction in female rat offspring. *Nature.* 467(7318), pp.963-966.

Nguyen, M.T., Delaney, D.P. and Kolon, T.F. 2009. Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. *Fertility and sterility.* 92(1), pp.182-187.

Nielsen, H. 2011. Working with RNA. *RNA: Methods and Protocols.* pp.15-28.

Nieschlag, E., Behre, H.M. and Nieschlag, S. 2010. “Andrology-Male reproductive health and dysfunction”. *Behre, Susan Nieschlag.* 3, pp.200-206.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R., Gillingwater, T. and Webb, J. 2004. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *The American Journal of Human Genetics.* 75(5), pp.822-831.

Nodine, M.D. and Bartel, D.P. 2012. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature.* 482(7383), pp.94-97.

Nolasco, S., Bellido, J., Gonçalves, J., Tavares, A., Zabala, J.C. and Soares, H. 2012. The expression of tubulin cofactor A (TBCA) is regulated by a noncoding antisense Tbca RNA during testis maturation. *PLoS One.* 7(8), pe42536.

O'connell, M., McClure, N. and Lewis, S. 2002. The effects of cryopreservation on sperm morphology, motility and mitochondrial function. *Human Reproduction.* 17(3), pp.704-709.

O'Donnell, L., Nicholls, P.K., O’Bryan, M.K., McLachlan, R.I. and Stanton, P.G. 2011. Spermiation: the process of sperm release. *Spermatogenesis.* 1(1), pp.14-35.

Ogawa, S., Araki, S., Araki, Y., Ohno, M. and Sato, I. 2000. Chromosome analysis of human spermatozoa from an oligoasthenozoospermic carrier for a 13; 14 Robertsonian translocation by their injection into mouse oocytes. *Human Reproduction.* 15(5), pp.1136-1139.

Oliva, R. 2006. Protamines and male infertility. *Human reproduction update.* 12(4), pp.417-435.

Olszańska, B. and Borgul, A. 1993. Maternal RNA content in oocytes of several mammalian and avian species. *Journal of Experimental Zoology.* 265(3), pp.317-320.

Ørom, U.A., Nielsen, F.C. and Lund, A.H. 2008. MicroRNA-10a binds the 5′ UTR of ribosomal protein mRNAs and enhances their translation. *Molecular cell.* 30(4), pp.460-471.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P. and Krawetz, S.A. 2002. Spermatozoal RNA profiles of normal fertile men. *The Lancet.* 360(9335), pp.772-777.
Ostermeier, G.C., Goodrich, R.J., Diamond, M.P., Dix, D.J. and Krawetz, S.A. 2005a. Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. *Fertility and Sterility.* 83(6), pp.1687-1694.
Ostermeier, G.C., Goodrich, R.J., Moldenhauer, J.S., Diamond, M.P. and Krawetz, S.A. 2005b. A Suite of Novel Human Spermatozoal RNAs. *Journal of andrology.* 26(1), pp.70-74.
Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P. and Krawetz, S.A. 2004. Reproductive biology: Delivering spermatozoan RNA to the oocyte. *Nature.* 429(6988), pp.154-154.
Paasch, U., Sharma, R.K., Gupta, A.K., Grunewald, S., Mascha, E.J., Thomas, A.J., Glander, H.-J. and Agarwal, A. 2004. Cryopreservation and thawing is associated with varying extent of activation of apoptotic machinery in subsets of ejaculated human spermatozoa. *Biology of Reproduction.* 71(6), pp.1828-1837.
Pacheco, S.E., Houseman, E.A., Christensen, B.C., Marsit, C.J., Kelsey, K.T., Sigman, M. and Boekelheide, K. 2011. Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One.* 6(6), pe20280.
Pall, G.S., Wallis, J., Axton, R., Brownstein, D.G., Gautier, P., Buerger, K., Mulford, C., Mullins, J.J. and Forrester, L.M. 2004. A novel transmembrane MSP-containing protein that plays a role in right ventricle development. *Genomics.* 84(6), pp.1051-1059.
Pang, K.C., Frith, M.C. and Mattick, J.S. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics.* 22(1), pp.1-5.
Pantano, L., Jodar, M., Bak, M., Ballescà, J.L., Tommerup, N., Oliva, R. and Vavouri, T. 2015. The small RNA content of human sperm reveals pseudogene-derived piRNAs complementary to protein-coding genes. *rna.*
Parmegiani, L., Cognigni, G.E., Bernardi, S., Troilo, E., Ciampaglia, W. and Filicori, M. 2010a. “Physiologic ICSI”: Hyaluronic acid (HA) favors selection of spermatozoa without DNA fragmentation and with normal nucleus, resulting in improvement of embryo quality. *Fertility and Sterility.* 93(2), pp.598-604.
Parmegiani, L., Cognigni, G.E., Ciampaglia, W., Pocognoli, P., Marchi, F. and Filicori, M. 2010b. Efficiency of hyaluronic acid (HA) sperm selection. *Journal of Assisted Reproduction and Genetics.* 27(1), pp.13-16.
Parrington, J., Jones, M., Tunwell, R., Devader, C., Katan, M. and Swann, K. 2002. Phospholipase C isoforms in mammalian spermatozoa: potential components of the sperm factor that causes Ca2+ release in eggs. *Reproduction.* 123(1), pp.31-39.
Peaston, A., Knowles, B. and Hutchison, K. 2007. Genome plasticity in the mouse oocyte and early embryo. *Biochemical Society Transactions.* 35(3), pp.618-622.
Pellati, D., Mylonakis, I., Bertoloni, G., Fiore, C., Andrisani, A., Ambrosini, G. and Armanini, D. 2008. Genital tract infections and infertility. *European Journal of Obstetrics & Gynecology and Reproductive Biology.* 140(1), pp.3-11.
Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., Lei, L., Han, C., Ning, L., Cao, Y., Zhou, Q., Chen, Q. and Duan, E. 2012. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res.* 22(11), pp.1609-1612.
Pennetta, G., Hiesinger, P.R., Fabian-Fine, R., Meinertzhagen, I.A. and Bellen, H.J. 2002. Drosophila VAP-33A directs bouton formation at neuromuscular junctions in a dosage-dependent manner. *Neuron.* 35(2), pp.291-306.
Pessot, C.A., Brito, M., Figueroa, J., Concha, I.I., Yañez, A. and Burzio, L.O. 1989. Presence of RNA in the sperm nucleus. *Biochemical and Biophysical Research Communications.* 158(1), pp.272-278.

Platts, A.E., Dix, D.J., Chemes, H.E., Thompson, K.E., Goodrich, R., Rockett, J.C., Rawe, V.Y., Quintana, S., Diamond, M.P. and Strader, L.F. 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Human Molecular Genetics.* 16(7), pp.763-773.

Ponjavic, J. and Ponting, C.P. 2007. The long and the short of RNA maps. *Bioessays.* 29(11), pp.1077-1080.

Ponjavic, J., Ponting, C.P. and Lunter, G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome research.* 17(5), pp.556-565.

Ponting, C.P., Oliver, P.L. and Reik, W. 2009. Evolution and Functions of Long Noncoding RNAs. *Cell.* 136(4), pp.629-641.

Prather, R. 1992. Nuclear control of early embryonic development in domestic pigs. *Journal of reproduction and fertility. Supplement.* 48, pp.17-29.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H. and Jensen, T.H. 2008. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science.* 322(5909), pp.1851-1854.

Premkumar, E. and Bhargava, P. 1972. Transcription and translation in bovine spermatozoa. *Nature.* 240(100), pp.139-143.

Puri, D., Dhawan, J. and Mishra, R.K. 2010. The paternal hidden agenda: epigenetic inheritance through sperm chromatin. *Epigenetics.* 5(5), pp.386-391.

Quentin, Y. 1994. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic acids research.* 22(12), pp.2222-2227.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V. and Whitelaw, E. 2003. Transgenerational inheritance of epigenetic states at the murine AxinFu allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences.* 100(5), pp.2538-2543.

Ramalho-Santos, J., Moreno, R.D., Wessel, G.M., Chan, E.K.L. and Schatten, G. 2001. Membrane Trafficking Machinery Components Associated with the Mammalian Acrosome during Spermiogenesis. *Experimental Cell Research.* 267(1), pp.45-60.

Rando, O.J. 2012. Daddy issues: paternal effects on phenotype. *Cell.* 151(4), pp.702-708.

Rassoulzadegan, M. and Cuzin, F. 2010. The making of an organ: RNA mediated developmental controls in mice. *Organogenesis.* 6(1), pp.33-36.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I. and Cuzin, F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature.* 441(7092), pp.469-474.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C. and Gongora, M.M. 2006. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome research.* 16(1), pp.11-19.

Redgrove, K.A., Nixon, B., Baker, M.A., Hetherington, L., Baker, G., Liu, D.-Y. and Aitken, R.J. 2012. The molecular chaperone HSPA2 plays a key role in regulating the expression of sperm surface receptors that mediate sperm-egg recognition. *PLoS One.* 7(11), pe50851.

Rejon, E., Bajon, C., Blaize, A. and Robert, D. 1988. RNA in the nucleus of a motile plant spermatozoid: Characterization by enzyme-gold cytochemistry and in situ hybridization. *Molecular reproduction and development.* 1(1), pp.49-56.

Ren, Y.-F., Li, G., Wu, J., Xue, Y.-F., Song, Y.-J., Lv, L., Zhang, X.-J. and Tang, K.-F. 2012. Dicer-dependent biogenesis of small RNAs derived from 7SL RNA. *PloS one.* 7(7), pe40705.

Ren, Y.-F., Li, G., Xue, Y.-F., Zhang, X.-J., Song, Y.-J., Lv, L., Wu, J., Fang, Y.-X., Wang, Y.-Q. and Shi, K.-Q. 2013. Decreased dicer expression enhances SRP-mediated protein targeting. *PloS one.* 8(2), pe56950.

Renart, J., Reiser, J. and Stark, G.R. 1979. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proceedings of the National Academy of Sciences.* 76(7), pp.3116-3120.

Ro, S., Park, C., Sanders, K.M., McCarrey, J.R. and Yan, W. 2007. Cloning and expression profiling of testis-expressed microRNAs. *Developmental Biology.* 311(2), pp.592-602.

Roberts, T.M. and King, K.L. 1991. Centripetal flow and directed reassembly of the major sperm protein (MSP) cytoskeleton in the amoeboid sperm of the nematode, Ascaris suum. *Cell motility and the cytoskeleton.* 20(3), pp.228-241.

Roberts, T.M. and Stewart, M. 1995. Nematode sperm locomotion. *Current opinion in cell biology.* 7(1), pp.13-17.

Roberts, T.M. and Stewart, M. 2000. Acting like Actin The Dynamics of the Nematode Major Sperm Protein (Msp) Cytoskeleton Indicate a Push-Pull Mechanism for Amoeboid Cell Motility. *The Journal of Cell Biology.* 149(1), pp.7-12.

Robertson, L. and Watson, P. 1986. Calcium transport in diluted or cooled ram semen. *Journal of reproduction and fertility.* 77(1), pp.177-185.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26(1), pp.139-140.

Rodgers, A.B., Morgan, C.P., Bronson, S.L., Revello, S. and Bale, T.L. 2013. Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. *The Journal of Neuroscience.* 33(21), pp.9003-9012.

Rodgers, A.B., Morgan, C.P., Leu, N.A. and Bale, T.L. 2015. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences.* 112(44), pp.13699-13704.

Rodríguez-Campos, A. and Azorín, F. 2007. RNA is an integral component of chromatin that contributes to its structural organization. *PloS one.* 2(11), pe1182.

Roy, A., Lin, Y.-N. and Matzuk, M.M. 2007. Genetics of idiopathic male infertility. *The Genetics of Male Infertility.* Springer, pp.99-111.

Rozen, S. and Skaletsky, H. 2000. *Primer3 on the WWW for general users and for biologist programmers, pp. 365Ð386. InS. Krawetz and S. Misener (eds.), Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, NJ.

Said, T.M. and Land, J.A. 2011. Effects of advanced selection methods on sperm quality and ART outcome: a systematic review. *Human Reproduction Update.* 17(6), pp.719-733.

Sakkas, D., Manicardi, G.C. and Bizzaro, D. 2003. Sperm nuclear DNA damage in the human. *Advances in Male Mediated Developmental Toxicity.* pp.73-84.

Salustri, A., Yanagishita, M. and Hascall, V.C. 1989. Synthesis and accumulation of hyaluronic acid and proteoglycans in the mouse cumulus cell-oocyte complex during follicle-stimulating hormone-induced mucification. *Journal of Biological Chemistry.* 264(23), pp.13840-13847.

Sasso, W.D.S. 1959. Existence of hyaluronic acid at the zona pellucida of the rabbit's ovum. *Cells Tissues Organs.* 36(4), pp.352-357.

Saunders, C.M., Larman, M.G., Parrington, J., Cox, L.J., Royse, J., Blayney, L.M., Swann, K. and Lai, F.A. 2002. PLCζ: a sperm-specific trigger of Ca2+ oscillations in eggs and embryo development. *Development.* 129(15), pp.3533-3544.

Saunders, C.M., Swann, K. and Lai, F.A. 2007. PLCζ, a sperm-specific PLC and its potential role in fertilization. In: *Biochemical Society Symposia*: Portland Press Limited, pp.23-36.

Schagdarsurengin, U. and Steger, K. 2016. Epigenetics in male reproduction: effect of paternal diet on sperm quality and offspring health. *Nat Rev Urol.* advance online publication.

Schenker, J.G. and Ezra, Y. 1994. Complications of assisted reproductive techniques. *Fertility and Sterility.* 61(3), pp.411-422.
Schier, A.F. 2007. The maternal-zygotic transition: death and birth of RNAs. *Science.* 316(5823), pp.406-407.
Schieve, L.A., Meikle, S.F., Ferre, C., Peterson, H.B., Jeng, G. and Wilcox, L.S. 2002. Low and very low birth weight in infants conceived with use of assisted reproductive technology. *New England Journal of Medicine.* 346(10), pp.731-737.
Schultz, G.A. and Heyner, S. 1992. Gene expression in pre-implantation mammalian embryos. *Mutation research/Reviews in Genetic Toxicology.* 296(1-2), pp.17-31.
Schultz, R.M. 2002. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Human Reproduction Update.* 8(4), pp.323-331.
Schuster, A., Tang, C., Xie, Y., Ortogero, N., Yuan, S. and Yan, W. 2016. SpermBase: A Database for Sperm-Borne RNA Contents. *Biology of Reproduction.* pbiolreprod. 116.142190.
Sciamanna, I., Barberi, L., Martire, A., Pittoggi, C., Beraldi, R., Giordano, R., Rosa Magnano, A., Hogdson, C. and Spadafora, C. 2003. Sperm endogenous reverse transcriptase as mediator of new genetic information. *Biochemical and Biophysical Research Communications.* 312(4), pp.1039-1046.
Sciamanna, I., Vitullo, P., Curatolo, A. and Spadafora, C. 2009. Retrotransposons, reverse transcriptase and the genesis of new genetic information. *Gene.* 448(2), pp.180-186.
Scott, A., Dinman, J., Sussman, D. and Ward, S. 1989. Major sperm protein and actin genes in free-living and parasitic nematodes. *Parasitology.* 98(03), pp.471-478.
Seli, E. and Sakkas, D. 2005. Spermatozoal nuclear determinants of reproductive outcome: implications for ART. *Human reproduction update.* 11(4), pp.337-349.
Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R. and Krawetz, S.A. 2013. Stability, delivery and functions of human sperm RNAs at fertilization. *Nucleic Acids Research.* 41(7), pp.4104-4117.
Sepsenwol, S. and Taft, S.J. 1990. In vitro induction of crawling in the amoeboid sperm of the nematode parasite, Ascaris suum. *Cell motility and the cytoskeleton.* 15(2), pp.99-110.
Shafeeque, C.M., Singh, R.P., Sharma, S.K., Mohan, J., Sastry, K.V.H., Kolluri, G., Saxena, V.K., Tyagi, J.S., Kataria, J.M. and Azeez, P.A. 2014. Development of a new method for sperm RNA purification in the chicken. *Animal Reproduction Science.* 149(3–4), pp.259-265.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome research.* 13(11), pp.2498-2504.
Shapiro, A.L., Viñuela, E. and Maizel, J.V. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and biophysical research communications.* 28(5), pp.815-820.
Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W. and Sun, F. 2016. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science.* 351(6271), pp.391-396.
Shi, W. 2014. Rsubread package: high-performance read alignment,quantification and mutation discovery.
Shi, W. and Liao, Y. 2016. Subread/Rsubread Users Guide.
Shiyanov, P., Nag, A. and Raychaudhuri, P. 1999. Cullin 4A associates with the UV-damaged DNA-binding protein DDB. *Journal of Biological Chemistry.* 274(50), pp.35309-35312.
Singaravelu, G. and Singson, A. 2011. New insights into the mechanism of fertilization in nematodes. *International review of cell and molecular biology.* 289, p211.

Singh, N.P., Muller, C.H. and Berger, R.E. 2003. Effects of age on DNA double-strand breaks and apoptosis in human sperm. *Fertility and Sterility.* 80(6), pp.1420-1430.

Sinnett, D., Richer, C., Deragon, J.-M. and Labuda, D. 1991. Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *Journal of Biological Chemistry.* 266(14), pp.8675-8678.

Skakkebék, N.E., Meyts, E.R.-D., JØRgensen, N., Carlsen, E., Petersen, P.M., Giwercman, A., Andersen, A.-G., Jensen, T.K., Andersson, A.-M. and MÜLler, J. 1998. Germ cell cancer and disorders of spermatogenesis: An environmental connection? *Apmis.* 106(1-6), pp.3-12.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. 1985. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry.* 150(1), pp.76-85.

Smyth, G.K. and Verbyla, A.P. 1996. A conditional likelihood approach to residual maximum likelihood estimation in generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological).* pp.565-572.

Sone, Y., Ito, M., Shirakawa, H., Shikano, T., Takeuchi, H., Kinoshita, K. and Miyazaki, S. 2005. Nuclear translocation of phospholipase C-zeta, an egg-activating factor, during early embryonic development. *Biochemical and Biophysical Research Communications.* 330(3), pp.690-694.

Steger, K., Failing, K., Klonisch, T., Behre, H.M., Manning, M., Weidner, W., Hertle, L., Bergmann, M. and Kliesch, S. 2001. Round spermatids from infertile men exhibit decreased protamine-1 and-2 mRNA. *Human Reproduction.* 16(4), pp.709-716.

Stewart, M., King, K.L. and Roberts, T.M. 1994. The motile major sperm protein (MSP) of Ascaris suum forms filaments constructed from two helical subfilaments. *Journal of molecular biology.* 243(1), pp.60-71.

Stowe, H.M., Calcatera, S.M., Dimmick, M.A., Andrae, J.G., Duckett, S.K. and Pratt, S.L. 2014. The bull sperm microRNAome and the effect of fescue toxicosis on sperm microRNA expression. *PloS one.* 9(12), pe113163.

Stutz, G., Zamudio, J., Santillán, M.E., Vincenti, L., De Cuneo, M.F. and Ruiz, R.D. 2004. The effect of alcohol, tobacco, and aspirin consumption on seminal quality among healthy young men. *Archives of Environmental Health: An International Journal.* 59(11), pp.548-552.

Suh, N. and Blelloch, R. 2011. Small RNAs in early mammalian development: from gametes to gastrulation. *Development.* 138(9), pp.1653-1661.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A. and Parkhomchuk, D. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science.* 321(5891), pp.956-960.

Sun, J.G., Jurisicova, A. and Casper, R.F. 1997. Detection of deoxyribonucleic acid fragmentation in human sperm: correlation with fertilization in vitro. *Biology of Reproduction.* 56(3), pp.602-607.

Suri, A. 2004. Sperm specific proteins-potential candidate molecules for fertility control. *Reproductive Biology and Endocrinology.* 2(1), p1.

Sutovsky, P. 2003. Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: Killing three birds with one stone. *Microscopy research and technique.* 61(1), pp.88-102.

Swann, K., Saunders, C., Rogers, N. and Lai, F. 2006. PLCζ (zeta): a sperm protein that triggers Ca 2+ oscillations and egg activation in mammals. In: *Seminars in cell & developmental biology*: Elsevier, pp.264-273.

Taft, R.J., Hawkins, P.G., Mattick, J.S. and Morris, K.V. 2011. The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization. *Epigenetics & chromatin.* 4(1), pp.1-14.

Takemoto, N., Yoshimura, T., Miyazaki, S., Tashiro, F. and Miyazaki, J.-i. 2016. Gtsf1l and Gtsf2 Are Specifically Expressed in Gonocytes and Spermatids but Are Not Essential for Spermatogenesis. *PloS one.* 11(3), pe0150390.

Tarr, D. and Scott, A.L. 2004. MSP domain proteins show enhanced expression in male germ line cells. *Molecular and biochemical parasitology.* 137(1), pp.87-98.
Tarr, D. and Scott, A.L. 2005. MSP domain protein-1 from Ascaris suum and its possible role in the regulation of major sperm protein-based crawling motility. *Molecular and Biochemical Parasitology.* 143(2), pp.165-172.
Telford, N.A., Watson, A.J. and Schultz, G.A. 1990. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. *Molecular reproduction and development.* 26(1), pp.90-100.
Teves, M.E., Guidobaldi, H.A., Uñates, D.R., Sanchez, R., Miska, W., Publicover, S.J., Garcia, A.A.M. and Giojalas, L.C. 2009. Molecular mechanism for human sperm chemotaxis mediated by progesterone. *PloS one.* 4(12), pe8211.
Tomes, C.N., Michaut, M., Blas, G.D., Visconti, P., Matti, U. and Mayorga, L.S. 2002. SNARE Complex Assembly Is Required for Human Sperm Acrosome Reaction. *Developmental Biology.* 243(2), pp.326-338.
Toole, B.P. 2001. Hyaluronan in morphogenesis. In: *Seminars in cell & developmental biology*: Elsevier, pp.79-87.
Toth, M. 2015. Mechanisms of non-genetic inheritance and psychiatric disorders. *Neuropsychopharmacology.* 40(1), pp.129-140.
Towbin, H., Staehelin, T. and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences.* 76(9), pp.4350-4354.
Tsunematsu, R., Nishiyama, M., Kotoshiba, S., Saiga, T., Kamura, T. and Nakayama, K.I. 2006. Fbxw8 Is Essential for Cul1-Cul7 Complex Formation and for Placental Development. *Molecular and Cellular Biology.* 26(16), pp.6157-6169.
Turker, M.S. 1999. The establishment and maintenance of DNA methylation patterns in mouse somatic cells. *Seminars in Cancer Biology.* 9(5), pp.329-337.
Turley, E.A., Noble, P.W. and Bourguignon, L.Y. 2002. Signaling properties of hyaluronan receptors. *Journal of Biological Chemistry.* 277(7), pp.4589-4592.
Uchida, T., Rossignol, F., Matthay, M.A., Mounier, R., Couette, S., Clottes, E. and Clerici, C. 2004. Prolonged Hypoxia Differentially Regulates Hypoxia-inducible Factor (HIF)-1α and HIF-2α Expression in Lung Epithelial Cells IMPLICATION OF NATURAL ANTISENSE HIF-1α. *Journal of Biological Chemistry.* 279(15), pp.14871-14878.
Ullu, E., Murphy, S. and Melli, M. 1982. Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an Alu sequence. *Cell.* 29(1), pp.195-202.
Ullu, E. and Tschudi, C. 1984. Alu sequences are processed 7SL RNA genes.
Urner, F. and Sakkas, D. 2003. Protein phosphorylation in mammalian spermatozoa. *Reproduction.* 125(1), pp.17-26.
van der Heijden, G.W. and Bortvin, A. 2009. Transient relaxation of transposon silencing at the onset of mammalian meiosis. *Epigenetics.* 4(2), pp.76-79.
van Dijk, E.L., Jaszczyszyn, Y. and Thermes, C. 2014. Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research.* 322(1), pp.12-20.
Vigneault, C., Gravel, C., Vallée, M., McGraw, S. and Sirard, M.-A. 2009. Unveiling the bovine embryo transcriptome during the maternal-to-embryonic transition. *Reproduction.* 137(2), pp.245-257.
Vogt, P.H. 2004. Molecular genetic of human male infertility: from genes to new therapeutic perspectives. *Current pharmaceutical design.* 10(5), pp.471-500.
Wagner, K.D., Wagner, N., Ghanbarian, H., Grandjean, V., Gounon, P., Cuzin, F. and Rassoulzadegan, M. 2008. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Developmental cell.* 14(6), pp.962-969.
Walser, C.B. and Lipshitz, H.D. 2011. Transcript clearance during the maternal-to-zygotic transition. *Current opinion in genetics & development.* 21(4), pp.431-443.
Wassarman, P.M. and Place, O.G.L.L. 1999. Mammalian Fertilization: Review Molecular Aspects of Gamete Adhesion, Exocytosis, and Fusion. *Cell.* 96, pp.175-183.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y. and Sasaki, H. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature.* 453(7194), pp.539-543.

Watson, P.F. 2000. The causes of reduced fertility with cryopreserved semen. *Animal Reproduction Science.* 60–61, pp.481-492.

Wedell, N. 2013. The dynamic relationship between polyandry and selfish genetic elements. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 368(1613).

Weir, M.L., Amira, K. and TRIMBLE, W.S. 1998. Identification of a human homologue of the vesicle-associated membrane protein (VAMP)-associated protein of 33 kDa (VAP-33): a broadly expressed protein that binds to VAMP. *Biochemical Journal.* 333(2), pp.247-251.

Werner, A., Carlile, M. and Swan, D. 2009. What do natural antisense transcripts regulate? *RNA biology.* 6(1), pp.43-48.

Werner, A. and Swan, D. 2010. What are natural antisense transcripts good for? *Biochemical Society Transactions.* 38(4), p1144.

Werren, J.H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences.* 108(Supplement 2), pp.10863-10870.

Whitfield, M.L., Zheng, L.-X., Baldwin, A., Ohta, T., Hurt, M.M. and Marzluff, W.F. 2000. Stem-loop binding protein, the protein that binds the 3′ end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. *Molecular and cellular biology.* 20(12), pp.4188-4198.

WHO. 2010. WHO laboratory manual for the examination and processing of human semen. 5th Edition.

Wu, W., Hu, Z., Qin, Y., Dong, J., Dai, J., Lu, C., Zhang, W., Shen, H., Xia, Y. and Wang, X. 2012. Seminal plasma microRNAs: potential biomarkers for spermatogenesis status. *Molecular Human Reproduction.* 18(10), pp.489-497.

Wykes, S.M., Visscher, D.W. and Krawetz, S.A. 1997. Haploid transcripts persist in mature human spermatozoa. *Molecular Human Reproduction.* 3(1), pp.15-19.

Yagci, A., Murk, W., Stronk, J. and Huszar, G. 2010. Spermatozoa bound to solid state hyaluronic acid show chromatin structure with high DNA chain integrity: an acridine orange fluorescence study. *Journal of andrology.* 31(6), pp.566-572.

Yan, N., Lu, Y., Sun, H., Tao, D., Zhang, S., Liu, W. and Ma, Y. 2007. A microarray for microRNA profiling in mouse testis tissues. *Reproduction.* 134(1), pp.73-79.

Yanagimachi, R. 1994. Mammalian fertilization. *The physiology of reproduction.* 1, pp.189-317.

Yanagimachi, R. 2005. Male Gamete Contributions to the Embryo. *Annals of the New York Academy of Sciences.* 1061(1), pp.203-207.

Yang, C.C., Lin, Y.S., Hsu, C.C., Tsai, M.H., Wu, S.C. and Cheng, W.T.K. 2010. Seasonal effect on sperm messenger RNA profile of domestic swine (Sus Scrofa). *Animal Reproduction Science.* 119(1–2), pp.76-84.

Yang, F., Li, R., Hong, A., Duan, F. and Li, Y. 2013. Generation and characterization of a polyclonal antibody against human high mobility group box 4. *Molecular medicine reports.* 8(5), pp.1460-1464.

Yang, K., Meinhardt, A., Zhang, B., Grzmil, P., Adham, I.M. and Hoyer-Fender, S. 2012. The Small Heat Shock Protein ODF1/HSPB10 Is Essential for Tight Linkage of Sperm Head to Tail and Male Fertility in Mice. *Molecular and Cellular Biology.* 32(1), pp.216-225.

Yao, C., Wang, Z., Zhou, Y., Xu, W., Li, Q., Ma, D., Wang, L. and Qiao, Z. 2010. A study of Y chromosome gene mRNA in human ejaculated spermatozoa. *Molecular reproduction and development.* 77(2), pp.158-166.

Yatsenko, A.N., Roy, A., Chen, R., Ma, L., Murthy, L.J., Yan, W., Lamb, D.J. and Matzuk, M.M. 2006. Non-invasive genetic diagnosis of male infertility using spermatozoal

RNA: KLHL10mutations in oligozoospermic patients impair homodimerization. *Human molecular genetics.* 15(23), pp.3411-3419.

Yazama, F. 2008. Continual maintenance of the blood-testis barrier during spermatogenesis: the intermediate compartment theory revisited. *Journal of Reproduction and Development.* 54(5), pp.299-305.

Ying, S.-Y. and Lin, S.-L. 2005. Intronic microRNAs. *Biochemical and biophysical research communications.* 326(3), pp.515-520.

Yoon, J.-H., Abdelmohsen, K. and Gorospe, M. 2013. Posttranscriptional Gene Regulation by Long Noncoding RNA. *Journal of Molecular Biology.* 425(19), pp.3723-3730.

Yoon, J.-H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, Jennifer L., De, S., Huarte, M., Zhan, M., Becker, Kevin G. and Gorospe, M. 2012. LincRNA-p21 Suppresses Target mRNA Translation. *Molecular Cell.* 47(4), pp.648-655.

Yoshimoto, T., Nakamura, S., Yamauchi, S., Muto, N., Nakada, T., Ashizawa, K. and Tatemoto, H. 2008. Improvement of the post-thaw qualities of Okinawan native pig spermatozoa frozen in an extender supplemented with ascorbic acid 2-O-α-glucoside. *Cryobiology.* 57(1), pp.30-36.

Yoshimura, T., Toyoda, S., Kuramochi-Miyagawa, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T. and Miyazaki, J.-i. 2009. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Developmental Biology.* 335(1), pp.216-227.

Zhao, C., Guo, X.J., Shi, Z.H., Wang, F.Q., Huang, X.Y., Huo, R., Zhu, H., Wang, X.R., Liu, J.Y. and Zhou, Z.M. 2009. Role of translation by mitochondrial-type ribosomes during sperm capacitation: An analysis based on a proteomic approach. *Proteomics.* 9(5), pp.1385-1399.

Zhao, Y., Li, Q., Yao, C., Wang, Z., Zhou, Y., Wang, Y., Liu, L., Wang, Y., Wang, L. and Qiao, Z. 2006. Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. *Human Reproduction.* 21(6), pp.1583-1590.

Zheng, H., Stratton, C.J., Morozumi, K., Jin, J., Yanagimachi, R. and Yan, W. 2007. Lack of Spem1 causes aberrant cytoplasm removal, sperm deformation, and male infertility. *Proceedings of the National Academy of Sciences.* 104(16), pp.6852-6857.

Zini, A., Boman, J.M., Belzile, E. and Ciampi, A. 2008. Sperm DNA damage is associated with an increased risk of pregnancy loss after IVF and ICSI: systematic review and meta-analysis. *Human Reproduction.* 23(12), pp.2663-2668.

Appendix I: Terminal and R Commands

1. Trim!Galore in Terminal Session

1.1 Start terminal

trim_galore—adapter(your adapter of choice)—clip_R19—clip_R29--trim|--paired—don't_gzip--fastqc

2. R Scripts

None mentioned commands were handled automatically as default

2.1 Building Indices using RSubread

2.1.1 Set the working directory

```
setwd('/Users/Name/Documents/alignments/genomes/species')
```

2.1.2 Read in RSubread:

```
library(Rsubread)
```

2.1.3 Read in target files

```
readfile1 <- readfile1 <- c('my_file1a.fq','my_file2a.fq')
readfile2 <- readfile2 <- c('my_file1b','my_file2b.fq')
```

2.1.4 Buildindex

```
buildindex(basename="my_index",reference="genome.fa",indexSplit=TRUE,memory=8000,TH_subread=24,colorspace=FALSE)
```

2.1.5 Alignment to the genome using RSubread

```
align(index = 'my_index', readfile1=readfile1,
readfile2=readfile2,
input_format="FASTQ",
output_format="BAM",
output_file=paste(readfile1,"subread",'BAM',sep="."),
nsubreads=10,
nthreads=2)
```

2.2 GenomicRanges

2.2.1 Read in GenomicRanges:

```
library(GenomicRanges)
```

2.2.2 Built Ranges and tile genome

```
annotation.chrInfo.gr <- makeGRangesFromDataFrame(annotation.chrInfo,
seqnames='seqnames',
start.field='start',
end.field='length',
strand.field='+')


seqlevels(annotation.chrInfo.gr) <- as.character(annotation.chrInfo$seqnames)
seqlengths(annotation.chrInfo.gr) <- annotation.chrInfo$length


annotation.100bp.bins <- tileGenome(seqlengths(speciesannotation.chrInfo.gr),
tilewidth = 100,cut.last.tile.in.chrom = TRUE)
```

2.2.3 Example of how to change column names

```
colnames(species.counts2RefSeqGenes$counts) <- c('Bt_sperm1', 'Bt_testis1',
'Bt_sperm2', 'Bt_testis2')
```

2.2.4 check content first

```
head(bosTau7.counts2RefSeqGenes$annotation)
```

2.2.5 built column out of a raw

```
newcolumn <- species.counts2RefSeqGenes$counts[,c(1,3,2,4)]

newcolumn <- cbind.data.frame(species.counts2RefSeqGenes$annotation[,1], test)
```

2.3 Feature counts

2.3.1 Feature counts for index

```
file_name <- featureCounts(files,
annot.ext='/annotation_file.gtf',
isGTFAnnotationFile=TRUE,
GTF.featureType="exon",
GTF.attrType="gene_id",
useMetaFeatures=FALSE,
allowMultiOverlap=FALSE,
isPairedEnd=TRUE,
```

requireBothEndsMapped=FALSE,
checkFragLength=TRUE,
minFragLength=50,
maxFragLength=10000,
nthreads=12,
ignoreDup=TRUE)

2.3.2 Feature counts for data

file_name <- featureCounts(files,annot.ext='annotation_file.gtf',

isGTFAnnotationFile=TRUE,GTF.featureType="exon",
GTF.attrType="gene_id",
useMetaFeatures=TRUE,
allowMultiOverlap=TRUE,
isPairedEnd=TRUE)

2.4 edgeR

2.4.1 Read in edgeR:

library(edgeR)

2.4.2 Building a DGEList

dge <- DGEList(counts=y,
genes=my_feature_counts_output$annotation[,c('GeneID','Length')],
group=c(1,2,3,1,2,3))

2.4.3 Data Exploration

2.4.3.1 MDS Plot

plotMDS(dgel.glm, labels=colnames(dgel.glm$counts),
col=as.numeric(dgel.glm$samples$group), cex=1, main='MDS plot Hs.testis-sperm RNA-seq data:\n counts to Hs_refGene.')

2.4.3.2 Common Dispersion Estimation and Biological coefficient and variation

estimateCommonDisp(dgelGenf)

2.4.3.3 BCV plot

plotBCV(dgel.glm, main='BCV plot Hs.testis-sperm RNA-seq data:\n counts to Hs_refGene.')

2.4.3.4 Spearman´s rank correlations

cor.test(dgel.glm$counts[,c(1,4)],dgel.glm$counts[,c(2,5)],method='spearman')

2.4.4 Differention Expression Gene Anaylsis

2.4.4.1 Differential Expression analysis human

File name: e.g.: lrt.frsh.min.frz <- glmLRT(fit, contrast=c(1,-1,0))

2.4.4.2 Total number of differentially expressed

summary(de <- decideTestsDGE(et, p=0.05))

2.4.4.3 Data output

write.table(data,file='H:\fileposition\filename,txt'.sep='\t',row.names=FALSE)

Appendix II: Solutions

Density gradient centrifugation

Component	Final quantity
KCl	31.0 mM
NaCl	1000 mM
NaH2PO4	3.0 mM
HEPES	100.00 mM

Table II-1: Composition of X10 buffer

Component	Final quantity
$CaCl_2$	2.0 mM
$MgCl_2$	0.4 mM
Lactic Acid	21.6 mM
$NaHCO_3$	25.0 mM
Dissolve the chemicals into 5 ml of X10 buffer and adjust the pH to 7.3. Add 45 ml of 100 % Percoll and adjust the pH to 7.2 – 7.4 again and adjust the Osmolality to 290 ± 10 mOsm afterwards.	

Table II-2: Composition of 90% Percoll

Component	Final Quantity
NaCl	100.0 mM
HEPES	10.0 mM
NaH_2PO4	0.3 mM
KCl	3.1 mM
$NaHCO_3$	25.0 mM
Lactic acid	21.6 mM
$CaCl_2$	2.0 mM
$MgCl_2$	0.4 mM
Sodium Pyruvate	1.0 mM
BSA	10 mg/ml
Adjust the pH to 7.3 after solving the chemicals	

Table II-3: Composition of 1x spTalp

Agarose gel electrophoresis

Component	Volume/Weight per sample
Bromophenolblue	250 mg
150 mM Tris pH 7.6	33 ml
Glycerol	60 ml
Distilled Water	7 ml

Table II-4: 10x Loading buffer components

Component	Volume/Weight per sample
Tris-base	108 g
Boric Acid	55 g
Na_2EDTA	7.4 g
Up to 1L	

Table II-5: 10x TBE buffer component

Protein extraction and concentration

Component	Final Quantity/Volume/Weight
Tris HCl pH 8	50 mM
NaCl	150 mM
EDTA	10 mM
NP-40	1% (v/v)
$C_{24}H_{39}NaO_4$	0.5% (w/v)
SDS	0.5% (w/v)
Proteinase Inhibitor Cocktail	1 µl
PMSF	10 µl

Table II-6: RIPA buffer components

Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis

Component	Final Volume
30 % (v/v) Acrylamide	2.08 ml
H_2O	1.57 ml
1.5 M Tris pH 8.8	1.25 ml
10% (w/v) SDS	50 µl
10% (w/v) APS	50 µl
TEMED	5 µl

Table II-7: Components of the 12.5% resolving gel

Component	Final Volume
30 % (v/v) Acrylamide	340 µl
H_2O	1.36 ml
1M Tris pH 6.8	250 µl
10% (w/v) SDS	20 µl
10% (w/v) APS	20 µl
TEMED	2 µl

Table II-8: Components of the 4% stacking gel

Component	Final Weight
Tris Base	30 g
SDS	10 g
Glycine	145 g

Table II-9: Components of a 1 x running buffer, made up in 1l

Component	Final Concentration/Percentage
Tris-HCL	250 mM
SDS (w/v)	10 %
Glycerol (v/v)	50 %
Bromophenol Blue (w/v)	0.02 µl

Table II-10: Components of 5x sample buffer

Protein transfer, blocking and detection

Component	Final Concentration/Percentage
Tris base	4 mM
Glycine	192 mM
Methanol (cold) (v/v)	20 %

Table II-11: Components of a 1x transfer buffer

Component	Final Weight/Percentage
Tris-HCl	2.4 g
Tris base	0.56 g
NaCl	8.8 g
Tween20 (v/v)	0.5 %
pH 7.6	

Table II-12: Components of 1x TBST

Appendix III: TC Stocks and Solutions

Stocks

Stock A

Compound	Quantity	Concentration
NaCl	6.36 g	1.07x103 mM
KCl	0.534 g	71.6 mM
KH2PO4	0.162 g	11.9 mM
MgSO4.7H2O	0.182 g	7.4 mM
Sodium lactate	0.991 ml	70 mM
ET water	90.9 ml	

Sterile filter at 4°C and store for 6 weeks

Stock B

Compound	Quantity	Concentration
NaHCO3	1.05 g	250 mM
ET water	50 ml	

Sterile filter at 4°C and store for 2 weeks

Stock C

Compound	Quantity	Concentration
Pyruvate	0.036 g	32.7 mM
ET water	10 ml	

Sterile filter at 4°C and store for 2 weeks

Stock D

Compound	Quantity	Concentration
CaCl2.2H2O	0.182 g	170 mM
ET water	10 ml	

Sterile filter and store at 4°C for 6 weeks

Stock G

Compound	Quantity	Concentration
D+ Glucose	0.108 g	60 mM
ET water	10 ml	

Sterile filter and store at 4°C for 6 weeks

Stock Gln

Compound	Quantity	Concentration
L- Glutamine	0.0292 g	200mM
ET water	1 ml	

Sterile filter and store at 4°C for 2 weeks

Stock H

Compound	Quantity	Concentration
Hepes	1.5 g	126 mM
Hepes Sodium salt	1.625 g	125 mM
ET water	50 ml	

Sterile filter and store at 4°C for 6 weeks

Stock L

Compound	Quantity	Concentration
Sodium-lactate syrup	0.47 ml	332.06 mM
ET water	9.53ml	

Sterile filter and store at 4°C for 6 weeks

Stock M

Compound	Quantity	Concentration
MgCl2.6H2O	0.1 g	49.19 mM
ET water	10 ml	

Sterile filter and store at 4°C for 6 weeks

Stock S2

Compound	Quantity	Concentration
NaCl	3.147 g	1.08 mM
KCl	0.267 g	71.62 mM
KH2PO4	0.081 g	11.9 mM
ET water	50 ml	

Sterile filter and store at 4°C for 6 weeks

Stock Heparin

Compound	Quantity	Concentration
Heparin sodium salt (H3149, porcine intestine) (or bovine)	0.02 g	360 U/ml or 2 mg/ml
ET water	10 ml	

Do not filter. Aliquot to 20µl and store at -20°C

Stock 10x TL

Compound	Quantity	Concentration
NaCl	1.6665 g	1.07x103mM
KCl	0.0595 g	71.6 mM
NaH2PO4	0.0155 g	5 mM
(or NaH2PO4.2H2O)	0.0178 g	5 mM
Gentamycin	1.25 ml	
ET water	24.75 ml	

Sterile filter and at 4°C store for 6 weeks

Stock Pen/Hyp

Compound	Quantity	Concentration
Penicillamine	0.003 g	200 µM
ET water	5ml	

Compound	Quantity	Concentration
Hypotaurine	0.0022 g	100 µM
ET water	10 ml	

Combine 5 ml Hypotaurine stock with 5 ml Penicillamine stock. Sterile filter. Aliquot to 300 µl and store at -20°C

Stock BSA FAF

Compound	Quantity	Concentration
BSA FAF	2 g	0.2 g/ml
ET water	10 ml	

Sterile filter and store at 4°C for 2 weeks

Stock BSA FrV

Compound	Quantity	Concentration
BSA FrV	2 g	0.2 g/ml
ET water	10 ml	

Sterile filter and at 4°C store for 2 weeks

Stock 47 mM Pyruvate

Compound	Quantity	Concentration
Sodium pyruvate	0.0517 g	47 mM
ET water	10 ml	

Prepare fresh. Sterile filter before usage.

Stock Penicillin/Streptomycin

Compound	Quantity	Concentration in media
Penicillin/Streptomycin	5 ml in 500 ml solution	50 IU/ml, 50 µg/ml

oFSH Stock

Compound	Quantity	Concentration in media
oFSH Stock (200 µg/ml; 2 IU/ml)	9 mg	0.001 IU/ml
10% (w/v) BSA, fraction V cell culture grade	4.5 ml	
MEM	40.5 ml	

Sterile filter and store at -20°C up to 3 month

oLH Stock

Compound	Quantity	Concentration in media
oLH Stock (200 µg/ml; 2 IU/ml))	9 mg	0.0005 IU/ml
10% (w/v) BSA, fraction V cell culture grade	4.5 ml	
MEM	40.5 ml	

Sterile filter and store at -20°C up to 3 months

Stock Holo Transferrin

Compound	Quantity	Concentration in media
Holo Transferrin Stock (5 mg/ml)	100 mg	5 µg/ml
MEM containing 0.1% (w/v) fraction V cell culture grade BSA	20 ml	

Sterile filter and store at -20°C up to 3 months

Stock Sodium Selenite

Compound	Quantity	Concentration in media
Sodium Selenite Stock (50 µg/ml)	1 mg	5 ng/ml
MEM containing 0.1% (w/v) fraction V cell culture grade BSA	20 ml	

Sterile filter and store at -20°C up to 3 months

Stock Insulin

Compound	Quantity	Concentration in media
Insulin stock (10 mg/ml)	10 mg	5 ng/ml
MEM containing 0.1% (w/v) fraction V cell culture grade BSA	20 ml	

Sterile filter and store at -20°C up to 3 months

Stock Long-R3 IGF1

Compound	Quantity	Concentration in media
100µg/ml Long R3 IGF1 Stock (100 µg/ml)	500 µg	1 ng/ml
MEM containing 0.1% (w/v) fraction V cell culture grade BSA	5 ml	

Sterile filter and store at -20°C up to 3 month

Hyaluronidase

Compound	Quantity	Concentration in media
Hyaluronidase from bovine testes; Type I-S (330 IU/mg)	12.12 mg	80 IU/ml
Holding Media	50 ml	

Sterile filter and store at -20°C. Pre-heat prior to usage to 39°C.

Media-Aspiration and IVM

Ovary Wash

Component	Volume
PBS (2x PBS tablets in 1L 18.2mΩ water)	1 L
Antibiotic/Antimycotic	10 ml

Sterile Filter and store at 4°C for 1 week

Follicle Isolation Medium

Component	Volume	Concentration
Hepes-MEM	470 ml	
Pen/Strep	5 ml	1000 IU/ml
20x BSA	25 ml	4 mg/ml

Sterile Filter and store at 4°C for 1 weekHolding Medium

Component	Volume	Concentration
ET Water	192.5 ml	
10x M199	25 ml	1x
Stock B	4 ml	4 mM
Stock H	21 ml	21.1 mM
Pen/Strep	2.5 ml	
~285 mOsm/kg		
Heparin	5mg or 0.2ml	0.02 mg/ml
Stock BSA	0.5ml	2 mg/ml

Using M199 + Glutamine (Gibco 10x 21180-021 or Sigma M2520)

Sterile Filter and store at 4°C for 1 week

Serum Free Maturation Media

Component	Volume for 10ml	Concentration
αMEM including NaHCO3	9ml	
200mM Gln	100 µl	2 mM
47mM Pyruvate stock	100 µl	0.47 mM
Holo-transferrin	10 µl	5 µg/ml
Sodium Selenite	1 µl	5 ng/ml
Insulin	10 µl	10 ng/ml
Long-R3 IGF-1	1 µl	10 ng/ml
FSH (2IU/ml)	3 µl	0.0006 IU/ml
LH (2IU/ml)	1.5 µl	0.0003 IU/ml
Pen/Strep	15 µl	100 IU/ml
Stock BSA FAF	200 µl	1 mg/ml

Media-IVF

90% Percoll solution

Component	Amount	Concentration	Notes
Stock A	5 ml	NaCl 107 mM KCl 7.16 mM KH2PO4 11.9 mM MgSO4.7H2O 7.4 mM Lactate 7 mM	Dissolve Hepes and bicarbonate in stock A Add Percoll Add stock D
Stock D	0.5 ml	1.53 mM	
Hepes free acid	126 mg	10.5 mM	
Hepes sodium salt	137 mg	10.5 mM	
NaHCO3	96 mg	22 mM	
Percoll	44.5 ml		

45% Percoll

Component	Volume
90% Percoll	2.0 ml
H-Talp	2.0 ml

Hepes TALP sperm wash media and Fertilisation TALP co-incubation media

Component	**Volume (H-TALP)**		**Component**	**Volume (Fert-TALP)**
ET water	18.8 ml		ET water	14.32 ml
Stock 10x TL	2.55 ml		Stock 10x TL	2 ml
Stock B	200 µl		Stock B	2 ml
Stock C	200 µl		Stock C	160 µl
Stock D	300 µl		Stock D	240 µl
Stock H	1.5 ml			
Stock L	750 µl		Stock L	600 µl
Stock M	250 µl		Stock M	200 µl
Check osmolarity ~285			Check osmolarity ~285	
Stock BSA FrV	500 µl		Stock BSA FAF	400 µl
			Heparin	100 µl
			Pen/Hyp	200 µl

Media-Embryo Culture

Component (H-SOF)	**Volume (20 ml)**	**mM**		**Component (SOFaaBSA)**	**Volume (10 ml)**	**mM**
ET water N.B start with 13ml	14.12 ml			ET water N.B start with 5ml	5.55 ml	
Stock S2 NaCl KCl NaH2PO4	2 ml	 111 7.16 1.19		Stock S2 NaCl KCl NaH2PO4	1 ml	 111 7.16 1.19
Stock B (NaHCO3)	400 µl	5		Stock B	1 ml	25
Stock C (Pyruvate)	200 µl	0.33		Stock C	100 µl	0.33
Stock D (CaCl2.2H2O)	200 µl	1.71		Stock D	100 µl	1.71
Stock G (Glucose)	500 µl	1.5		Stock G	250 µl	1.5
				Stock GLN	1 ml	0.1
Stock H (Hepes)	1.6 ml	20				
Stock L (Lactate)	200 µl	33.2		Stock L	100 µl	33.2
Stock M (MgCl2.7H2O)	200 µl	4.9		Stock M	100 µl	4.9
				100x NEAA	100 µl	
				50x EAA	200 µl	
Check osmolarity ~285				Check osmolarity ~285		
Stock BSA FrV	400 µl	4mg/ml		Stock BSA FAF	400 µl	8mg/ml
Pen/Strep	200 µl			Pen/Strep	100 µl	

Appendix IV: Suppliers and chemicals, materials, reagents

1. Suppliers

Advanced Instruments, Inc. Two Technology Way Norwood MA 02062 United States	**Agilent Technologies LDA UK Limited** Lakeside Cheadle Royal Business Park Stockport Cheshire SK8 3GR United Kingdom
Ambion® **Thermo Fisher Scientific (Life Technologies)** Inchinnan Business Park 3 Fountain Drive Paisley PA4 9RF United Kingdom	**Amsbio** 184 Park Drive, Milton Park, Abingdon OX14 4SE United Kingdom
BDH Poole Dorset BH15 1TD United Kingdom	**Beckman Coulter Ltd** Oakley Ct/Kingsmead Business Park/Frederick Pl High Wycombe HP11 1JU United Kingdom
Bio-Rad Laboratories Ltd Bio-Rad House, Maxted Road, Hemel Hempstead, Herts, HP2 7DX United Kingdom	**Clontech Laboratories, Inc.** A Takara Bio Company 1290 Terra Bella Ave. Mountain View CA 94043 United States

Cook Medical O'Halloran Road National Technology Park Limerick Ireland	**Covaris Ltd.** Unit 3, Brighton Office Campus Hunns Mere Way Woodingdean Brighton BN2 6AH United Kingdom
Cook Medical O'Halloran Road National Technology Park Limerick Ireland	**Covaris Ltd.** Unit 3, Brighton Office Campus Hunns Mere Way Woodingdean Brighton BN2 6AH United Kingdom
Fisher Scientific UK Ltd Bishop Meadow Road Loughborough LE11 5RG United Kingdom	**Greiner Bio-one Ltd** Brunel Way Stroudwater Business Park Stonehouse Glos. GL10 3SX United Kingdom
Epicenter and Illumina Saffron Walden Cambridgeshire CB10 1XL United Kingdom	**Millipore (UK) Limited** Suite 3&5 Building 6, Croxley Green Business Park, Watford WD18 8YH United Kingdom
Origio Ltd 7 Castlefield Rd Reigate RH2 0SA United Kingdom	**Next Advance Inc.** 1548 Burden Lake Rd Averill Park NY 12018 United States of America

New England Biolabs Ltd 75-77 Knowl Piece Hitchin SG4 0TY United Kingdom	**NuGEN** 201 Industrial Road Suite 310 San Carlos CA 94070 United States of America
Nunc distributed by Sigma-Aldrich	**QIAGEN Ltd** Skelton House Lloyd Street North Crawley Manchester United Kingdom
Sarstedt Ltd. 68 Boston Road Beaumont Leys Leicester LE4 1AW United Kingdom	**Scientific Laboratory Supplies Limited** Wilford Industrial Estate Ruddington Lane Wilford Nottingham NG11 7EP United Kingdom
Sigma-Aldrich Company Ltd. The Old Brickyard New Road Gillingham Dorset SP8 4XT United Kingdom	**Takara Bio** 2 Avenue du Président John Fitzgerald Kennedy, 78100, Saint-Germain-en-Lave, France
Thermo Fisher Scientific (Life Technologies) 3 Fountain Drive Inchinnan Business Park Paisley PA4 9RF United Kingdom	**VWR International Ltd** Hunter Boulevard Magna Park Lutterworth Leicestershire LE17 4XN United Kingdom

3. Chemicals, Enzymes, Antibodies, Materials

Chemicals, Enzymes, Antibodies, Materials	**Supplier**	**Cat No.**
Acetic Acid	Fisher Scientific Ltd	A/0400/PB17
Acrylamide/bis-acrylamide, 40% solution	Sigma-Aldrich	A7168
Adhesion slides, Polysine®	VWR	631-0107
Agilent RNA 6000 Pico Kit	Agilent Technologies	5067-1513
Agilent High Sensitive DNA Kit	Agilent Technologies	5067-4626
Ammonium persulfate	Sigma-Aldrich	A3678
Amicon ®Ultra-0.5 Centrifugal Filter Devices 3K	Millipore (UK) Limited	UFC500396
BCA protein assay	Thermo Fisher Scientific	23225
Brilliant Blue R Staining Solution	Sigma-Aldrich	B8647-1EA
Bromophenol Blue sodium salt	Sigma-Aldrich	B6131
BSA FAF	Sigma-Aldrich	A6003
BSA Fr V	Sigma-Aldrich	A9418
BSA for Immunocytochemistry	Sigma-Aldrich	A7906
Calcium Chloride Dihydrate	BDH	260355U
Calcium Chloride Dihydrate ($CaCl_2.2H_2O$)	Sigma-Aldrich	C7902
Chloroform	Sigma-Aldrich	288306-
ColorPlus Prestained Protein Ladder, Broad Range (10-230 kDa)	New England Biolabs UK Ltd	P7711S

Chemicals, Enzymes, Antibodies, Materials	**Supplier**	**Cat No.**
Covaris Microtubes for shearing	Covaris	S20045
DAPI	Sigma-Aldrich	32670
Diethylpyrocarbonate	Sigma-Aldrich	40718
Diethyl pyrocarbonate (DEPC)	Sigma-Aldrich	40718
DNA ladder, 100 bp	New England Biolabs UK Ltd	N0467L
dNTP Mix 100mM	BIOLINE	BIO-39028
Di-Sodium Hydrogen Orthophosphate Dihydrate ($Na_2HPO_4.2H_2O$)	Sigma-Aldrich	71638
DPX mounting media	BDH	360294H
EGF (bovine)	Sigma-Aldrich	E4127
Ethanol (Mol.Bio grade)	Sigma-Aldrich	51976-500ML-F
Ethylenediaminetetraacetic acid	Sigma-Aldrich	E6758
FCS	Sigma-Aldrich	F9665
Femto Maximum Sensitivity Substrate	Thermo Fisher Scientific	34095
FGF (bovine)	Sigma-Aldrich	F3133
FSH (ovine)	Sigma-Aldrich	F8174-1VL
G26 needle	SLS	SYR6250
Gentamycin Solution	Sigma-Aldrich	G1272
Giemsa stain	Sigma-Aldrich	G5637
goat anti-mouse lgG-TRITC	abcam	Ab6786
Glucose	Sigma-Aldrich	G6152
Glycerol	Sigma-Aldrich	G9012
Glycoblue	Thermo Scientific	AM9515

Chemicals, Enzymes, Antibodies, Materials	**Supplier**	**Cat No.**
Heparin (bovine)	Sigma-Aldrich	H0777
Heparin (porcine)	Sigma-Aldrich	H3149
HEPES free acid	Sigma-Aldrich	H4034
HEPES sodium salt	Sigma-Aldrich	H7006
Insulin	Sigma-Aldrich	I6634
Isopropanol, 99.5%, for HPLC gradient grade	Fisher Scientific	10722295
L(+)-lactic acid free acid 30% solution	Sigma-Aldrich	L1875-
L-Glutamine	Sigma-Aldrich	G8540
LH (ovine)	Sigma-Aldrich	L5269-1VL
LONG®R^3 IGF-I (human)	Sigma-Aldrich	I1271-.1MG
Low Input Library Prep Kit	Clontech/Takara	634947
M199 liquid	Sigma-Aldrich	M4530
M199 liquid 10x	Sigma-Aldrich	M0650
Magnesium Chloride Anhydrous	Sigma-Aldrich	M8266
Magnesium Sulphate Heptahydrate ($MgCl_2.7H_2O$)	Sigma-Aldrich	M2643
MEM 100x	Sigma-Aldrich	M7145
Methanol	Fisher Scientific Ltd	M/4056/17
MinElute®Reaction Clean Up Kit	QIAGEN	28204
Mineral Oil	Sigma-Aldrich	M8410
Monosodium Phosphate	BDL	10245
anti-MOSPD3	abvnova	H00064598-M05
Neubauer Chamber	Thermo Fisher Scientific	

Chemicals, Enzymes, Antibodies, Materials	Supplier	Cat No.
Nonyl phenoxypolyethoxylethanol 40	BHD	56009
Oligomycin	Sigma-Aldrich	O4876-5MG
Ovation® RNA-Seq System V2	NuGEN	Version 7102
Ovation® Ultralow Library Systems	NuGEN	Version 0344
Phosphate Buffered Saline Tablets	Gibco	18912-014
Penicillamine	Sigma-Aldrich	P4875
Penicillin/Streptomycin 10mg/ml	Sigma-Aldrich	P4333-100ML
Percoll®	Fisher	10607095
Phenylmethylsulfonyl fluoride	VWR	A0999,0005
Prestained Protein Marker	NEB	P7712
Potassium Chloride (KCl)	BHD	10198
Protease Inhibitor Cocktail	Sigma-Aldrich	P2714
Polypropylene tube	Greiner	
Polyvinylidene difluoride membrane	Bio-Rad Laboratories Ltd	1620177
Qubit® dsDNA HS Assay Kit	Thermo Fisher Scientific	Q3285
Qubit HS RNA Kit	Thermo Fisher Scientific	Q32852
Qubit BR RNA Kit	Thermo Fisher Scientific	Q10210
Quinns Advantage ™Sperm Freezing Medium	Origio	ART-8022
Ribo-Zero™Magnetic Kit	Epicenter, now Illumina	MRZH11124

Chemicals, Enzymes, Antibodies, Materials	**Supplier**	**Cat No.**
RNAse Block	Agilent Technologies LDA UK Limited	300151
RNeasy MiniKit	QIAGEN	74104
Slim Milk	VWR	84615.0500
SMARTer® Universal Low Input RNA Kit for Sequencing	Clontech	634938
Sperm Preparation Medium PURECEPTION 100% ISOTONIC	Origio	ART-2100
Sperm Washing Medium QUINN'S SPERM WASH	Origio	ART-1006
Sodium Bicarbonate ($NaHCO_3$)	Sigma-Aldrich	S5761
Sodium Chloride (NaCl)	Fisher Scientific	BP 358-212
Sodium Dodecyl Sulfate	SIGMA	L4390
Sodium Deoxycholate	Sigma-Aldrich	D6750
Sodium Lactate Syrup	Sigma-Aldrich	L1375
Sodium Pyruvate	Sigma-Aldrich	P2256
Sodium selenite	Sigma-Aldrich	S5261
0.5 mm stainless steel beads	Next Advance	SSB05-RNA
SuperScript® III Reverse Transcriptase (2000U)	Thermo Fisher Scientific	18080093
TURBO DNA-free ™ Kit	Thermo Fisher Scientific	AM1907
Transferrin	Sigma-Aldrich	T0665
Trizma HCl	Melford	T3253
Trizma Base	Sigma-Aldrich	33742

Chemicals, Enzymes, Antibodies, Materials	**Supplier**	**Cat No.**
Trizol	Thermo Fisher Scientific	15596018
Tween 20	Sigma-Aldrich	P9416
β-mercaptoethanol	Sigma-Aldrich	M3148
N,N,N′,N′-Tetramethylethylenediamine	Sigma-Aldrich	T9281
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	BDH	442854V

Appendix V: PCR

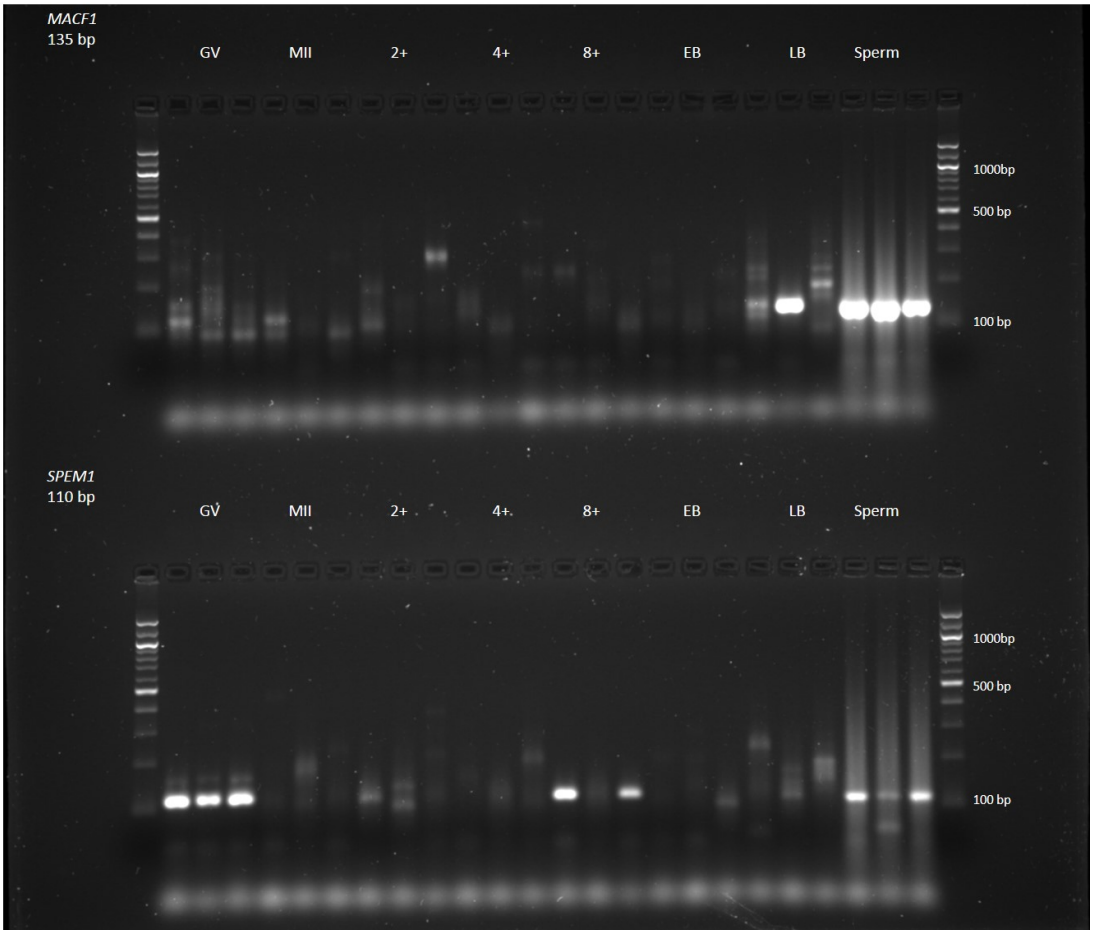


Figure V-1: Embryonic fate of spermatozoal RNA 1: a) *MACF1* (135 bp); b) *SPEM1* (110 bp). Transcripts were analysed in all collected embryo stages.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

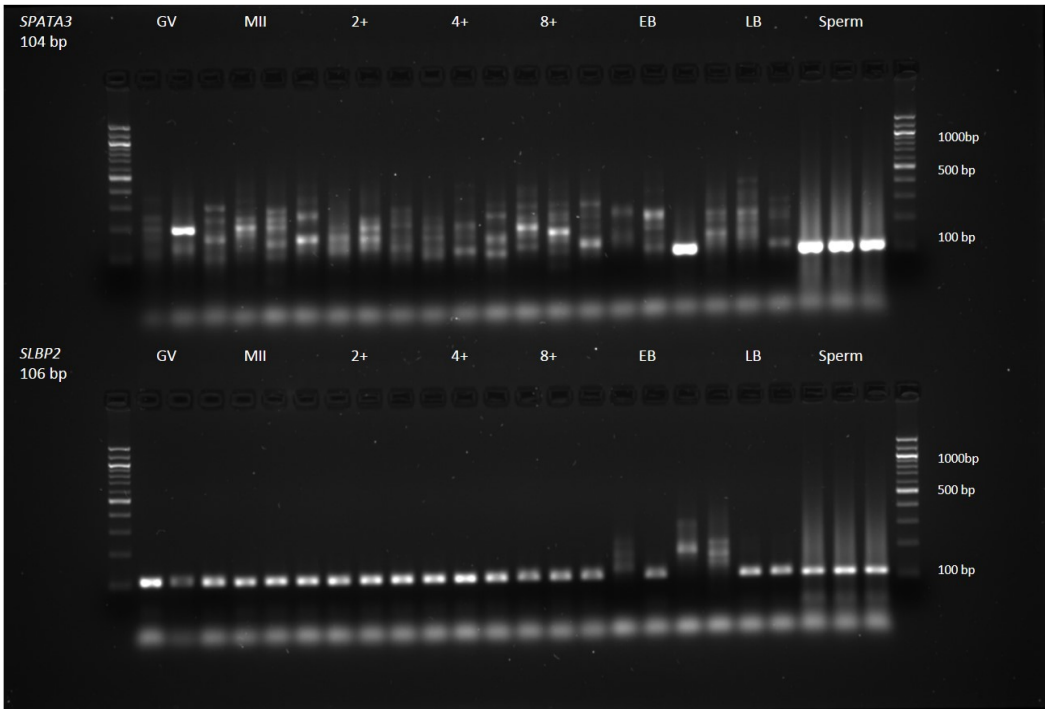


Figure V-2: Embryonic fate of spermatozoal RNA 2. a) *SPATA3* (104 bp); b) *SLBP2* (106 bp).

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

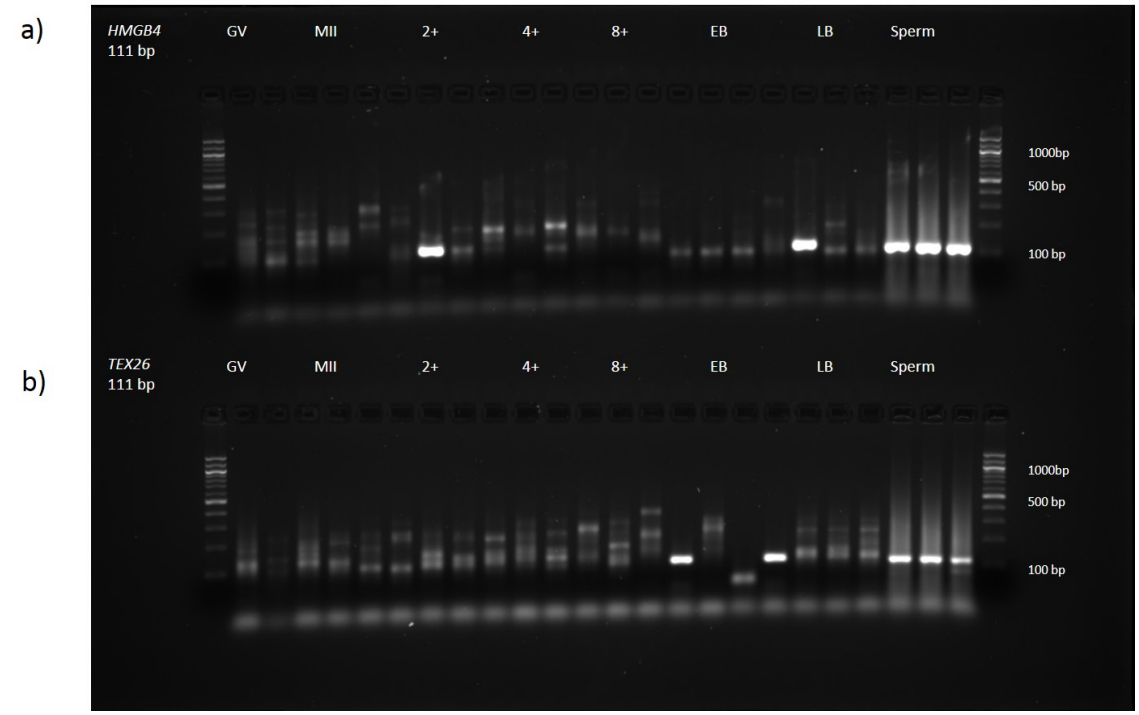


Figure V-3: Embryonic fate of spermatozoal RNA 3. a) *HMGB4* (111 bp); b) *TEX26* (111bp). Most chosen transcripts showed more than one amplified transcript and the spermatozoal RNA transfer can just be speculated.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

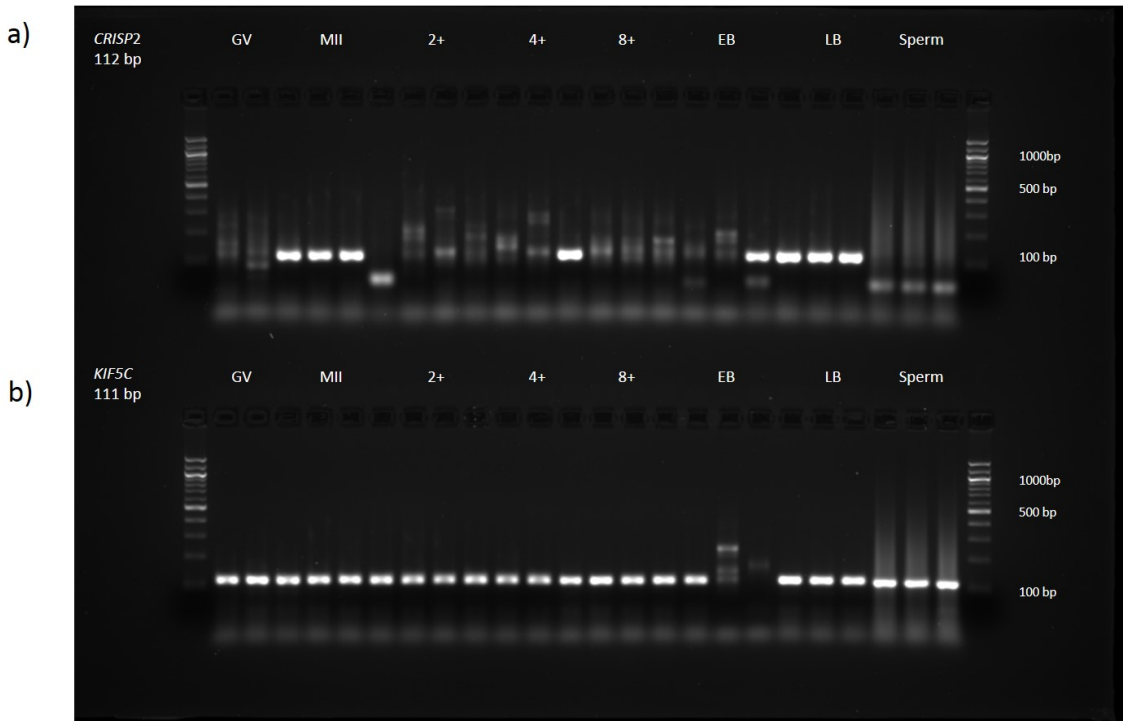


Figure V-4: Embryonic fate of spermatozoal RNA 4. a) *CRISP2* (112 bp); b) *KIF5C* (111 bp).

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

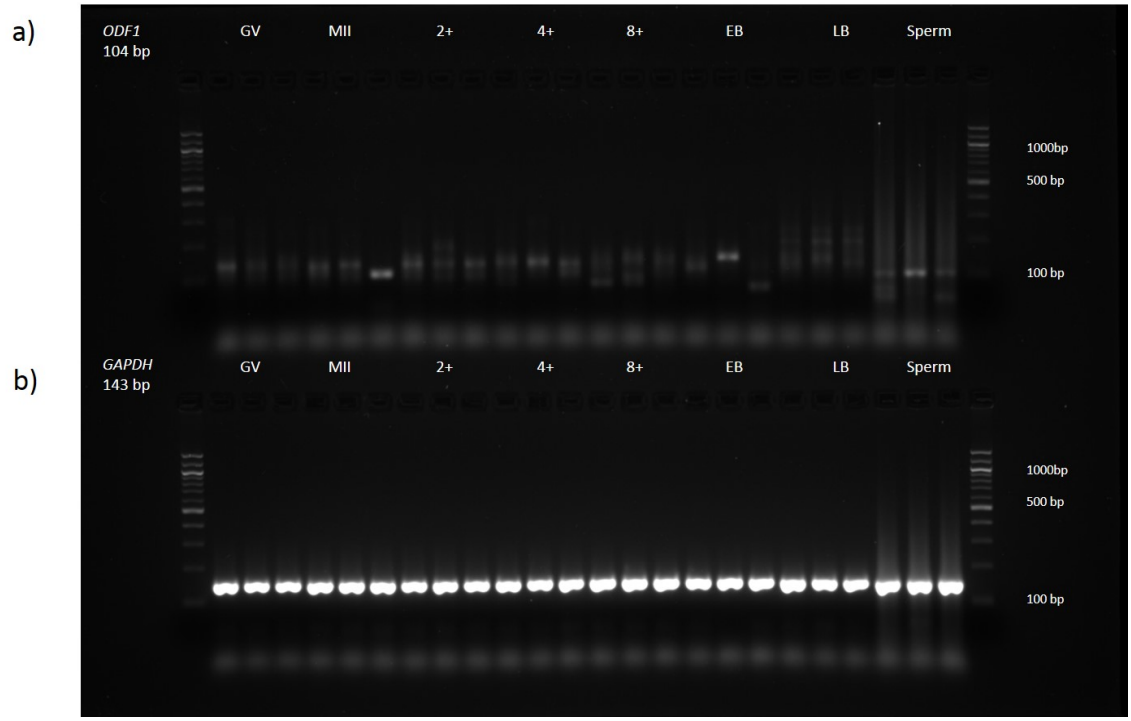


Figure V-5: Embryonic fate of spermatozoal RNA 5. a) *ODF1* (104 bp); b) *GAPDH* (143 bp) as housekeeper.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

a)

b)

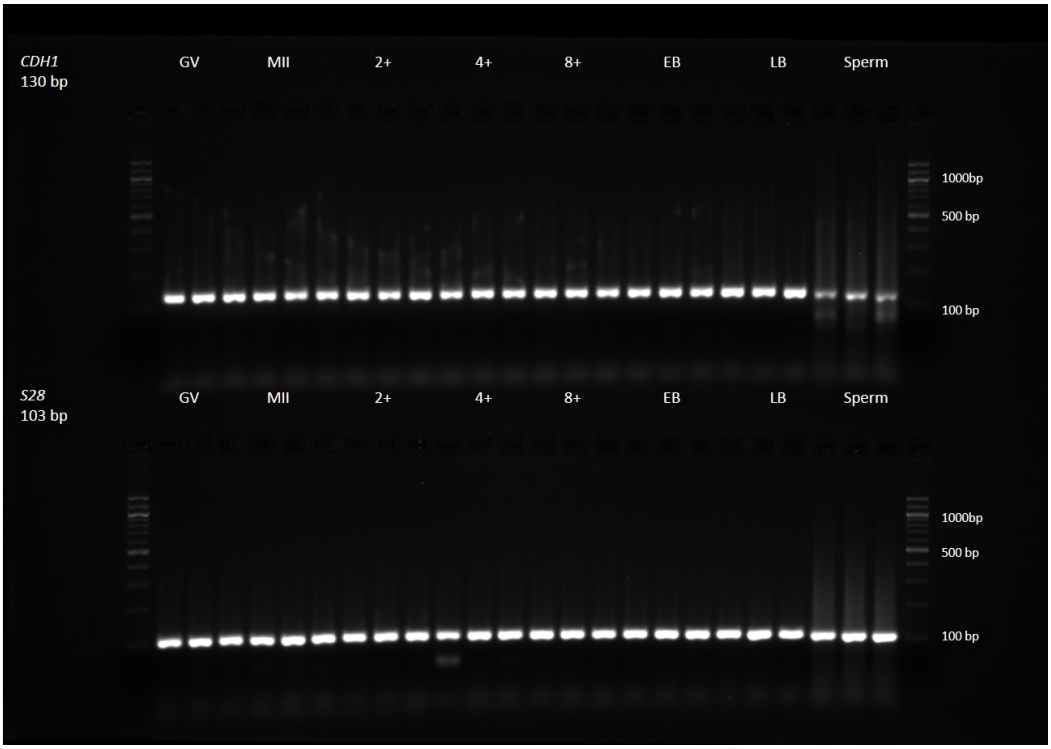


Figure V-6: Embryonic fate of spermatozoal RNA 6. a) *CDH1* (130 bp); b) *S28* (103 bp); as housekeepers.

Transcripts were analysed in all collected embryo stages: GV=Germinal vesicle, MII= Metaphase II oocytes, 2+=2 plus cell stage, 4+=4 plus cell stage, 8+=8 plus cell stage, EB=Early blastocysts and LB=Late blastocysts and sperm=Spermatozoa.

Appendix VI: SeqMonk Table

Chr	Feature	ID	HA selected	Un selected	S/US	M2 ≥ 600?
2	null		581.26	0.00	#DIV/0!	No
17	RBFOX3-002	ENST00000338834	472.21	0.00	#DIV/0!	No
15	RP11-69H14.6-005	ENST00000560134	342.21	0.00	#DIV/0!	No
1	RP11-345P4.5-001	ENST00000422725	302.59	0.00	#DIV/0!	No
7	STYXL1-201	ENST00000248600	228.82	0.00	#DIV/0!	No
5	null		160.92	0.00	#DIV/0!	No
12	null		164.68	0.00	#DIV/0!	No
6	null		187.71	0.00	#DIV/0!	No
15	RP11-69H14.6-005	ENST00000560134	159.23	0.00	#DIV/0!	No
18	B4GALT6-001	ENST00000306851	184.54	0.00	#DIV/0!	No
X	null		157.11	0.00	#DIV/0!	No
6	LEMD2-001	ENST00000293760	151.66	0.00	#DIV/0!	No
8	null		147.55	0.00	#DIV/0!	No
12	RP11-1020M18.10-001	ENST00000548135	145.90	0.00	#DIV/0!	No
18	MYO5B-201	ENST00000285039	150.28	0.40	375.70	No
1	RP11-345P4.5-001	ENST00000422725	195.94	0.62	315.71	No
20	null		163.73	1.00	163.73	No
12	R3HDM2-001	ENST00000347140	273.30	2.11	129.62	No
17	CDK12-201	ENST00000430627	210.21	2.06	101.87	No
17	CDK12-201	ENST00000430627	190.65	2.12	89.95	No
X	SH3KBP1-201	ENST00000379702	211.79	3.22	65.67	No
6	DTNBP1-001	ENST00000344537	152.22	2.33	65.45	No
18	ATP9B-202	ENST00000426216	152.78	2.45	62.34	No
6	ASCC3-001	ENST00000369162	502.43	8.58	58.56	No
15	C15orf39-001	ENST00000360639	272.90	4.87	56.04	No
16	HSD11B2-001	ENST00000326152	177.10	3.40	52.09	No
1	ARF1-001	ENST00000272102	178.31	3.77	47.27	No
11	POU2AF1-007	ENST00000531398	159.46	3.44	46.41	No
13	null		400.59	8.66	46.25	No
17	MINK1-202	ENST00000355280	152.61	3.40	44.89	No
8	PAG1-001	ENST00000220597	494.90	12.97	38.16	No

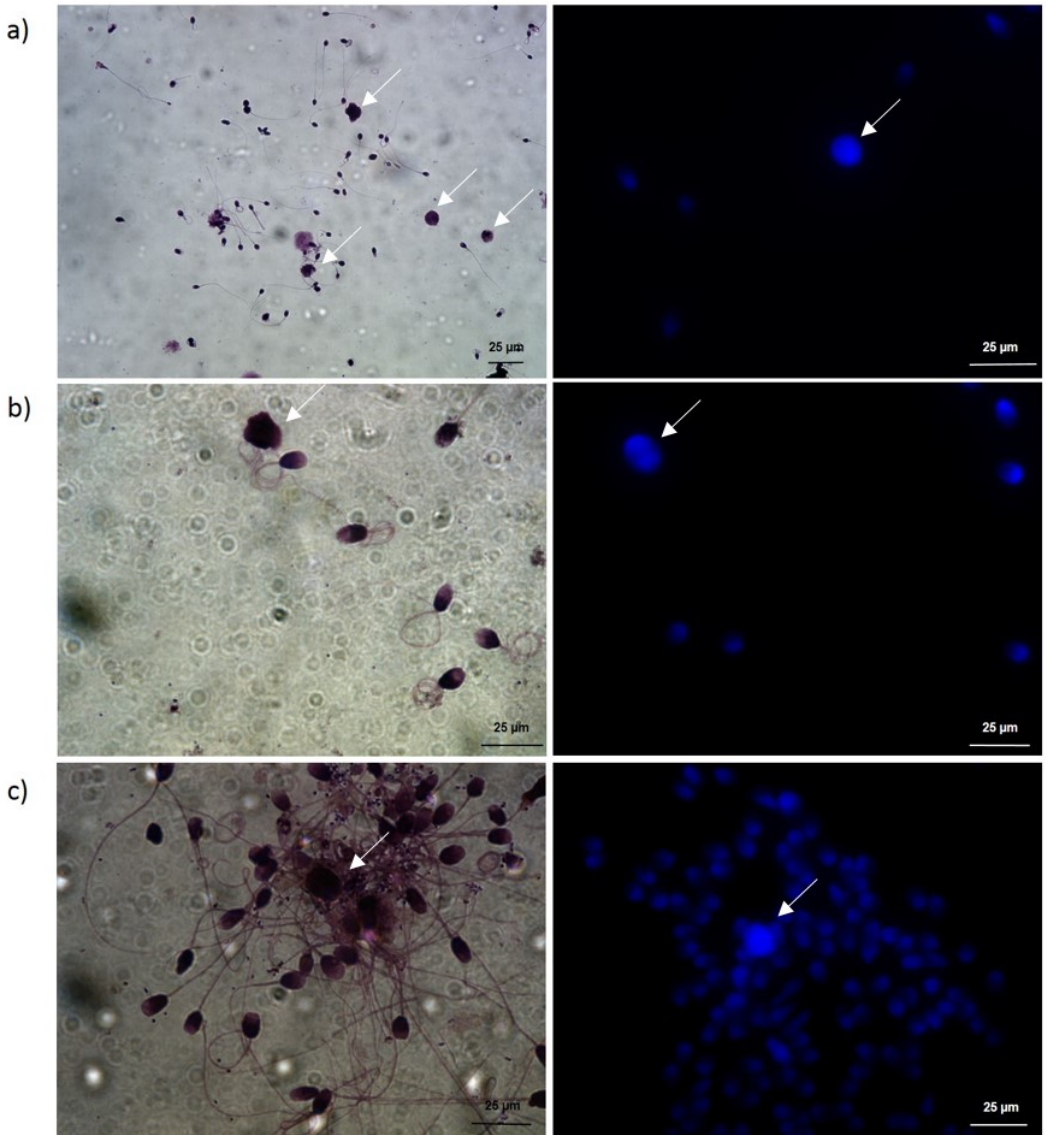
Chr	Feature	ID	HA selected	Un selected	S/US	M2 ≥ 600?
16	IL4R-203	ENST00000380925	226.74	6.18	36.71	No
12	CD163L1-001	ENST00000313599	181.50	6.29	28.84	No
12	KIAA0528-001	ENST00000333957	145.45	5.12	28.39	No
15	LACTB-001	ENST00000261893	156.72	6.15	25.48	No
6	CMAHP-001	ENST00000377993	166.70	6.88	24.22	No
9	ZCCHC7-003	ENST00000336755	164.47	6.89	23.86	No
10	null		193.38	8.56	22.58	No
12	CNOT2-007	ENST00000552231	203.17	9.17	22.15	No
22	TBC1D22A-001	ENST00000337137	240.14	12.44	19.30	No
7	null		380.43	19.73	19.28	No
6	MAN1A1-001	ENST00000368468	145.04	7.59	19.10	No
13	null		184.91	11.58	15.97	No
10	null		187.62	11.81	15.88	No
10	null		226.46	15.46	14.64	No
19	null		161.32	11.11	14.52	No
14	NDUFB1-004	ENST00000553514	166.27	11.96	13.91	No
2	PCBP1-AS1-001	ENST00000435880	152.70	11.17	13.67	No
1	KIAA1522-002	ENST00000401073	184.67	13.60	13.58	No
10	null		178.28	13.25	13.45	No
9	B4GALT1-201	ENST00000535206	172.98	13.15	13.15	No
5	STK10-001	ENST00000176763	169.86	13.65	12.45	No
22	C22orf45-001	ENST00000326341	175.34	14.43	12.15	No
7	null		389.03	33.29	11.69	No
2	null		186.01	16.21	11.47	No
19	ZNRF4-201	ENST00000222033	293.48	25.88	11.34	No
7	CUX1-003	ENST00000437600	158.64	14.27	11.12	No
7	ZNF800-004	ENST00000485577	150.79	13.65	11.04	No
2	null		144.20	13.26	10.88	No
7	CUX1-003	ENST00000437600	143.90	13.48	10.68	No
7	TTYH3-001	ENST00000258796	208.18	19.81	10.51	No
7	DOCK4-202	ENST00000352877	212.37	20.27	10.48	No
5	GPBP1-001	ENST00000424459	143.61	13.72	10.47	No
1	CAPZB-203	ENST00000401084	153.56	14.80	10.38	No
16	UBFD1-201	ENST00000219638	165.50	16.27	10.17	No
3	TGFBR2-001	ENST00000295754	158.34	16.15	9.80	No

Chr	Feature	ID	HA selected	Un selected	S/US	M2 ≥ 600?
3	NUP210-001	ENST00000254508	155.59	15.90	9.78	No
11	POU2AF1-007	ENST00000531398	163.38	16.79	9.73	No
15	ANP32A-201	ENST00000358235	165.55	17.50	9.46	No
2	AFTPH-003	ENST00000238856	166.80	17.78	9.38	No
11	NXF1-201	ENST00000294172	212.79	22.70	9.37	No
16	null		197.08	21.42	9.20	No
12	CD163L1-001	ENST00000313599	199.81	21.99	9.08	No
12	null		164.30	18.15	9.05	No
2	null		212.23	23.64	8.98	No
6	TREML4-001	ENST00000341495	142.21	15.98	8.90	No
7	null		210.84	23.72	8.89	No
3	GBE1-001	ENST00000429644	174.70	19.66	8.89	No
22	AP1B1-005	ENST00000482818	251.58	28.44	8.85	No
16	N4BP1-201	ENST00000262384	271.74	30.85	8.81	No
3	ATG7-002	ENST00000446450	168.46	19.61	8.59	No
3	HCLS1-001	ENST00000314583	222.84	26.28	8.48	No
15	ANKDD1A-001	ENST00000380230	180.29	22.50	8.01	No
15	RP11-330L19.4-006	ENST00000560837	160.05	20.28	7.89	No
15	FURIN-001	ENST00000268171	143.39	18.44	7.78	No
11	IFITM10-004	ENST00000382123	227.22	29.69	7.65	No
3	SLC6A6-002	ENST00000427436	186.64	24.59	7.59	No
17	HS3ST3B1-001	ENST00000360954	157.52	21.01	7.50	No
19	GADD45B-201	ENST00000215631	592.92	79.84	7.43	No
1	SLC45A3-002	ENST00000460934	217.45	30.78	7.07	No
1	null		276.19	39.28	7.03	No
2	null		208.51	30.25	6.89	No
1	null		456.58	66.65	6.85	No
20	SRC-203	ENST00000445403	279.14	40.90	6.83	No
6	RP3-393E18.2-001	ENST00000430078	204.24	29.94	6.82	No
9	ABL1-202	ENST00000444970	192.24	28.23	6.81	No
16	FBRS-001	ENST00000356166	217.18	32.03	6.78	No
8	PAG1-001	ENST00000220597	184.42	28.52	6.47	No

Chr	Feature	ID	HA selected	Un selected	S/US	M2 ≥ 600?
20	B4GALT5-001	ENST00000371711	356.74	55.91	6.38	No
11	CD44-003	ENST00000263398	167.81	26.32	6.37	No
3	CCDC13-001	ENST00000310232	244.60	39.50	6.19	No
17	PGS1-201	ENST00000262764	190.87	30.89	6.18	No
6	ZFAND3-001	ENST00000287218	203.02	33.44	6.07	No
9	RAPGEF1-001	ENST00000266110	175.56	29.07	6.04	No
11	MAML2-001	ENST00000524717	250.16	41.61	6.01	No
X	null		244.13	41.45	5.89	No
16	CORO1A-001	ENST00000219150	188.50	32.21	5.85	No
12	MLL2-001	ENST00000301067	215.76	36.99	5.83	No
22	BID-002	ENST00000317361	179.22	30.81	5.82	No
4	GPRIN3-001	ENST00000333209	179.65	30.94	5.81	No
22	GGA1-202	ENST00000414350	259.32	44.67	5.81	No
17	FASN-201	ENST00000306749	195.34	34.04	5.74	No
10	null		254.54	44.43	5.73	No
8	RHOBTB2-002	ENST00000519685	254.97	44.56	5.72	No
8	ZHX2-001	ENST00000314393	387.74	68.71	5.64	No
3	NUP210-001	ENST00000254508	296.19	52.85	5.60	No
1	null		375.39	67.51	5.56	No
16	TXNDC11-202	ENST00000356957	204.50	37.35	5.47	No
1	null		537.61	99.20	5.42	No
6	TREML4-001	ENST00000341495	217.97	40.34	5.40	No
12	LRMP-006	ENST00000550945	410.01	79.63	5.15	No
15	RAB8B-001	ENST00000321437	191.15	37.54	5.09	No
19	SEMA6B-202	ENST00000301293	221.45	43.78	5.06	No
8	ZHX2-001	ENST00000314393	223.84	44.57	5.02	No
17	MLLT6-001	ENST00000325718	279.79	56.03	4.99	No
10	null		247.68	50.00	4.95	No
4	TACC3-003	ENST00000485989	344.99	69.82	4.94	No
15	C15orf39-001	ENST00000360639	460.15	95.47	4.82	No
12	RILPL2-201	ENST00000280571	320.14	67.05	4.77	No
16	PIGQ-004	ENST00000409527	319.21	67.52	4.73	No
17	PHOSPHO1-001	ENST00000310544	284.94	61.24	4.65	No
11	AHNAK-011	ENST00000530124	389.53	85.83	4.54	No
20	TM9SF4-002	ENST00000398022	495.92	109.28	4.54	No

Chr	Feature	ID	HA selected	Un selected	S/US	M2 ≥ 600?
14	TNFAIP2-001	ENST00000560869	430.72	103.20	4.17	No
19	GADD45B-201	ENST00000215631	469.11	116.20	4.04	No
20	PPDPF-001	ENST00000370179	528.62	137.08	3.86	No
17	GRB2-001	ENST00000316804	430.60	117.00	3.68	No
1	SPRR3-201	ENST00000331860	92.29	346.80	0.27	No
9	RAPGEF1-001	ENST00000266110	16.04	187.00	0.09	No
17	RBFOX3-002	ENST00000338834	1519.62	0.00	#DIV/0!	Yes
12	RAB35-001	ENST00000229340	1073.15	29.32	36.60	Yes
8	PAG1-001	ENST00000220597	837.54	29.73	28.17	Yes
12	LRMP-006	ENST00000550945	608.28	23.08	26.35	Yes
7	MOSPD3-001	ENST00000223054	2285.60	446.64	5.12	Yes
7	MOSPD3-001	ENST00000223054	760.99	166.59	4.57	Yes
10	MTPAP-002	ENST00000488290	675.28	149.69	4.51	Yes
7	MOSPD3-001	ENST00000223054	750.57	182.42	4.11	Yes
16	PIGQ-004	ENST00000409527	659.94	175.61	3.76	Yes
20	PI3-001	ENST00000243924	1891.21	529.50	3.57	Yes
19	JUNB-201	ENST00000302754	1626.97	457.67	3.55	Yes
19	IFI30-201	ENST00000407280	1425.43	407.50	3.50	Yes
20	TGM2-002	ENST00000361475	722.68	207.24	3.49	Yes
14	ZFP36L1-003	ENST00000555997	921.07	281.77	3.27	Yes
19	PPP1R15A-201	ENST00000200453	797.91	245.93	3.24	Yes
14	NFKBIA-001	ENST00000216797	1231.26	380.73	3.23	Yes
1	WDTC1-201	ENST00000319394	1605.66	527.52	3.04	Yes
16	PRM2-001	ENST00000241808	112932.70	95473.23	1.18	Yes
1	HFM1-002	ENST00000370425	327456.80	309757.30	1.06	Yes
16	PRM2-001	ENST00000241808	522273.00	567659.60	0.92	Yes
13	null		253942.22	303906.88	0.84	Yes
5	DHFR-001	ENST00000439211	73235.16	95689.96	0.77	Yes
16	PRM2-001	ENST00000241808	142147.84	193087.90	0.74	Yes
1	SMCP-001	ENST00000368765	40266.57	55924.11	0.72	Yes
21	null		22188.37	48853.45	0.45	Yes

Appendix VII: Monitoring for somatic cell contamination



Appendix VII-1: Giemsa and DAPI staining's. a) swim up fractions; b) HA-bound spermatozoa and c) DGC 90% and 45% fraction. Somatic cells (indicated by an arrow) were found in all spermatozoal selection approaches for human. Scale Bar: 25 µm