UNIVERSITY OF LEEDS

This is a repository copy of *Potential sperm contributions to the murine zygote predicted by in silico analysis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/146679/

Version: Accepted Version

Article:
Ntostis, P, Carter, D, Iles, D et al. (3 more authors) (2017) Potential sperm contributions to the murine zygote predicted by in silico analysis. Reproduction, 154 (6). pp. 777-788. ISSN 1470-1626

https://doi.org/10.1530/REP-17-0097

(c) 2017, Society for Reproduction and Fertility. This manuscript version has not been copy-edited, formatted or proofed. Consequently, Bioscientifica accepts no responsibility for any errors or omissions it may contain. The definitive version is now freely available at [https://doi.org/10.1530/REP-17-0097]

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Reproduction

Potential sperm contributions to the murine zygote predicted by in silico analysis.

Journal:	*Reproduction*
Manuscript ID	REP-17-0097.R2
mstype:	Research paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Ntostis, Panagiotis; University of Leeds, Leeds Institute of Cardiovascular and Metabolic Medicine; National and Kapodistrian University of Athens - Faculty of Medicine, Department of Medical Genetics Carter, Deborah; University of Leeds, St James' Biomedical Services (SCIF) Iles, David; University of Leeds, Leeds Institute of Cardiovascular and Metabolic Medicine Huntriss, J.D.; University of Leeds, Leeds Institute of Cardiovascular and Metabolic Medicine Tzetis, Maria; National and Kapodistrian University of Athens - Faculty of Medicine, Department of Medical Genetics Miller, David; University of Leeds, Leeds Institute of Cardiovascular and Metabolic Medicine
Keywords:	sperm RNA, RNA sequencing, in silico analysis, zygote formation, Gene expression

SCHOLARONE™
Manuscripts

For Peer Review Only

Title: Potential sperm contributions to the murine zygote predicted by *in silico* analysis.

Authors and affiliations: Panagiotis Ntostis[1,3], Deborah Carter[2], David Iles[1], John Huntriss[1], Maria Tzetis[3], David Miller[1] *

[1] Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, Leeds, West Yorkshire, LS2 9JT, UK

[2] Leeds Institute of Molecular Medicine, University of Leeds, Leeds, West Yorkshire, LS9 7TF, UK

[3] Department of Medical Genetics, St. Sophia's children hospital, School of Medicine, National and Kapodistrian University of Athens, Athens, Attiki, 11527, Greece

Correspondence should be addressed to [1]D Miller; Email: d.miller@leeds.ac.uk

Short Title: Paternal / maternal interaction in the zygote.

Keywords: Sperm RNA, RNA sequencing, in silico analysis, zygote formation, gene expression.

For Review Only

Abstract

Paternal contributions to the zygote are thought to extend beyond delivery of the genome and paternal RNAs have been linked to epigenetic transgenerational inheritance in different species. In addition, sperm-egg fusion activates several downstream processes that contribute to zygote formation, including PLC zeta-mediated egg activation and maternal RNA clearance. Since a third of the preimplantation developmental period in the mouse occurs prior to the first cleavage stage, there is ample time for paternal RNAs or their encoded proteins potentially to interact and participate in early zygotic activities. To investigate this possibility, a bespoke next generation RNA sequencing pipeline was employed for the first time to characterise and compare transcripts obtained from isolated murine sperm, MII eggs and pre cleavage stage zygotes. Gene network analysis was then employed to identify potential interactions between paternally and maternally derived factors during the murine egg to zygote transition involving RNA clearance, protein clearance and post-transcriptional regulation of gene expression. Our *in silico* approach looked for factors in sperm, eggs and zygotes that could potentially interact co-operatively and synergistically during zygote formation. At least five sperm RNAs (*Hdac11*, *Fbxo2*, *Map1lc3*, *Pcbp4* and *Zfp821*) met these requirements for a paternal contribution, which with complementary maternal co-factors suggest a wider potential for extra-genomic paternal involvement in the developing zygote.

For Review Only

Introduction

Assuming fertilisation is successful, spermatozoal entry into the egg triggers a series of events that ends with the transformation of the terminally differentiated egg into the totipotent zygote. Alongside the paternal genome, the sperm also delivers non-genomic factors including the microtubule organising centre or centriole (excepting rodents), the oocyte-activating factor, PLC-zeta (Saunders et al, 2002; Barroso, et al. 2009) and a complex repertoire of RNAs to the egg (Ostermeier, et al. 2004, Yuan, et al. 2015). Both sperm and egg are transcriptionally silent (Braun 2000, Richter and Lasko 2011) and the egg-to-zygote transition (EZT) occurs in the absence of transcription (Evsikov, et al. 2006). While somatic nuclear cloning (Gurdon and Melton 2008) and the generation of viable gynogenetic mice (Kono, et al. 2004) suggest that maternal factors alone are sufficient to guide early embryo development, these processes are grossly inefficient and structural or signaling factors from the sperm may complement maternal factors that could participate in and aid the early programming of embryonic development (Jodar, et al. 2015, Miller 2015).

Paternal RNAs can epigenetically affect transgenerational inheritance through specific small non-coding RNAs (sncRNAs) and associated RNA-binding proteins (Chen, et al. 2016, Rodgers, et al. 2015). In *Caenorhabditis elegans,* a hybrid strain crossing showed that approximately 10% of embryonic RNA is of paternal origin with functional importance during EZT and possibly embryogenesis (Stoeckius, et al. 2014b). An equivalent paternal contribution to the mammalian zygote will be small by comparison, but evidence of the potential for sperm RNAs (or their translated proteins) to contribute to and participate in zygote formation is strong and worthy of further investigation.

For Review Only

Early molecular processes in the zygote can be classified into three main categories including maternal clearance, chromatin remodeling and eventually zygotic genome activation (ZGA) (Lee, et al. 2013). Maternal clearance is the process of removing maternal factors including RNAs and proteins essential for oogenesis that become surplus to requirements after fertilisation (Tadros and Lipshitz 2009). Post-transcriptional regulation plays a role during EZT and includes the destruction of maternal mRNAs guided by their 3' untranslated (3' UTR) sequences (Giraldez 2010). Compared with approximately 2000 proteins reported in pre cleavage stage zygotes of *M. musculus,* over 3500 proteins have been identified in metaphase II eggs (Wang, et al. 2010, Yurttas, et al. 2010). During the embryonic development, this removal of maternal factors is guided mainly by ubiquitin-dependent degradation pathways and by autophagy (Marlow 2010).

While transcriptionally inert, MII eggs and zygotes are likely to be translationally active (Fang, et al. 2014, Potireddy, et al. 2006), leaving open the possibility for sperm RNAs to be translated into proteins following their introduction to the egg (Fang, et al. 2014). We reasoned that a potential non-genomic paternal contribution would most likely involve interactions with maternal factors responsible for the regulation of gene expression prior to the EZT and the clearance of maternal factors prior to embryonic genome activation. The main objective of the study, therefore was to see if *in silico* analysis of RNA sequencing data obtained from sperm, MII eggs and pre cleavage stage zygotes (PCZ but henceforth referred to as zygotes) using an identical bespoke protocol, could highlight potential interactions between paternal and maternal cofactors brought together by fertilisation. Herein, we focus on five, full length mRNAs present at high levels in murine sperm that with associated maternal cofactors, fit the requirements for a potential paternal, non-genomic metabolic contribution to the zygote.

For Review Only

Materials and Methods

Study design

RNAs isolated from pooled murine spermatozoa, single MII eggs and single zygotes were sequenced and compared. To help identify paternal RNAs with the potential to participate in EZT events, we looked firstly for candidate RNAs that were highly represented in sperm, thus increasing the probability of being delivered to and translated by the zygote or being translated into protein during the late stages of spermatogenesis and delivered to the MII egg at fertilisation. Secondly, considering the highly fragmented nature of sperm RNA, no less than 80% of the exonic regions of at least one gene isoform of a 'candidate' paternal cofactor had to be covered by RNA-Seq reads, giving a greater potential for the RNA to be functional. Thirdly, sperm RNAs with good sequence coverage were only considered further if their ontological descriptions suggested functions other than spermatogenesis.

A bespoke identical library preparation method and sequencing pipeline was applied to all samples allowing accurate comparative assessment of RNAs across the different samples. The library kit used (Ovation single cell RNA-Sequencing system, NuGEN, CA, USA), has a mix of oligo-dT and random primers targeting a wider range of transcripts, including those with varying poly(A) tail lengths, typically encountered in gamete and zygotic mRNAs (Paranjpe, et al. 2013). Cytoscape's GeneMANIA module (see below) was then employed for the *in silico* analysis investigating potential interactions between gene products of paternal and maternal origin (Warde-Farley, et al. 2010).

Ethics

Experiments involving the use of animals were regulated under the Home Office, UK Animals Scientific Procedures Act (ASPA) under license service PPL 40/3391 approved by the University of Leeds AWERC (Animal Welfare Ethical Review

reproduction@bioscientifica.com
For Review Only

Committee). All animals were culled using cervical dislocation in accordance to Schedule 1 of the ASPA.

Mouse gamete and zygote harvest

Groups of C57BL/6J females were super-ovulated with a 5 IU dose of pregnant mares' serum gonadotropin (PMSG) (Sigma Aldrich, MO, USA) injected intraperitoneally (I/P) on day 1, followed at day 3 by a 5IU dose of human chorionic gonadotropin (HCG) (Sigma Aldrich, MO, USA) I/P and mated with vasectomised males to provide MII eggs. The zygote groups were mated to proven C57BL/6J stud males immediately after hCG dosing and checked the following day for post coital plugs. Plugged females were pooled and used for zygote harvest. Both zygote and MII egg groups were sacrificed on day 4. Oviducts from the zygote and egg groups were harvested separately and suspended in M2 media (Sigma Aldrich, MO, USA). Dissected oviducts were placed into a pre-heated dish of synthetic Human Tubal Fluid (HTF) media (Irvine scientific, CA, USA) with bovine serum albumin (BSA) (Sigma Aldrich, MO, USA). Cumulus masses were released into the HTF/BSA medium and transferred into a drop of hyaluronidase (Sigma Aldrich, MO, USA) following which, a wide bore pipette was used to strip the eggs and zygotes of their cumulus cells. These were in turn collected by mouth pipette and washed through sequential drops of M2 media (Sigma Aldrich, MO, USA).

Sperm harvest

The epididymides of fertile C57BL/6J males were dissected out and transferred into pre-warmed HTF (Irvine Scientific, CA, USA). Using a sterile 26G needle, small incisions were made in the cauda and sperm were allowed to swim out before collection by gentle aspiration. Spermatozoa were washed in HTF (Irvine, CA, USA), filtered through an 80-micron mesh (Sigma Aldrich, MO, USA) and centrifuged at 500xg prior to resuspension and centrifugation through a two-layer (65%-50%)

For Review Only

discontinuous percoll gradient (GE Healthcare Biosciences, Uppsala, Sweden) at 300xg, employing the *ProInsert* technology (Nidacon International AB, Gothenburg, Sweden) to facilitate the selective isolation of pelleted spermatozoa while preventing possible contamination by somatic cells (Fourie, et al. 2012). Spermatozoa were pelleted at 500xg and washed twice in Dulbecco's phosphate-buffered saline (DPBS) (Thermo Scientific, MA, USA) Approximately 1 million spermatozoa were harvested before the second wash and Giemsa stained (Sigma Aldrich, MO, USA) to visually confirm lack of other cell types using a Leica Leitz DMRB microscope (Mazurek Optical Services, Southam, UK).

RNA isolation and library construction

Sperm RNA was extracted using the method described by Goodrich (Goodrich, et al. 2013) with modifications. Briefly, 10^7 spermatozoa were placed in RLT buffer (Qiagen, Hilden, Germany) with 1.5% β-mercaptoethanol (Sigma Aldrich, MO, USA) and 0.5mm nuclease free stainless steel beads. Following homogenisation with a DisruptorGenie™ cell disruptor (Thermofisher Scientific, MA, USA), an equal volume of chloroform was added followed by centrifugation at 12,000xg (4^oC), allowing recovery of the RNA. Prior to library construction, any residual genomic DNA was removed from the samples by digestion with Turbo DNase (Thermofisher Scientific, MA, USA) following the manufacturer's instructions. Quantitative Real-Time PCR (qRT-PCR) using *Prm2* and *Map1lc3a* intron spanning primers with SybrGreen PCR mastermix (Applied Biosystems, CA, USA) was employed to monitor for DNA contamination.

Sperm RNA quality assessment was carried out using the RNA-6000 pico assay (Agilent, CA, USA) on a 2100 Bioanalyzer (Agilent, CA, USA), where the absence of clearly defined peaks from 18S and 28S ribosomal RNAs (low RIN score) indicates corresponding absence of contaminating somatic cell RNA (supplementary Figure 1).

For Review Only

As additional QC, Real-Time qPCR primers for the Melanoma-Associated Antigen D2 (*Maged2*) were used to confirm potential contaminating somatic cell RNA in these preparations (principally from Leydig and Sertoli cells; Chalmel, et al. 2007). Only spermatozoal cDNAs shown to be free of genomic DNA and somatic RNA contamination were used for library construction.

Mouse eggs and zygotes were processed at the single cell level. Each cell was transferred by mouth pipette into lysis buffer, after being immersed in DPBS (LIFE Technologies, USA) in a washing step containing 0.1% BSA (Sigma-Aldrich, MO, USA). Following first and second strand synthesis and processing using the Ovation single cell RNA-Seq system (NuGEN, USA), Illumina adaptor sequences were ligated to the sperm, egg and zygote cDNAs. Two rounds of library amplification were carried out and the fragment distribution was checked using the Agilent high sensitivity DNA assay on the 2100 Bioanalyser (Agilent Technologies, CA, USA). The libraries were quantified using Picogreen assay (Thermo Scientific, MA, USA) on a FLUOstar Galaxy plate reader (MTX Lab Systems, USA) and pooled. The Illumina HiSeq 2500 and 3000 platforms were employed for RNA-Seq.

Bioinformatics Analysis

Spermatozoa, MII eggs and zygotes from a minimum of three biological replicates each were sequenced using either 50bp (single-ended) or 150bp paired-end reads. RNA-Seq data underwent automated adapter and quality trimming using Trim Galore! v0.4, ignoring reads with MAPQ<20 (Krueger 2015). The reads above this threshold were mapped to the *Mus musculus* reference genome (mm10) using the subjunc function of the Rsubread package version 1.20.3 (Liao, et al. 2013b). The output BAM format files were sorted using Samtools version 1.3 (Li, et al. 2009) and duplicate reads removed using the Picard MarkDuplicates tool version 2.1.1 (Broad Institute. (2010), available online at http://broadinstitute.github.io/picard). BedGraph

reproduction@bioscientifica.com
For Review Only

and bigwig files were generated using Bedtools version 2.25.0 (Quinlan and Hall 2010), and the function bedGraphToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/macOSX.x86_64/bedGraphToBigWig). After removal of duplicate unmapped and incorrectly paired reads using Samtools version 1.3 (Li, et al. 2009), the reads were visualized on the UCSC genome browser (Kent, et al. 2002). The numbers of reads assigned to genomic features were counted using the featureCounts function of Rsubread (Liao, et al. 2013a). For paired-end libraries, we required both read mates to be uniquely mapped in the correct orientation. All remaining options were set to featureCounts default. Differential RNA representation in MII egg and zygote RNA-Seq libraries was tested using the edgeR exact test (Robinson, et al. 2010) and only genes represented at levels ≥10 counts-per-million reads (CPM) in at least 6 out of 7 MII and zygote libraries were included in the downstream analysis. The only exception to this rule was for five maternal transcripts with reads just below 10 CPM, represented across all exons, that were also included. Data normalisation was based on the trimmed mean of M value (TMM) using the calcNormFactorsfunction (Robinson and Oshlack 2010).

Ontological analysis, gene networks and molecular interactions

Ontological descriptions of RNAs from sperm, MII eggs and zygotes were derived by DAVID v6.8 (Huang, et al. 2009), with a subsequent focus on biological processes. Gene networks involving candidate spermatozoal and maternal factors were identified by the Cytoscape module GeneMANIA v.3.4.1; (Warde-Farley, et al. 2010). GeneMANIA uses publically available data sets, encompassing physical and molecular interactions, co-expression, co-localisation and molecular pathways.

NGS validation using quantitative real-Time PCR

For Review Only

Following first-strand cDNA synthesis, the cDNA of mouse MII egg, zygote, spermatozoal and testicular RNA (positive control) was amplified by long distance PCR, using the SMART-Seq v4 ultra low input RNA kit (Clontech, USA). Validatory quantitative real-time PCR was carried out as required using gene-specific primers (supplementary Table 1) and SYBR green on an ABI 7900HT Real-time PCR system (Applied Biosystems, CA, USA) over 40 cycles according to the manufacturer's instructions. The annealing temperature per primer pair ranged between 59^0C and 62^0C.

Results

1. RNA characterisation and ontological profiles

The average correctly paired and mapped reads per sperm RNA sample was calculated at 20 ± 2 x 10^6. The average number for MII eggs was 18 ± 1.5 x 10^6 and for the zygotes 20 ± 1.5 x 10^6 per sample. RefSeq IDs for sperm, MII eggs and zygotes, alongside differentially expressed RNAs using EdgeR for MII eggs and zygotes are listed in the supplementary info (RNA lists)**.** While we cannot be certain that RNAs common to sperm and zygotes originated in the fertilising sperm, our initial approach was to look for sperm RNAs that were absent in MII eggs but present in zygotes. Figure 1**A** shows Venn diagrams for overlaps between sperm, MII egg and zygote RNAs ≥10 CPM. In aggregate, 5,368, 5,148 and 1,918 RNAs were reported, respectively from MII eggs, zygotes and sperm with 75 shared between sperm and zygotes that were either absent altogether or present in MII eggs at well below threshold reporting levels. The same sperm list compared with EdgeR normalised reads for MII egg and zygote RNAs yielded 56 RNAs shared between sperm and zygotes (Figure 1**B**). Closer scrutiny of the read data, however, showed all but four of these 'shared' RNAs were detected in MII eggs albeit at low levels of

For Review Only

expression (<10 CPM). The four absent in MII eggs were more fragmented in sperm than in zygotes, suggesting they were not sperm-specific.

An alternative approach focused simply on highly abundant sperm RNAs with good exon coverage that were essentially absent in both MII eggs and zygotes. To help narrow down the list of hundreds of possible RNAs to pursue in this regard, functional annotation clustering (supplementary info; FAC sheets) was employed to provide a general overview of MII egg and zygote RNAs using the lists of differentially expressed MII egg and zygote RNAs generated by EdgeR alongside the list of sperm RNAs selected on the basis of high representation and good exon coverage. A graphical representation of the numbers of genes in the main ontological annotations (biological processes) for sperm, MII eggs and zygotes is shown in supplementary Figure 2.

As shown in Table 1 and in supplementary info (BP; EdgeR sheet), the expected top sperm annotation related to spermatogenesis as a differentiation process, with associated weaker enrichments in processes associated with lipid metabolism and DNA condensation. Enrichment for annotation relating to the control of transcription dominated the ontological descriptions for both MII eggs and zygotes, which was expected considering the similarity between them (Table 1 and supplementary info; BP EdgeR sheet). Focusing on differential expression between the two, however, revealed interesting differences (supplementary info; BP EdgeR MII or PCZ Up sheet). MII eggs showed specific enrichments in activities relating to mRNA processing, while the cell cycle showed the strongest enrichment in zygotes. Processes relating to ubiquitination and transcription were more apparent in differentially up-regulated RNAs from zygotes but not MII eggs, suggesting that clearance activity and perhaps renewed RNA processing triggered by fertilisation may have already commenced in zygotes at the time of harvesting.

For Review Only

Using the functional annotation of highly expressed RNAs as guidance alongside closer inspection of the selected gene lists from sperm, MII eggs and zygotes, potential interacting partners relating to clearance of maternal factors were revealed. Considering the RNAs' high expression in sperm compared with MII eggs and zygotes, their exclusion from the dominant spermatogenesis ontology and their relative freedom from fragmentation as assessed by exon coverage, five sperm RNAs were selected for follow up (Table 2). These include the histone deacetylase 11 (*Hdac11*), the Rbx1-SCF E3 ubiquitin-ligase component F-box protein 2 (*Fbxo2*), the microtubule-associated protein 1A/1B light chain 3A (*Map1lc3a*), the poly (rC)-binding protein 4 (*Pcbp4*) and the zinc finger protein 821 (*Zfp821*). These five sperm RNAs were in turn interrogated using GeneMANIA for all known interacting partners, which returned approximately 100 genes of which 37 were either present in the up-regulated zygotic transcripts or in the list of shared (MII & zygote) maternal RNAs (Table 2). Together, these paternal and maternal RNAs comprise the gene network profile shown in Figure 2. The network's functional annotation was dominated by strong enrichment in processes related to ubiquitin-mediated degradation pathways (supplementary info; BP EdgeR sheet), reflecting the ontology of up-regulated transcripts in zygotes.

2. Predicting and providing evidence for potential parental interactions

Network analysis (Figure 2) suggested that paternal (♂) and maternal (♀) cofactors could interact in pathways leading to EZT. An example is illustrated in Figure 3 for the gamete-specific cofactors of the multiple component SCF E3 ubiquitin ligase complex which includes *Fbxo2* (also known as *Fbs1*♂; **A**) alongside Cullin 1 (*Cul1*♀; **B**), Ring Box 1 (*Rbx1*♀; **C)** and S-Phase Kinase-Associated Protein 1A (*Skp1a*♀; **D**). These genes are indicated in Figure 2 by boxes. Note that reads covering all exons for *Fbxo2*, were strongly represented in the sperm RNA libraries but with few or no reads from either MII egg or zygote libraries. In contrast, with the exception of *Fbxo2*,

reproduction@bioscientifica.com
For Review Only

RNAs encoding the other cofactors of the SCF-E3 ubiquitin ligase complex were highly represented in both MII eggs and zygotes but not in sperm. Additional predicted interactions between *Fbxo2*♂ and the maternally expressed *Fbxo5*♀ and *Fbxo34*♀ were also suggested (Table 2 and Figure 2; boxes). Real-time qPCR confirmed the expression of paternal factors in sperm and testis (Figure 4) and although products were generated for Map1lc3a from all sources (panel **A**), Ct data confirmed that the RNA was considerably more abundant in sperm (see panel **B**). A 142bp product from *Maged 2* was only detected in RNA from eggs and testis, indicating that sperm libraries were free of contamination by RNAs from testis-derived somatic cells. All PCR products were obtained from samples after 40 PCR cyles and so the products shown in panel **A** are only qualitative. The corresponding Ct values give more quantitative assessments.

UCSC tracks are shown in supplementary Figure 3 for a number of additional, GeneMANIA suggested potential paternal-maternal interacting cofactors. *Hdac11*♂ (**A**), which was highly expressed in sperm, could interact withthe Mitotic Checkpoint Serine/Threonine Kinase B, (*Bub1*♀; **B**) and cell division cycle protein 20 (*Cdc20*♀; Figure 2; ovals). GeneMANIA also suggested potential co-localisation and co-expression between *Hdac11*♂, *Hdac2*♀ and *Hdac8* (Figure 2; ovals) with *Hdac11*♂ and *Hdac2*♀ having shared protein domains. Predicted interactions with the *Nelfcd*♀ and *Aamp*♀ were also highlighted (Figure 2; ovals). Reads from a long terminal repeat (LTR♂) region located within the 8th intron of *Hdac11*, which could potentially be expressed independently of *Hdac11*♂ RNA were also noted (supplemental Figure 3A).

The ubiquitin-like protein Microtubule Associated protein 1, light chain 3 alpha (*Map1lc3a*♂; **C**) is involved in autophagosome formation and GeneMania indicated functional interactions with several maternal factors, including *Map1b*♀ (**D**), *Atg3*♀

For Review Only

(**E**) and *Atg10*♀ (Figure 2, boxes with course dashed lines) of which *Atg3*♀ and *Map1b*♀ showed high levels of expression in MII eggs and zygotes with good exonic representation and were absent in sperm. In addition, *Pcbp4*♂ (**F**), which may complement the heterogeneous nuclear ribonucleoprotein K (*Hnrnpk*♀; **G**) and Quaking (*Qk*♀; **H**; and Figure 2, pentangles) were highly represented in MII eggs and zygotes but not in sperm. The corresponding maternal proteins are involved in post-transcriptional regulation of gene expression, protein degradation and the cell cycle. *Pcbp4* ♂ also has predicted interactions with *Pcbp1*♀ and co-localises with the *Pcbp2*♀ isoform. Finally, GeneMANIA reported predicted interactions between *Zfp821*♂ (**I**), which may be involved in transcriptional regulation, *Fchsd2*♀ and *Rimlkb*♀ (Figure 2; boxes with fine dashed lines). *Zpf821*♂ is highly expressed in sperm but not MII eggs and zygotes.

Discussion

Existing sequencing germ line and zygote datasets are not fully complementary and are therefore difficult to compare (they either omit sperm or MII eggs or zygotes from their analysis) and are derived from library construction methods that differ between the various cell types (Johnson, et al. 2015, Tang, et al. 2010, Xue, et al. 2013). To avoid introducing methodological effects and bias, we used a bespoke pipeline that included construction of our own libraries for sequencing and analysis. Sperm contain far less RNA than either MII eggs or zygotes; therefore, sperm libraries were unavoidably derived from sperm-specific pooled samples while MII egg and zygote libraries were from individual cells. The equivalent read counts obtained from the three sources demonstrate the care taken to assure quantitative equivalence of input RNA. We found, however, that relying on comparisons between the three to select sperm-specific factors were problematic, because although representation may have been too low to report the RNA as present in MII eggs or zygotes (for example), we frequently encountered reads indicating fragmented RNA in both regardless. We

reproduction@bioscientifica.com

For Review Only

focused, therefore, on highly abundant RNAs with full-length transcripts in sperm, MII eggs and zygotes and with demonstrably reciprocal representation (in sperm but not MII eggs or zygotes and the reverse) following inspection of RNA-Seq tracks on the UCSC browser and where necessary, confirming by qRT-PCR.

In silico analysis of our RNA sequencing data supports the possible complementation of maternal with paternal factors introduced at fertilization. Five highly expressed sperm RNAs were considered based on their relative low abundance or absence in MII eggs or zygotes. All factors potentially interacting with translated products from these RNAs were mapped out using the pathway and network analyses tools in GeneMANIA. These factors were then matched to corresponding maternal cofactors to help identify those with a greater potential to participate in EZT pathways. As the predicted interactions were more likely to be between proteins, where possible we checked for a corresponding proteomic record of the RNAs in question (Skerget, et al. 2015, Wang, et al. 2013, Wang, et al. 2010). Sperm RNAs could either be translated into proteins during late spermatogenesis or if delivered to the MII egg, in the pre cleavage stage zygote (Fang, et al. 2014). We also searched for reproductive effects of existing knockout models for the corresponding genes of parental factors where available (supplementary Table 2).

The current study provides evidence of a novel role for paternally introduced factors in murine zygotic RNA/protein clearance (Sato and Sato 2013, Stitzel and Seydoux 2007). The RBX1-SCFE3 ubiquitin ligase complex, for example, plays an important role during gametogenesis and mouse embryogenesis, catalysing the ubiquitination of proteins during cytoplasmic turnover, which are then destined for proteasomal degradation (Jia and Sun 2009, Sato and Sato 2013). The F-box family includes FBXO2, which is an E3 ligase adaptor protein targeting glycosylated proteins for degradation. Our network analysis, showed that all RBX1-SCF E3 ubiquitin ligase

For Review Only

components but one (*Fbxo2*♂) were maternally expressed (McCall, et al. 2005) and on fertilisation, FBXO2♂ could complete the complex and be active in the EZT. The FBXO2 protein is also present in mature sperm (Wang, et al. 2013) and has been linked to idiopathic male infertility (Bieniek, et al. 2016). Similarly, MAP1LC3A♂ is a ubiquitin-like modifier (Cherra, et al. 2010) with potential autophagic interactions with ATG3♀ and MAP1B♀. MAP1B protein is also present in eggs and zygotes. Both *Atg3* and *Map1b* KO studies show lethality one day after birth (supplementary Table 2).

Quaking (QK)♀, HNRNPK♀ and PCBP1/2/4♀/♂ showed predicted *in silico* interactions as part of the post-transcription regulatory process. In *C. elegans,* the first wave of degradation of egg factors involves PES4 (Stoeckius, et al. 2014a), a member of the PCBP family of RNA-binding proteins that post-transcriptionally regulate alternative polyadenylation at a global level (Ji, et al. 2013). Both members of the Poly-(rC) binding protein family, PCBP4♂ and PCBP1/2♀, detected in our analysis, are mammalian orthologues of the nematode PES4 protein. Potential interactions between HDAC11♂, BUB1B♀ and CDC20♀ were predicted by our analysis and both *Hdac11* and *Bub1b* were detected in high levels in sperm and eggs/zygotes, respectively. HDAC11♂ is involved in epigenetic repression, transcriptional regulation and embryonic development (Bagui, et al. 2013, Haberland, et al. 2009, Sahakian, et al. 2015). HDAC11 activates BUB1B by deacetylation, which in turn lifts the inhibition of the CDC20/APC complex, activating its ubiquitin ligase activity (Watanabe, 2014). Although the fertility rate of mice homozygous for *Hdac11* deletion has not been reported, *Cdc20* KO mice showed 2-cell embryo arrest and *Bub1b* KO mice show developmental arrest in early gestation (E8.5) (supplementary Table 2).

For Review Only

A particularly interesting finding was the expression of a long terminal repeat (LTR) transposable element, located entirely within the 8th intron of *Hdac11* (intragenic) in sperm, which was absent in both MII eggs and zygotes. LTR RNAs are expressed abundantly in mouse eggs and zygotes where they are thought to augment the regulation of host gene expression (Göke, et al. 2015, Peaston, et al. 2004). Spermatozoal LTRs transferred into the oocyte during fertilization, could lead to new retrotransposition events and possibly genetic alterations (Kitsou, et al. 2016). Paternally derived *Hdac11* LTRs may have maternal targets that together participate in the regulation of zygotic gene expression.

In conclusion, our data supports the argument favouring extra-genomic contributions by the fertilising sperm to the zygote. In addition to the inheritance of acquired traits propagating transgenerationally via sperm RNA (Chen, et al. 2016, Gapp, et al. 2014), our data and its analysis provides evidence for a role of paternal RNAs or proteins in maternal clearance during EZT. Sperm may deliver signals or factors that can potentially interact locally with maternal cofactors and act, perhaps as a 'last minute' checkpoint or gateway for embryonic genome activation (EGA). The hypothesis of confrontation and consolidation with regard to the uniquely invasive nature of sperm entry to the egg falls into this latter category (Bourc'his and Voinnet 2010, Miller 2015). Figure 5 shows a model for how a sperm factor introduced at fertilisation might complement a maternal cofactor or pathway required for the EZT. At least one such sperm-borne factor, PLC zeta♂ is already known to activate the oocyte (Saunders, et al. 2002). A similar approach to ours could be employed to investigate paternal/maternal interactions in humans. However, to confirm the potential biological relevance of the suggested interacting cofactors reported in this study, additional work such as RNA knock down upon or conditional gene knock out prior to fertilization would require the mouse model. In view of the renewed concern over rising human male infertility (Levine 2017) and the rapid rise and expansion of

For Review Only

infertility treatment by ICSI, further research into extra-genomic paternal contributions using model systems is fully justified.

For Review Only

Declaration of interest

The authors have no conflict of interest to declare.

Funding

This work was supported by a FP7, Marie Curie Marie Curie International Reproductive Biology Early Research Training (Reprotrain) network fellowship to Panagiotis Ntostis (Grant number 289880, 2012).

Ntostis carried out the experimental work and wrote the original manuscript, and Ntostis and Iles carried out the bioinformatic analysis reported in this paper. Carter assisted with the animal work, Huntriss and Tzetis provided critical reading of the manuscript and Miller and Iles designed the original experimental plan and Miller revised and wrote the final manuscript.

For Review Only

References

Bagui, TK, SS Sharma, L Ma, and WJ Pledger 2013 Proliferative status regulates HDAC11 mRNA abundance in nontransformed fibroblasts. *Cell Cycle* **12** 3433-3441.

Barroso, G, C Valdespin, E Vega, R Kershenovich, R Avila, C Avendano, and S Oehninger 2009 Developmental sperm contributions: fertilization and beyond. *Fertil Steril* **92** 835-848.

Bieniek, JM, AP Drabovich, and KC Lo 2016 Seminal biomarkers for the evaluation of male infertility. *Asian journal of andrology* **18** 426.

Bourc'his, D, and O Voinnet 2010 A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* **330** 617-622.

Braun, RE 2000 Temporal control of protein synthesis during spermatogenesis. *Int J Androl* **23 Suppl 2** 92-94.

Chalmel, F, AD Rolland, C Niederhauser-Wiederkehr, SS Chung, P Demougin, A Gattiker, J Moore, J-J Patard, DJ Wolgemuth, and B Jégou 2007 The conserved transcriptome in human and rodent male gametogenesis. *Proceedings of the National Academy of Sciences* **104** 8346-8351.

Chen, Q, W Yan, and E Duan 2016 Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nature Reviews Genetics*.

Cherra, SJ, SM Kulich, G Uechi, M Balasubramani, J Mountzouris, BW Day, and CT Chu 2010 Regulation of the autophagy protein LC3 by phosphorylation. *The Journal of cell biology* **190** 533-539.

Evsikov, AV, JH Graber, JM Brockman, A Hampl, AE Holbrook, P Singh, JJ Eppig, D Solter, and BB Knowles 2006 Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes & development* **20** 2713-2727.

Fang, P, P Zeng, Z Wang, M Liu, W Xu, J Dai, X Zhao, D Zhang, D Liang, X Chen, S Shi, M Zhang, L Wang, Z Qiao, and H Shi 2014 Estimated diversity of messenger RNAs in each murine spermatozoa and their potential function during early zygotic development. *Biol Reprod* **90** 94.

Fourie, J, N Loskutoff, and C Huyser 2012 Elimination of bacteria from human semen during sperm preparation using density gradient centrifugation with a novel tube insert. *Andrologia* **44** 513-517.

Gapp, K, A Jawaid, P Sarkies, J Bohacek, P Pelczar, J Prados, L Farinelli, E Miska, and IM Mansuy 2014 Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nature neuroscience* **17** 667-669.

For Review Only

Giraldez, AJ 2010 microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. *Current opinion in genetics & development* **20** 369-375.

Göke, J, X Lu, Y-S Chan, H-H Ng, L-H Ly, F Sachs, and I Szczerbinska 2015 Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell stem cell* **16** 135-141.

Goodrich, RJ, E Anton, and SA Krawetz 2013 Isolating mRNA and small noncoding RNAs from human sperm. *Spermatogenesis: Methods and Protocols* 385-396.

Gurdon, JB, and DA Melton 2008 Nuclear reprogramming in cells. *Science* **322** 1811-1815.

Haberland, M, RL Montgomery, and EN Olson 2009 The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews Genetics* **10** 32-42.

Huang, DW, BT Sherman, and RA Lempicki 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4** 44.

Ji, X, J Wan, M Vishnu, Y Xing, and SA Liebhaber 2013 αCP Poly (C) binding proteins act as global regulators of alternative polyadenylation. *Molecular and cellular biology* **33** 2560-2573.

Jia, L, and Y Sun 2009 RBX1/ROC1-SCF E3 ubiquitin ligase is required for mouse embryogenesis and cancer cell survival. *Cell division* **4** 16.

Jodar, M, E Sendler, SI Moskovtsev, CL Librach, R Goodrich, S Swanson, R Hauser, MP Diamond, and SA Krawetz 2015 Absence of sperm RNA elements correlates with idiopathic male infertility. *Science translational medicine* **7** 295re296-295re296.

Johnson, GD, P Mackie, M Jodar, S Moskovtsev, and SA Krawetz 2015 Chromatin and extracellular vesicle associated sperm RNAs. *Nucleic acids research* **43** 6847-6859.

Kent, WJ, CW Sugnet, TS Furey, KM Roskin, TH Pringle, AM Zahler, and D Haussler 2002 The human genome browser at UCSC. *Genome research* **12** 996-1006.

Kitsou, C, L Lazaros, S Bellou, G Vartholomatos, P Sakaloglou, E Hatzi, S Markoula, K Zikopoulos, T Tzavaras, and I Georgiou 2016 Exogenous retroelement integration in sperm and embryos affects preimplantation development. *Reproduction* **152** 185-193.

For Review Only

Kono, T, Y Obata, Q Wu, K Niwa, Y Ono, Y Yamamoto, ES Park, J-S Seo, and H Ogawa 2004 Birth of parthenogenetic mice that can develop to adulthood. *Nature* **428** 860-864.

Krueger, F 2015 Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Lee, MT, AR Bonneau, and AJ Giraldez 2013 Zygotic genome activation during the maternal-to-zygotic transition. *Annual review of cell and developmental biology* **30** 581-613.

Levine, H, Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Mindlis, I., Pinotti, R., Swan, R 2017 Temporal trends in sperm count: a systematic review and meta-regression analysis,. *Human Reproduction Update*.

Li, H, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, and R Durbin 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25** 2078-2079.

Liao, Y, GK Smyth, and W Shi 2013a featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* btt656.

Liao, Y, GK Smyth, and W Shi 2013b The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41** e108-e108.

Marlow, FL 2010. Maternal control of development in vertebrates. pp. 1-196: Morgan & Claypool Life Sciences.

McCall, CM, J Hu, and Y Xiong 2005 Recruiting substrates to cullin 4-dependent ubiquitin ligases by DDB1. *Cell Cycle* **4** 27-29.

Miller, D 2015 Confrontation, Consolidation, and Recognition: The Oocyte's Perspective on the Incoming Sperm. *Cold Spring Harbor perspectives in medicine* **5** a023408.

Ostermeier, GC, D Miller, JD Huntriss, MP Diamond, and SA Krawetz 2004 Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429** 154-154.

Paranjpe, SS, UG Jacobi, SJ van Heeringen, and GJC Veenstra 2013 A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. *BMC genomics* **14** 762.

Peaston, AE, AV Evsikov, JH Graber, WN De Vries, AE Holbrook, D Solter, and BB Knowles 2004 Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Developmental cell* **7** 597-606.

For Review Only

Potireddy, S, R Vassena, BG Patel, and KE Latham 2006 Analysis of polysomal mRNA populations of mouse oocytes and zygotes: dynamic changes in maternal mRNA utilization and function. *Developmental biology* **298** 155-166.

Quinlan, AR, and IM Hall 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26** 841-842.

Richter, JD, and P Lasko 2011 Translational control in oocyte development. *Cold Spring Harb Perspect Biol* **3** a002758.

Robinson, MD, DJ McCarthy, and GK Smyth 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26** 139-140.

Robinson, MD, and A Oshlack 2010 A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11** R25.

Rodgers, AB, CP Morgan, NA Leu, and TL Bale 2015 Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences* **112** 13699-13704.

Sahakian, E, JJ Powers, J Chen, SL Deng, F Cheng, A Distler, DM Woods, J Rock-Klotz, AL Sodre, and J-I Youn 2015 Histone deacetylase 11: a novel epigenetic regulator of myeloid derived suppressor cell expansion and function. *Molecular immunology* **63** 579-585.

Sato, M, and K Sato 2013 Dynamic regulation of autophagy and endocytosis for cell remodeling during early development. *Traffic* **14** 479-486.

Saunders, CM, MG Larman, J Parrington, LJ Cox, J Royse, LM Blayney, K Swann, and FA Lai 2002 PLCζ: a sperm-specific trigger of Ca 2+ oscillations in eggs and embryo development. *Development* **129** 3533-3544.

Skerget, S, MA Rosenow, K Petritis, and TL Karr 2015 Sperm Proteome Maturation in the Mouse Epididymis. *PloS one* **10** e0140650.

Stitzel, ML, and G Seydoux 2007 Regulation of the oocyte-to-zygote transition. *Science* **316** 407-408.

Stoeckius, M, D Grün, M Kirchner, S Ayoub, F Torti, F Piano, M Herzog, M Selbach, and N Rajewsky 2014a Global characterization of the oocyte□to□ embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. *The EMBO journal* e201488769.

Stoeckius, M, D Grün, and N Rajewsky 2014b Paternal RNA contributions in the Caenorhabditis elegans zygote. *The EMBO journal* **33** 1740-1750.

Tadros, W, and HD Lipshitz 2009 The maternal-to-zygotic transition: a play in two acts. *Development* **136** 3033-3042.

For Review Only

Tang, F, C Barbacioru, E Nordman, B Li, N Xu, VI Bashkirov, K Lao, and MA Surani 2010 RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature protocols* **5** 516-535.

Wang, G, Y Guo, T Zhou, X Shi, J Yu, Y Yang, Y Wu, J Wang, M Liu, and X Chen 2013 In-depth proteomic analysis of the human sperm reveals complex protein compositions. *Journal of proteomics* **79** 114-122.

Wang, S, Z Kou, Z Jing, Y Zhang, X Guo, M Dong, I Wilmut, and S Gao 2010 Proteome of mouse oocytes at different developmental stages. *Proceedings of the National Academy of Sciences* **107** 17639-17644.

Warde-Farley, D, SL Donaldson, O Comes, K Zuberi, R Badrawi, P Chao, M Franz, C Grouios, F Kazi, and CT Lopes 2010 The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research* **38** W214-W220.

Xue, Z, K Huang, C Cai, L Cai, C-y Jiang, Y Feng, Z Liu, Q Zeng, L Cheng, and YE Sun 2013 Genetic programs in human and mouse early embryos revealed by single-cell RNA [thinsp] sequencing. *Nature* **500** 593-597.

Yuan, S, A Schuster, C Tang, T Yu, N Ortogero, J Bao, H Zheng, and W Yan 2015 Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development* dev. 131755.

Yurttas, P, E Morency, and SA Coonrod 2010 Use of proteomics to identify highly abundant maternal factors that drive the egg-to-embryo transition. *Reproduction* **139** 809-823.

For Review Only

Figure legends

Figure 1. Venn diagrams for the cross-representation of sperm, MII egg and zygote RNAs. Panel **A** shows the overlaps between all RNAs, aggregated from all biological replicates at ≥ 10 CPM from each of the three sources. Panel **B** shows similar overlaps, except that the selected lists for MII eggs and zygotes were obtained after testing with Edge R, which normalises the data and identifies differentially expressed RNAs that are significantly 'up' in MII eggs or in zygotes. These over-represented RNAs are then added to the list of RNAs common to both MII eggs and zygotes. Complete RNA lists are provided in the supplementary information.

Figure 2. GeneMANIA network nodes. The nodes represent paternal (blue) and maternal (black) factors and their predicted interactions (interconnecting lines). Colours signify the interaction type including co-expression (purple), physical interactions (red), shared protein domains (green), co-localisation (blue) and all predicted interactions (orange). Nodes with boxes around them belong to the E3 ubiquitin ligase complex. Other encircling borders indicate additional inter-parental interactions supported by the RNA-seq data (see text for details).

Figure 3. Alignment of RNA sequencing reads (pile ups) across representative components of the E3 ubiquitin ligase complex. As tracked on the UCSC genome browser for all spermatozoal (sperm 1-3), egg (MII 1-4) and zygote (PCZ 1-3) biological replicates, reads for components of the E3 ubiquitin ligase complex are shown for *Fbxo2*♂ (**A**) alongside *Cul1*♀ (**B**), *Rbx1*♀ (**C**) and *Skp1a*♀ (**D**). Genes are depicted at the foot of each diagram with exons shown as filled blocks.

Figure 4. Real-time qPCR. Real-time qPCR was carried out on 400 pg of cDNA using primers for the five RNA-seq predicted paternal factors *Hdac11*, *Fbxo2*, *Zfp821, Pcbp4* and *Map1lc3a*, the maternal *Hdac2* and the testis-expressed somatic cell RNA control *Maged2*. QPCR products are shown for sperm (S), MII eggs (M), Zygotes (Z) and Testis (T) cDNAs in panel **A** with corresponding Ct values shown in panel **B.** A DNA ladder is shown for gel calibration with 100 bp and 500 bp markers indicated. Note that very high Ct values (>37) corresponding with samples not supporting specific PCR products and assumed to be PCR artifacts are plotted as 0.

For Review Only

Figure 5. Alternative potential pathways for a paternal contribution to the zygote. The first panel **(A**) depicts the MII egg and the spermatozoon just before fertilisation with i, a metabolic pathway that needs one or more paternal factors to be fulfilled (either RNA or protein); ii, a protein complex that needs a paternal factor to be functional; iii, the incoming fertilizing spermatozoon. The second panel (**B**) depicts the MII egg and spermatozoon after fertilization with, i a functional metabolic pathway following the insertion of a paternal factor; ii, an activated protein complex due to the addition of the missing paternal factor, such as *Fbxo2* in the Rbx1-E3 ubiquitin ligase. The disintegrating sperm membrane with arrows illustrating the released sperm-borne factors into the MII ooplasm is also shown (iii).

Supplementary Figure 1. Bioanalyser traces for the three sperm RNA pools are shown. Note the absence of 28S and 18S rRNA peaks, low R.I.N. and the generally short RNAs that make up the profile, all typical of sperm RNA and showing no evidence of contamination from other (somatic) cell sources.

Supplementary Figure 2. Gene ontology descriptions for parent bioprocesses derived from all RNAs (≥10 CPM) reported in sperm (green), MII eggs (blue) and zygotes (orange). The general similarity between MII eggs and zygotes compared with sperm is clear. This figure is best downloaded for onscreen viewing and magnification.

Supplementary Figure 3. Alignment of RNA sequencing reads (pile ups) across other paternally and maternally expressed components. As tracked on the UCSC genome browser for all spermatozoal (sperm 1-3), egg (MII 1-4) and zygote (PCZ 1-3) biological replicates, reads are shown for *Hdac11*♂ (**A**), *Bub1b*♀ (**B**), *Mapl1c3a*♂ (**C**), *Map1b*♀ (**D**), *Atg3*♀ (**E**), *Pcbp4*♂(**F**), *Hnrnpk*♀ (**G**), *Qk*♀ (**H**) and *Zfp821*♂ (**I**) are shown. Genes are depicted at the foot of each diagram with exons shown as filled in blocks. A box indicates the LTR in *Hdac11*

For Review Only

Table 1. Top biological processes for source and differentially expressed RNAs. RNA lists from each source (sperm, MII eggs and zygotes) and differentially expressed MII-Zygote genes flagged by EdgeR analysis (MII up and PCZ UP) were submitted to DAVID for ontological analysis. Bioprocesses are reported alongside uncorrected p values, Benjamini corrected p values and False Discovery Rates (FDRs).

Table 2. Paternally-derived factors and their potential maternal cofactors in *Mus musculus*. Column 1 shows spermatozoal factors with good exonic representation as revealed by RNA-seq and UCSC browsing. Column 2 gives the associated gene name and MGI accession number. Column 3 briefly depicts their functionality as described in UniProt. Column 4 shows the potentially interacting maternal factors as predicted by GeneMANIA.

Supplementary Table 1. Primer names, oligonucleotide sequences, annealing temperatures and expected product sizes for real-time qPCR.

Supplementary Table 2. Paternally-derived factors and their *Mus musculus* knock-out phenotypes. Column 1 shows spermatozoal genes with good exonic representation as revealed by RNA-seq and UCSC browsing. Column 2 indicates any available KO studies for the paternally-derived factors. Column 3 shows the maternal factors that may have potential interactions with the paternal cofactors based on GeneMANIA. Column 4 shows KO studies reported for maternal factors listed in column 3, with some conclusions for each study.

Full supporting supplementary information.xlsx (supplementary info). Details of original RefSeq accession numbers for the most highly expressed RNAs (≥10 CPM) in sperm, MII eggs and zygotes (RNA Lists (All) sheet). The gene ontology data for these lists is listed in the BP All sheet. The RNA Lists (EdgeR) sheet includes the most highly represented sperm RNAs with good exon coverage (column A) and lists of Edge R-determined more highly represented RNAs in MII eggs (column C), zygotes (column D) or not differentially represented in either (column E). The RNAs participating in the GeneMania network (Figure 2) are listed in column G. Column B indicates the sperm-specific RNAs chosen for further study. Other sheets contain more complete lists of functional annotation clustering and enrichment for sperm (FAC_Sperm), MII eggs (FAC_MII) and zygote (FAC_Zygote), RNAs common to MII egg and zygotes (FAC_MII&Zygote). The BP EdgeR sheet includes

For Review Only

bioprocesses for the selected sperm, MII egg and zygote RNA lists and RNAs in the GeneMANIA network. The BP EdgeR MII or PCZ Up sheet includes lists of differentially expressed RNAs higher in MII eggs or Zygotes.

For Review Only